Understanding Medical Education

EVIDENCE, THEORY, AND PRACTICE

THIRD EDITION

EDITED BY

Tim Swanwick | Kirsty Forrest | Bridget C. O'Brien

WILEY Blackwell

Understanding Medical Education

EVIDENCE, THEORY, AND PRACTICE

Understanding Medical Education

EVIDENCE, THEORY, AND PRACTICE

THIRD EDITION

EDITED BY

Tim Swanwick
Dean of Education and Leadership Development
NHS Leadership Academy
Health Education England, London, UK

Kirsty Forrest
Dean of Medicine
Faculty of Health Sciences and Medicine, Bond University, Gold Coast, QLD, Australia

Bridget C. O'Brien
Associate Professor, Department of Medicine
Education Researcher, Center for Faculty Educators
School of Medicine, University of California, San Francisco, CA, USA

WILEY Blackwell

This edition first published 2019

Edition History
(1e, 2010), (2e 2014).

Registered Office(s)
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Office
9600 Garsington Road, Oxford, OX4 2DQ, UK

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Library of Congress Cataloging-in-Publication Data

Names: Swanwick, Tim, editor. | Forrest, Kirsty, editor. | O'Brien, Bridget C., editor. | Association for the Study of Medical Education, sponsoring body.
Title: Understanding medical education : evidence, theory, and practice / edited by Tim Swanwick, Kirsty Forrest, Bridget C. O'Brien.
Description: 3rd edition. | Hoboken, NJ : Wiley-Blackwell, 2018. | Includes bibliographical references and index. |
Identifiers: LCCN 2018027439 (print) | LCCN 2018027968 (ebook) | ISBN 9781119373858 (Adobe PDF) | ISBN 9781119373834 (ePub) | ISBN 9781119373827 (pbk.)
Subjects: | MESH: Education, Medical–methods | Education, Medical–organization & administration
Classification: LCC R845 (ebook) | LCC R845 (print) | NLM W 18 | DDC 610.76–dc23
LC record available at https://lccn.loc.gov/2018027439

Cover Design: Wiley
Cover Image: ©Steve Debenport/E+/Getty Images

Set in 9/11pts Palatino by SPi Global, Pondicherry, India

10 9 8 7 6 5 4 3 2 1

Contents

Contributors

Lindsay Baker
Scientist
Li K Shing Knowledge Institute
Researcher
Centre for Faculty Development
St. Michael's Hospital
Assistant Professor
Faculty of Medicine
University of Toronto
Toronto, Ontario, Canada

Alexis Battista
Assistant Professor of Medicine
Department of Medicine & F. Edward Hébert School of Medicine
Uniformed Services University of the Health Sciences
Bethesda, MD, USA

Jo Bishop
Associate Dean
Student Affairs and Service Quality
Faculty of Health Sciences and Medicine
Bond University
Gold Coast, Queensland, Australia

Nicole J. Borges
Professor of Neurobiology and Anatomical Sciences
University of Mississippi Medical Center
Jackson, MS, USA

Katharine A. M. Boursicot
Director
Health Professional Assessment Consultancy, Singapore
Singapore

William P. Burdick
Associate Vice President for Education
Foundation for Advancement of International Medical Education and Research (FAIMER)
Clinical Professor of Department of Emergency Medicine
Drexel University College of Medicine
Philadelphia, PA, USA

Craig Campbell
Principal Senior Advisor, Competency-based Continuing Professional Development
Office of Specialty Education
Royal College of Physicians and Surgeons of Canada
Ottawa, Ontario, Canada

Madeline Carter
Senior Lecturer
Department of Psychology, Faculty of Health and Life Sciences
Northumbria University
Newcastle, UK

Olivia Carter-Pokras
Professor of Epidemiology and Associate Dean for Diversity and Inclusion
Department of Epidemiology and Biostatistics
School of Public Health
University of Maryland, College Park
College Park, MD, USA

Neville Chiavaroli
Senior Lecturer
Department of Medical Education
Melbourne Medical School
University of Melbourne
Melbourne, Australia

Jennifer Cleland
Professor of Medical Education Research
Centre for Health care Education Research and Innovation
University of Aberdeen
King's College
Aberdeen, UK

Deborah Cohen
Professor and Director
Student Support
Centre for Psychosocial Research, Occupational and Physician Health
School of Medicine
Cardiff University
Cardiff, UK

Ian Cooper
Language and Communications Specialist
School of Medicine
Cardiff University
Cardiff, UK

Sayra Cristancho
Scientist
Centre for Education Research & Innovation
Assistant Professor
Departments of Surgery and Faculty of Education
Schulich School of Medicine & Dentistry
Western University,
London, Ontario, Canada

Richard L. Cruess
Professor of Surgery
The Centre for Medical Education
McGill University
Montreal, Quebec, Canada

Sylvia R. Cruess
Professor of Medicine
The Centre for Medical Education
McGill University
Montreal, Quebec, Canada

Mary Jane Dalton
Senior Lecturer
School of Medicine and Public Health
University of Newcastle
Callaghan, New South Wales, Australia

Anique B. H. de Bruin
Associate Professor
School of Health Professions Education
Maastricht University
Maastricht, The Netherlands

André F. De Champlain
Chief Research Psychometrician
Research and Development, Medical Council of Canada
Ottawa, Ontario, Canada

Nisha Dogra
Emeritus Professor of Psychiatry Education
Greenwood Institute of Child Health, Department of Neuroscience, Psychology and Behaviour, College of Life Sciences
University of Leicester
Leicester, UK

Erik Driessen
Associate Professor of Medical Education
Department of Educational Development and Research
Faculty of Health, Medicine and Life Sciences
Maastricht University
Maastricht, The Netherlands

Steven J. Durning
Professor and Director
Division of Health Professions Education, Department of Medicine
Uniformed Services University of the Health Sciences
Bethesda, MD, USA

Rachel H. Ellaway
Professor of Medical Education
Department of Community Health Sciences
Co-Director
Office of Health and Medical Education Scholarship
Cumming School of Medicine
University of Calgary
Calgary, Alberta, Canada

Caroline Elton
Director
Career Planning for Doctors and Dentists
London, UK

Kevin W. Eva
Senior Scientist
Centre for Health Education Scholarship
Professor
Director of Education Research & Scholarship
Department of Medicine
University of British Columbia
Vancouver, British Columbia, Canada

Eamonn Ferguson
Professor of Health Psychology
University of Nottingham
Nottingham, UK

Kirsty Forrest
Dean of Medicine
Faculty of Health Sciences and Medicine
Bond University
Gold Coast, QLD, Australia

Della Freeth
Executive Director of Education
Royal College of Physicians
London, UK

Janet Grant
Emerita Professor
The Open University
Honorary Professor
University College London Medical School
Director
CenMEDIC (Centre for Medical Education in Context)
London, UK
Special adviser to the President, World Federation for Medical Education
Ferney-Voltaire, France

Matthew C. E. Gwee
Medical Educationalist and Chairman
International and Education Programs
Centre for Medical Education, Yong Lo Lin School of Medicine
National University of Singapore, Singapore
Singapore

Graeme Horton
Senior Lecturer
School of Medicine and Public Health
University of Newcastle
Newcastle, New South Wales, Australia

Wendy Hu
Associate Dean
Academic (learning & innovation)
Chair of Medical Education
School of Medicine
Western Sydney University
Penrith, New South Wales, Australia

Chien-Da Huang
Physician Educator
Chang Gung Medical Education Research Centre (CG-MERC)
Chang Gung Memorial Hospital
Linkou, Taiwan (ROC)
Deputy Director
Department of Internal Medicine
Associate Professor
Department of Thoracic Medicine and Medical Education
Chang Gung Memorial Hospital
Chang Gung University College of Medicine
Taipei, Taiwan (ROC)

Jan Illing
Professor of Medical Education Research
School of Medical Education
Newcastle University, Newcastle, UK

Dahn Jeong
Research Associate
Office of Continuing Professional Development
Department of Innovation in Medical Education
Faculty of Medicine, University of Ottawa
Ottawa, Ontario, Canada

Brian Jolly
Professor
School of Medicine and Public Health
University of Newcastle
Callaghan, New South Wales, Australia
School of Rural Medicine
University of New England
Armidale, Australia

David M. Kaufman
Professor
Faculty of Education
Associate Member
Faculty of Health Sciences
Associate Member
Department of Gerontology
Simon Fraser University
Burnaby, British Columbia, Canada

Tara J. Kennedy
Developmental Pediatrician and Clinical Leader
Pediatric Autism Rehabilitation Services
Stan Cassidy Centre for Rehabilitation
Fredericton, New Brunswick, Canada

Diane N. Kenwright
Head of Department
Department of Pathology and Molecular Medicine,
University of Otago
Wellington, New Zealand

Simon Kitto
Full Professor
Department of Innovation in Medical Education
Director of Research
Office of Continuing Professional Development
Faculty of Medicine, University of Ottawa
Ottawa, Ontario, Canada
Assistant Professor
Department of Surgery, University of Toronto
Toronto, Ontario, Canada

John Launer
Programme Director for Educational Innovation in Primary Care
(North Central and East London)
Health Education England
London, UK

Lorelei Lingard
Director and Senior Scientist
Centre for Education Research & Innovation
Professor
Department of Medicine
Schulich School of Medicine & Dentistry
Faculty of Education
Western University
London, Ontario, Canada

Bridget Lock
Associate Director of Medical Education
South London Health care NHS Trust
London, UK

Andrew Long
Vice President (Education)
Royal College of Paediatrics and Child Health
Honorary Senior Lecturer
University College London
Consultant Paediatrician
Great Ormond Street Hospital
London, UK

Chris Lovato
Professor
School of Population & Public Health
Director
Evaluation Studies Unit, Faculty of Medicine
University of British Columbia
Vancouver, British Columbia, Canada

Lauren A. Maggio
Associate Professor and Associate Director
Division of Health Professions Education, Department of Medicine
Uniformed Services University of the Health Sciences
Bethesda, MD, USA

Peter McCrorie
Professor of Medical Education
Institute of Medical and Biomedical Education
St George's University of London
London, UK
Founding Chair and Professor of Medical Education
Department of Medical Education, University of Nicosia Medical School
Nicosia, Cyprus

Judy McKimm
Professor and Director of Strategic Educational Development
College of Medicine
Swansea University Medical School
Swansea, UK

Lynn Monrouxe
Director
Chang Gung Medical Education Research Centre (CG-MERC)
Chang Gung Memorial Hospital
Linkou, Taiwan (ROC)

Sandra Monteiro
Assistant Professor
Department of Health Research Methods, Evidence and Impact
McMaster University
Hamilton, Ontario, Canada

Clare Morris
Reader in Medical Education Research and Development
Institute of Health Sciences Education,
Barts and the London School of Medicine and Dentistry
Queen Mary University of London
London, UK

Debra Nestel
Professor of Simulation Education in Health care
Monash University
Clayton, Victoria, Australia
Professor of Surgical Education
Department of Surgery
University of Melbourne
Melbourne, Victoria, Australia

Stella L. Ng
Scientist
The Wilson Centre
Director of Research
Centre for Faculty Development
Arrell Family Chair in Health Professions Teaching
St. Michael's Hospital
Assistant Professor
Faculty of Medicine
University of Toronto
Toronto, Ontario, Canada

John J. Norcini
President and CEO
Foundation for Advancement of International Medical
Education and Research (FAIMER)
Philadelphia, PA, USA

Geoff Norman
Professor Emeritus
Clinical Epidemiology and Biostatistics
McMaster University
Hamilton, Ontario, Canada

Bridget C. O'Brien
Associate Professor
Department of Medicine
Education Researcher
Center for Faculty Educators
University of California
San Francisco, CA, USA

Hirotaka Onishi
Assistant Professor
Department of International Cooperation for Medical Education
International Research Centre for Medical Education
University of Tokyo
Tokyo, Japan

Fiona Patterson
Founding Director
Work Psychology Group
London, UK
Principal Researcher
University of Cambridge
Cambridge, UK

Linda Peterson
Professor and Senior Evaluation Adviser
Retired, Faculty of Medicine
University of British Columbia
Vancouver, British Columbia, Canada

David Price
Senior Vice President
ABMS Research and Education Foundation
American Board of Medical Specialties
Chicago, IL, USA
Professor
Department of Family Medicine,
University of Colorado School of Medicine
Denver, CO, USA

Scott Reeves
Professor
Faculty of Health, Social Care and Education
Kingston University and St. George's University of London
London, UK

Trudie E. Roberts
Director of Leeds Institute of Medical Education
University of Leeds
Leeds, UK

Melody Rhydderch
GP Training Academic Development Lead
Wales Deanery
Cardiff, UK

Dujeepa D. Samarasekera
Director
Centre for Medical Education, Yong Lo Lin School of Medicine
National University of Singapore, Singapore
Singapore

Maggi Savin-Baden
Professor of Education
Institute of Education
University of Worcester
Worcester, UK

Lambert W.T. Schuwirth
Strategic Professor for Medical Education
Flinders University
Adelaide, Australia
Professor for Innovative Assessment
Maastricht University, Maastricht
The Netherlands
Professor of Medical Education
Chang Gung University
Taoyuan, Taiwan
Professor of Medicine (Education),
Uniformed Services University for the Health Professions
Bethesda, MD, USA

Matthew Sibbald
Assistant Professor
Department of Medicine
Director
Centre for Simulation-Based Learning
McMaster University
Hamilton, Ontario, Canada

Linda Snell
Professor of Medicine
Centre for Medical Education and Department of Medicine
McGill University
Montréal, Quebec, Canada

Daisuke Son
Assistant Professor
Department of Medical Education Studies
International Research Centre for Medical Education,
University of Tokyo
Tokyo, Japan

John Spencer
Emeritus Professor of Primary Care and Clinical Education
Newcastle University
Newcastle, UK

Yvonne Steinert
Professor of Family Medicine
Richard and Sylvia Cruess Chair in Medical Education
Director
Centre for Medical Education
Faculty of Medicine
McGill University
Montreal, Quebec, Canada

Tim Swanwick
Dean of Education and Leadership Development
NHS Leadership Academy
Health Education England
London, UK

Jools Symons
Patient & Public Involvement Manager
Faculty of Medicine and Health
Leeds University
Leeds, UK

Olle ten Cate
Professor
Directorate of Education and Training
Utrecht University
Utrecht, The Netherlands
Adjunct Professor of Medicine
University of California
San Francisco, CA, USA

Jill Thistlethwaite
Adjunct Professor
School of Communication
University of Technology
Sydney, Australia

Aliki Thomas
Associate Professor and Research Scientist
Centre for Medical Education
School of Physical and Occupational Therapy
Faculty of Medicine, McGill University
Montreal, Quebec, Canada

Cees P. M. van der Vleuten
Scientific Director
School of Health Professions Education
University of Maastricht
Maastricht, The Netherlands

Jan van Tartwijk
Professor of Education
Department of Education
Utrecht University
Utrecht, The Netherlands

Claire Vogan
Associate Professor
Swansea University Medical School
Swansea, Wales, UK

Marjo Wijnen-Meijer
Associate Professor
TUM Medical Education Center
Technical University of Munich
Munich, Germany

Tim Wilkinson
Professor of Medicine and Director of the MB ChB programme
Otago Medical School
University of Otago
Christchurch, New Zealand

Diana F. Wood
Director of Medical Education and Clinical Dean
University of Cambridge School of Clinical Medicine
Cambridge, UK

Zareen Zaidi
Associate Professor
Department of Medicine
University of Florida College of Medicine
Gainesville, FL, USA

Lara Zibarras
Associate Director for Research, Development and Dissemination
Work Psychology Group
Senior Lecturer in Organisational Psychology
City, University of London
London, UK

Foreword to the Third Edition

What can one say about a book that within less than 10 years is its third edition? The need for this new edition reflects that *Understanding Medical Education* is the authoritative and comprehensive resource in modern medical education practice. To borrow from Parmenides, 'nothing comes from nothing' and so it is useful to reflect on both the origins of *Understanding Medical Education* and how it has evolved through its various iterations. During the early 'noughties', the Association for the Study of Medical Education (ASME) approached leading experts in the field of medical education to contribute to a series of standalone monographs on their topics of expertise. These monographs proved incredibly popular, so much so that it became clear that there was a need for a definitive guide to medical education presented, for the first time, in a single core textbook. Having commissioned and edited the original series, Tim Swanwick was invited to take on the gargantuan task of bringing everything together in one place, approaching authors to revise their contributions in light of new research evidence and emerging thinking, and sourcing other well-known figures and rising stars as contributors. *Understanding Medical Education* was a cutting-edge 'one-stop shop' presented in simple language and applicable across the entire spectrum of health professions education. It was an instant 'hit', adopted rapidly by medical educators across five continents with translations available in a number of different languages.

Yet nothing stands still. After many centuries of little change, medical education and medicine have shifted dramatically in recent years. Medical practice, society, health care systems, and expectations from patients are changing, and medical education has to also change to keep up. For example, ways of working with patients and colleagues are different. There are changes in how we deliver education and training linked to changing health care practices, particularly fewer opportunities to learn in the workplace because of system changes such as regulated hours for junior doctors. There have been major advances in research and treatments, and hence views of what is good clinical practice. These rapid changes mean that medical education must prepare today's medical students and doctors in training to work in very different ways from those of the past. Best practice in medical curricula, methods of instruction, assessment, and so on have to change and evolve in order to reflect the needs of contemporary medical practice. It is extremely challenging for educators to keep up with the literature, read journal articles and book chapters: the synthesis of the latest, most relevant, and essential material in medical education is if anything more necessary today than it was at the time of the first edition.

The second edition of *Understanding Medical Education* was published only four years after the first, reflecting the rapidity of change. This third edition, has kept pace with the continuing and hectic evolution of medical education. The content illustrates the journey that medical education has taken over recent years, and hints at the challenges that lie ahead. The content also reflects the responsiveness of the *Understanding Medical Education* project, a feature which will help those delivering medical education and training reflect on how things have changed since they were in the classroom and clinic, and help them break free of what Whitehead and colleagues have called the 'carousel of ponies'. This colourful analogy suggests that there are returning themes in medical education, circling round and round in the continual rediscovery of discursive 'truths'. Getting off this carousel requires both knowledge and reflection. *Understanding Medical Education*'s five sections of Foundations, Teaching and Learning, Assessment and Selection, Research and Evaluation, and Faculty and Learners focus ostensibly on knowledge. Yet running throughout the book is also a strong acknowledgement of the necessity of considered and scholarly reflection on the process of medical education. By this I mean the need to think not just about the 'what' or 'how' (to assess in the workplace, to introduce portfolios, design a curriculum, etc.) but also the 'why' (are we introducing something new, what can we learn from pedagogic shifts, and so on). *Understanding Medical Education* provides a resource which will help educators reflect on the complexity of medical education, to question discourses and practices in a way which will help them develop as professionals and move medical education ever forwards.

Long-term fans will also notice that *Understanding Medical Education* has extended its editorial team. As the current Chair of the Association for the Study of Medical Education (ASME), and the person responsible for commissioning the third edition of UME, I believed it was critically important to ensure that the book explicitly reflected ASME's explicit 'UK-based internationally facing' mission. This mission is reflected in editors, contributors, and readers of ASME's journals, *Medical Education* and *The Clinical Teacher*, and our other indispensable resource, *Researching Medical Education*. In support of this aim, I was delighted to invite Bridget C. O'Brien from the US and Kirsty Forrest from Australia to join Tim Swanwick as co-editors. Their international collaboration on this edition illustrates the great benefits of working together to share knowledge and networks.

Understanding Medical Education synthesises the latest knowledge, evidence, and best practices in the field. It provides a snapshot of how far we have come as a field. It is the essential resource for established educators and those new to the field. This extensively revised and extended third edition should be on the desk of every medical educator.

Professor Jennifer Cleland
Centre for Health care Education Research and Innovation,
University of Aberdeen, Aberdeen, UK
Chair of the Association for the Study of Medical Education

Preface

Understanding Medical Education was launched by the Association for the Study of Medical Education as a series of monographs in September 2006. In 2010 these monographs were brought together into a single textbook, providing a unique and comprehensive guide to the theoretical and academic bases to modern medical education practice.

As well as providing practical guidance for clinicians, teachers, and researchers, *Understanding Medical Education* is designed to meet the needs of all newcomers to medical education, including those studying at certificate, diploma, or masters level; *Understanding Medical Education* aims to be both accessible and useful to the reader. The intention is that after reading one of the chapters the reader will not only be better informed about their field of interest, but able to assimilate their new knowledge into their clinical teaching or academic activities.

Following a rigorous process of expert peer review, this third edition sees major updates of all existing chapters and some completely new ones, including contributions on the science of learning, knowledge synthesis, and learner support and well-being. The third edition also comes with a brand new foreword from Professor Jennifer Cleland, Chair of Medical Education Research at the University of Aberdeen and Chair of Council for the Association for the Study of Medical Education.

Understanding Medical Education remains the first port of call for anyone engaged in medical education as an academic discipline. The book is a unique resource which should prove invaluable for anyone involved in the development of health care professionals, in whatever discipline, wherever they are in the world.

An online edition of the complete book together with individual chapter downloads is available at http://onlinelibrary.wiley.com.

Editors

Tim Swanwick, MA, FRCGP, MA(Ed), FAcadMEd, has a broad range of experience in health care education and is based in London (UK), where he is Dean of Education and Leadership Development for Health Education England and the NHS Leadership Academy. Tim has a variety of academic interests including work-based learning, faculty development, professional support, academic careers and clinical leadership, and has taught, researched and published widely.

Kirsty Forrest, MBChB, BSc Hons, FRCA, MMEd, FAcadMEd, FANZCA, is based in Gold Coast, Queensland, Australia, as Dean of Medicine at Bond University. She is co-author and editor of a number of medical textbooks. Kirsty has been involved in educational research and management for many years both at the undergraduate and postgraduate levels and she continues to work clinically as a consultant anaesthetist.

Bridget C. O'Brien, PhD, is an associate professor and education researcher at the University of California, San Francisco, where she supervises doctoral students in the UCSF-University of Utrecht doctoral program and directs the San Francisco Veterans Affairs Fellowship in Health Professions Education Evaluation and Research. She co-authored *Educating Physicians: A Call for Reform of Medical School and Residency* (2010) and has published numerous peer-reviewed research papers and articles. Her research focuses on workplace learning across the continuum of health professions education.

Association for the Study of Medical Education

The Association for the Study of Medical Education (ASME) was established in 1957 by the UK General Medical Council to promote and conduct research into medical education. ASME's goals are to:

- Promote high-quality research into medical education.
- Provide opportunities for developing medical educators.
- Disseminate good evidence-based educational practice.
- Inform and advise Governmental and other organisations on medical education matters.
- Develop relationships with other organisations and groupings in health care education.

ASME's mission is to meet the needs of teachers, trainers, and learners in medical education by supporting research-informed best practice across the continuum of medical education.

Acknowledgements

Producing a textbook such as this is a team effort, and thanks must be extended to the ASME Executive and *Understanding Medical Education* editorial advisory board for their advice and guidance, to the ASME team for their administrative support, and to James Watson and the editorial team at Wiley for helping to make the third edition of *Understanding Medical Education* such an impressive and attractive volume.

ICONS

Throughout the *Understanding Medical Education* series readers will come across a number of icons in the margin. These graphic devices serve to highlight certain insert boxes where the author wishes to take the reader off into a particular area in greater detail (Focus on), explore the evidence behind a particular concept (Where's the evidence), provide practical advice (How to) or summarise the main points of the paper (Key messages).

FOCUS ON

WHERE'S THE EVIDENCE?

HOW TO

KEY MESSAGES

Part 1

Foundations

1 Understanding Medical Education

Tim Swanwick

Dean of Education and Leadership Development, NHS Leadership Academy, Health Education England, London, UK

It was the nuclear physicist and father of the hydrogen bomb, Edmund Teller, who wrote (perhaps rather alarmingly) 'Confusion is no bad thing; it is the first step towards understanding' [1, p. 79]. Newcomers to the field of medical education could be forgiven for being confused. Medical education is a busy, clamorous place, where a host of pedagogical practices, educational philosophies, and conceptual frameworks collide. It is a place where academic journals vie for attention, institutions and professional bodies compete for political leverage, and the wheel of reform and 'improvement' revolves faster than, and often independently of, the cycle of evaluation and research. And it is a place of increasing accountability and regulation because of its proximity to one of the prime socio-political concerns of government, that of the health of its people.

It was the desire to develop evidence-based policy and practice in this complex arena that led to the establishment of the Association for the Study of Medical Education (ASME) in 1957. The past 60 years have seen a burgeoning of literature in the field. This is both a help *and* a challenge to the clinician taking on responsibilities for teaching, assessment, and educational supervision. The range and diversity of relevant theory and research are now almost overwhelming, and in 2006 ASME recognised the need for a succinct yet comprehensive guide to the vast literature now underpinning best practice in medical education. *Understanding Medical Education* aims to be that guide.

What is Medical Education?

Medical education as we know it today spans three sectors: undergraduate, postgraduate, and the continuing professional development of established clinicians. However, it has not always been that way, and Abraham Flexner – the centenary of whose seminal report on the transformation of the American medical school system was celebrated earlier this decade [2] – would not have recognised the attention currently given to the design, management, and quality assurance of structured training in the postgraduate years, still less the need to instigate regulatory systems to ensure the ongoing personal and professional development of practising clinicians.

Medical education's ultimate aim is to supply society with a knowledgeable, skilled, and up-to-date cadre of health care professionals who put patient care above self-interest, and who undertake to maintain and develop their expertise over the course of a lifelong career. Medicine has a privileged position in society and, as a result, medical education is itself set apart from the main body of higher education. In many countries it luxuriates in separate funding streams and higher rates of remuneration for its clinical teachers; it is the beneficiary of status and patronage through its colleges, academies, and professional institutions; and it is a formidably powerful, and predominantly conservative, political lobby, more than occasionally a source of frustration for those who seek to modernise health services.

Within the confines of this academic and political preserve lies the discipline of medical education; although one could question whether medical education is a discipline in its own right, or an idiosyncratic collection of concepts appropriated from other educational fields and perfused with a technical rationality borne out of the dominance of bioscience within medicine [3, 4]. There are certainly a number of predominant educational assumptions, such as experiential learning and reflective practice, and favoured curricular approaches borrowed from other fields – witness the enthusiastic transplantation of competency-based education from vocational training [5]. But medical education is not just a 'magpie', taking ideas wherever they can be found, but has made, and continues to make, its own significant advances and contributions to the wider educational literature. Many of these unique and major developments are expounded within this book: problem-based learning, simulation, structured assessments of clinical competence, supervision, and the use of technology to enhance learning, to name but a few.

Challenges and Preoccupations

Another characteristic of medical education is that it is, as Cooke and her colleagues note, 'in a perpetual state of unrest' [6, p. 1339]. A constant stream of reports issues from regulators, commissions, inquiries, and task forces – all urging reform. This may just reflect the sluggish response to change and innate conservatism of the profession and its educational institutions. This is not, as it happens, a new phenomenon. In the UK, George Pickering, writing as far back as 1956, offers us the wry observation that 'no country

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

has produced so many excellent analyses of the present defects of medical education as has Britain, and no country has done less to implement them' [7]. Britain is not alone in this regard and from the other side of the Atlantic, Warren Anderson – in a special centenary 'Flexner' edition of *Medical Education* – questions 'whether the current proliferation of literature about reforms in medical education can lead to real change, or whether it constitutes a self-referential agitation that, in the aggregate, holds little promise' [8, p. 29]. Despite such reservations, the frequency of such reports increases, and the clarion calls to action grow ever louder. So what are the current preoccupations of medical education and society's expectations of it?

To 'begin at the beginning'; getting the right students and later on the right trainees training in the right specialty is crucial. In a competitive and litigious environment, the importance of having demonstrably fair selection processes is unarguable. A good person/job fit is essential to productivity, quality, and job satisfaction. In Chapter 26, Fiona Patterson and her colleagues identify just how difficult getting all this right can be. Predicting who will make a good doctor is critically dependent on what the role of the doctor will be 10–15 years into the future, something that is increasingly uncertain. So are there generic attributes that we can select for? What selection methods should we use? And to encourage the recruitment of well-rounded practitioners, should entry to medical school be graduate only?

Having selected the right students and, with luck, matched the right trainees to the most suitable postgraduate training programme, how and what are they to learn, and how can the *quality* of their education and training be ensured? An array of approaches to teaching and learning are described in the central section of this book framed by a discussion by Janet Grant on approaches to curriculum (Chapter 5) and Linda Snell and colleagues on the importance of good instructional design (Chapter 6). A concise summary of relevant, and guiding, educational theory is provided by David Kaufman in Chapter 4, preceded by a summary of the emerging insights, for medical education, of the relatively recent field of cognitive neuroscience (Chapter 3). And in Chapter 7, Diane Kenwright and Tim Wilkinson address the thorny concept of 'quality' – how do we know what we're doing is any good?

One of medical education's evolving 'special interests' has been assessment. Indeed it is often involvement in professional assessment, either formative or summative, that first draws clinicians into the world of medical education. Chapters 20–25 recount the increasing sophistication of assessment instruments in medical education, how validity is ensured and standards are set, the growing acceptance of the need for programmatic approaches, and the perennial challenge in professional education of balancing assessment for learning and assessment for accountability.

It was Flexner's mentor, William Osler, who brought students and patients closer together through his educational philosophy that medicine was 'learned by the bedside and not in the classroom' [9, p. 188] and through the practical introduction of residency programmes. Both are now threatened by concerns over patient safety, expansion of medical student numbers, regulatory requirements on working hours, and a staggeringly accelerated patient throughput. Patients undergoing gall bladder operations in Osler's day were in hospital for several weeks – the procedure now is carried out on a day-patient basis. At almost every stage of training, learners see fewer patients, do less to them, and, as a consequence, find themselves increasingly unprepared for practice [10]. This, as pointed out by Clare Morris in Chapter 12 and by John Launer in Chapter 13, requires new ways of thinking about work-based learning and the mediating role of the trainer or supervisor.

A related concern is patient safety. Medicine is not only faster-paced, it is also more hazardous. As Cyril Chantler has succinctly put it: 'Medicine used to be simple, ineffective and relatively safe. Now it is complex, effective and potentially dangerous' [11, p. 1178]. One of the responses to reduced opportunities for contact with patients and more hazardous interventions has been the widespread adoption of simulation across all fields and stages of medical education. The availability of sophisticated technologies now enables high-fidelity reproduction of complex patient scenarios. Students and doctors in training no longer need to carry out procedures for the first time on real patients – the skills of ophthalmoscopy, venepuncture, and catheterisation can all be learned in the skills laboratory. Full-immersion scenarios also offer the opportunity to work on non-technical areas such as team working, leadership, and situational awareness. However, questions remain about transfer to the authentic setting – an issue that is explored in depth by Alexis Battista and Debra Nestel in Chapter 11.

Growing concerns over patient safety have influenced not only the way medicine is practised – with the widespread introduction of protocols, checks, and audit – but also the degree to which doctors are now publicly accountable. In the UK, for instance, high-profile cases (such as Bristol [12], Alder Hey [13], Shipman [14], and, more recently, the Francis Inquiry [15]) have ushered in a new era of public accountability, while 2013 saw the introduction of relicensing for all medical practitioners in Britain, with regulators coming under increasing and critical pressure [16]. Patient safety issues also permeate undergraduate medicine. Protecting patients within a teaching and learning environment, while producing competent doctors who will maintain their knowledge, attitudes, and skills, is a major challenge for those who design undergraduate curricula.

Increasing accountability is just one facet of a new social compact with patients; a compact that is no longer based on blind and unquestioning trust but on true partnership [17]. As John Spencer, writing with Judy McKimm and Jools Symons, highlights in Chapter 15, we see increased patient involvement across the board in both teaching and learning, and also in decision-making about how medical education is organised, governed, and its resources allocated. Patients are now also intimately involved in the selection and assessment of both undergraduate students and postgraduate trainees, and feedback from patients is a routine feature of continuing professional development and reaccreditation processes.

One of the corollaries of the above is that there is a growing recognition of the need to professionalise clinical teaching [18]. The pressures for this are channelled through professional bodies, but also arise from an increase in the expectations of students and doctors in training about the quality of the learning opportunities they are afforded. Clinical teachers and others with responsibilities for medical education increasingly look for academic support and accreditation of their expertise, and one of the target groups of *Understanding Medical Education* are newcomers to medical education, whether undergraduate or postgraduate, including those studying at certificate, diploma, and master's levels. As Yvonne Steinert describes in Chapter 36 – on faculty development – the professional credentialing of medical educators is a burgeoning industry in Europe and North America and reflects a more general trend of the 'professionalisation' of medical education. Professionalisation has produced a new breed of scholarly educators and, coming as they do from a bioscientific background, a desire for evidence-informed medical education practice.

This raises questions about the nature of medical education research and again, as is highlighted in the five chapters on research and evaluation (Chapters 27–31), we see worlds colliding. In a recent exchange in ASME's academic journal, *Medical Education*, a series of articles considered whether it is helpful to construe medical education as a medical or a social science [19, 20]. Monrouxe and Rees capture the essence of the debate:

> Medical education research has benefited from its association with 'hard' medical science in that this has encouraged the engagement of clinicians in research activities. However, this gain is offset by a particular loss represented by the failure (of some) to understand that medical education is about people, and the way we think, act and interact in the world. Medical education research is not a poor relation of medical research; it belongs to a different family altogether [20, p. 198].

Curricula at the undergraduate level continue to evolve. Postgraduate medical education too is also in the throes of perpetual curricular change, with many specialties previously taught to implicit and informal curricula now articulating explicit and public curriculum statements of intent for the first time. Curriculum delivery is also challenged by the emerging possibilities of technology, many of which are addressed in a new chapter by Rachel Ellaway in which she explores the relationship between technology and learning (Chapter 10).

There are macro-political concerns too, around the responsiveness of medical education to societal needs [21]. In Chapter 35, Nisha Dogra and Olivia Carter-Pokras consider medical education's engagement with increasing diversity – considering patients and citizens as well as students and the workforce. Changing demographics are also profoundly influencing patterns of demand, with developed countries already experiencing the effects of an ageing population with complex health care needs. And across the increasingly interdependent world, we see a health inequalities gap that shows no signs of narrowing, with both emerging and developed health care systems struggling to cope [22]. Rising patient expectations and an ease of access to information present challenges not only in how care is delivered, but where and by whom. There are nostradamic predictions of future global shortages of health care workers [23] – an 18 million shortfall by 2035 – with little sign of a reversal of the maldistributive trend of doctors eschewing remote and rural locations in favour of large conurbations, and an imbalance of education and training outputs causing shortages in generalist and community-based specialties [24]. Managers within all health care systems are also waking up to the fact that the majority of their future employees already work in their health services and that significant investment may need to be diverted from training new and inexperienced practitioners into developing and supporting their existing workforce. Chapter 19 examines the complex issues that surround continuing professional development and there is an acknowledgement of the need to retain and support learners and staff, and provide support for their career decisions, in Chapters 32–34.

In Chapter 17, a new addition to this volume, Sylvia and Richard Cruess explore a central concern in medical education – the development of professional identity. But, what is 'a doctor' (or any other health care professional, for that matter)? With significant overlaps in knowledge and skills developing, what unique features does a doctor bring to the bedside or office, and what do we mean by a professional in the twenty-first century? Friedson argues that the professions, societal groups based on expertise, altruism, and self-scrutiny, will never disappear, but will merely shrink in size, as much of their work is taken on by a deprofessionalised operating core of medical technicians [25]. Others, such as Donald Berwick, disagree and see 'the reinvention of professionalism in a world on new terms of engagement; complexity, interdependence, pervasive hazard, a changing distribution of power and control and borne on the back of technology, distributed, democratised capacities …' [26, p. 130].

What is certain is that at no point in the past has the medical profession had to engage so actively with these debates, and the question 'What are we educating for?' has never been so important, something that my co-editors, Bridget C. O'Brien and Kirsty Forrest, and their colleagues explore in Chapter 2.

Scholarship and the Pursuit of Excellence

Understanding Medical Education began life as a series of free-standing monographs. The aim of the series was to provide an authoritative, up-to-date, and comprehensive resource summarising the theoretical and academic bases to modern medical education practice. It is now a best-selling textbook worldwide and although the majority of its expert authors come from Europe, Australasia, and North America, it offers a global perspective on contemporary practice and scholarship.

Boyer's expanded definition of 'scholarship' takes us beyond the narrow confines of research to consider the need to recognise and reward not only the scholarship of 'discovery' but also to recognise and reward the integration of new knowledge, its application to social practice, and teaching and learning [27]. This is a hugely important distinction for medical education, as the vast majority of medical educators are not researchers, nor indeed do they have the opportunity to work across disciplinary boundaries to integrate new knowledge. What they can be, and often are, are excellent teachers and scholarly agents of change and improvement within medical education (see Chapter 37). This highlights a perennial problem in medical education, namely that funding for academic institutions – despite recent attempts to redress the issue [28] – is linked strongly to research outputs. Similarly, teaching in clinical settings usually plays 'second fiddle' to clinical productivity. This has led to a situation where both academic and service institutions continue to emphasise staff involvement in activities other than teaching, such that teaching remains largely unrewarded and unrecognised. This is a challenge that professional bodies such as the UK's Academy of Medical Educators have set out to address [29].

Medical education is complicated, contested, and political. In a complex, uncertain, and networked world we need to make the best decisions about education, training, and development that we can and, as our final chapter outlines, engage in the leadership of change and improvement in an informed and intelligent way. For that, we need both scholarly medical educators *and* educational scholars. We hope that this book will continue to contribute to their development.

References

1 Teller, E., Teller, W., and Talley, W. (1991). Hypotheses non fingo. In: *Conversations on the Dark Secrets of Physics*, 69–80 chapter 5. Cambridge, MA: Perseus Publishing.

2 Lexner, A. (1910). *Medical Education in the United States and Canada: A Report to the Carnegie Foundation for the Advancement of Teaching*. New York: Carnegie Foundation for the Advancement of Teaching.

3 Swanwick, T. (2013). Doctors, science and society. *Medical Education* 47: 7–9.

4 Whitehead, C. (2013). Scientist or science-stuffed? Discourses of science in North American medical education. *Medical Education* 47: 26–32.

5 Royal College of Physicians and Surgeons of Canada (2014). Competence by Design: Reshaping Canadian Medical Education. http://www.royalcollege.ca/rcsite/documents/educational-strategy-accreditation/royal-college-competency-by-design-ebook-e.pdf (accessed 31 May 2018).

6 Cooke, M., Irby, D.M., Sullivan, W., and Ludmerer, K.M. (2006). American medical education 100 years after the Flexner report. *The New England Journal of Medicine* 355: 1339–1344.

7 Pickering, G.W. (1956). The purpose of medical education. *British Medical Journal* 2 (4968): 113–116.

8 Anderson, W. (2011). Outside looking in: observations on medical education since the Flexner report. *Medical Education* 45: 29–35.

9 Osler, W. (2003). *The Quotable Osler* (ed. M.E. Silverman, T.J. Murray, and C.S. Bryan). Philadelphia, PA: American College of Physicians – American Society for Internal Medicine.

10 Illing, J., Davies, C., and Bauldauf, B. (2008). *How Prepared Are Medical Graduates to Begin Practice. A Comparison of Three Diverse Medical Schools. Report to Education Committee of the General Medical Council*. London: General Medical Council.

11 Chantler, C. (1999). The role and education of doctors in the delivery of healthcare. *Lancet* 353: 1178–1181.

12 Bristol Royal Infirmary Enquiry (2001). Learning from Bristol: The Report of the Public Inquiry into Children's Heart Surgery at the Bristol Royal Infirmary 1984–1995. http://webarchive.nationalarchives.gov.uk/20090811143822/http://www.bristol-inquiry.org.uk/final_report/the_report.pdf (accessed 31 May 2018).

13 Redfern, M. (2001). The Royal Liverpool Children's Inquiry Report (The Alder Hey Report). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/250934/0012_ii.pdf (accessed 31 May 2018).

14 Smith, J. (2004). Shipman Fifth Report. Safeguarding Patients: Lessons from the Past – Proposals for the Future. http://webarchive.nationalarchives.gov.uk/20090808160144/http://www.the-shipman-inquiry.org.uk/fifthreport.asp (accessed 31 May 2018).

15 Francis, R. (2013). *Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry*. London: Stationery Office. https://www.gov.uk/government/publications/report-of-the-mid-staffordshire-nhs-foundation-trust-public-inquiry (accessed 4 August 2017).

16 Professional Standards Authority (2015). Rethinking Regulation. www.professionalstandards.org.uk/publications/detail/rethinking-regulation (accessed 4 August 2017).

17 Karazivan, P., Dumez, V., Flora, L. et al. (2015). The patient-as-partner approach in health care: a conceptual framework for a necessary transition. *Academic Medicine* 90 (4): 437–441.

18 Swanwick, T. (2009). Teaching the teachers – no longer an optional extra. *British Journal of Hospital Medicine* 70: 176–177.

19 Bligh, J. and Brice, J. (2008). What is the value of good medical education research? *Medical Education* 42 (7): 652–653.

20 Monrouxe, L.V. and Rees, C.E. (2009). Picking up the gauntlet: constructing medical education as a social science. *Medical Education* 43: 196–198.

21 Boelen, C. (2016). Why should social accountability be a benchmark for excellence in medical education? *Educación Médica* 17 (3): 101–105.

22 Frenk, J., Chen, L., Bhutta, Z.A. et al. (2010). Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. *Lancet* 376: 1923–1958.

23 World Health Organization (2016). Working for Health and Growth: Investing in the Health Workforce. Report of the High-Level Commission on Health Employment and Economic Growth. http://www.who.int/hrh/com-heeg/reports/en (accessed 4 August 2017).

24 Specialty and Geographic Distribution of the Physician Workforce: What Influences Medical Student & Resident Choices? http://www.graham-center.org/dam/rgc/documents/publications-reports/monographs-books/Specialty-geography-compressed.pdf (accessed 4 August 2017).

25 Friedson, E. (1988). *Profession of Medicine: A Study of the Sociology of Applied Knowledge*. Chicago, IL: University of Chicago Press.

26 Berwick, D. (2009). The epitaph of profession. *British Journal of General Practice* 59: 128–131.

27 Boyer, E.L. (1997). *Scholarship Reconsidered: Priorities of the Professoriate*. San Francisco, CA: Jossey-Bass.

28 Higher Education Funding Council for England (2016). The Teaching Excellence Framework. www.hefce.ac.uk/lt/tef (accessed 4 August 2017).

29 Bligh, J. and Brice, J. (2007). The academy of medical educators: a professional home for medical educators in the UK. *Medical Education* 41: 625–627.

2 A Global View of Structures and Trends in Medical Education

Bridget C. O'Brien[1], *Kirsty Forrest*[2], *Marjo Wijnen-Meijer*[3], *and Olle ten Cate*[1,4]

[1]Department of Medicine, Center for Faculty Educators, University of California, San Francisco, CA, USA
[2]Faculty of Health Sciences and Medicine, Bond University, Gold Coast, QLD, Australia
[3]TUM Medical Education Center, Technical University of Munich, Munich, Germany
[4]Center for Research and Development of Education, University Medical Center, Utrecht University, Utrecht, The Netherlands

KEY MESSAGES

- The educational pathway from secondary school to unrestricted medical practice shows roughly six structural routes worldwide.
- All pathways will likely be affected by educational system innovations, globalisation, health care systems, social and cultural values, and technology.
- Each of these forces pushes and pulls medical education in different directions, which results in disparate views and uncertainty about the purpose of medical education.
- Change is one constant feature of medical education that we can anticipate. The speed of developments in health care and education will require programmes, learners, and educators to adapt throughout the continuum of training and practice, as a core quality.

Introduction

This book, the third edition of *Understanding Medical Education*, aims to provide a more global perspective on medical education. This chapter provides context for subsequent chapters. In the first section we describe six structural models of medical education around the world. In the second section, we consider the purpose of medical education and the complexity of defining and working toward a shared sense of social accountability in an increasingly globalised world. In the third section, we discuss current trends in medical education, identified by thought leaders in the field. We speculate where these trends may take us in the next 10 years, and then conclude with some overarching reflections on the themes presented in the chapter and questions for further consideration.

Medical Education Pathways Worldwide

The training of medical doctors is well established in virtually every country around the globe; to the public, physicians are physicians, no matter where they train. Yet when we look in detail, the pathways students must follow to become licensed appear to vary considerably. In many industrialised societies, the current structure of medical training was established between 100 and 150 years ago, when university studies in medicine were combined with the guild-like models of barber-surgeon training. A theoretical training phase followed by a phase of practical apprenticeships became a dominant model in the first half of the twentieth century. After World War II a large expansion of postgraduate medical specialty training emerged, and in parallel newer educational models of undergraduate education were introduced. Several solutions to transition problems, from theory to practice, from undergraduate to postgraduate, and from training to unrestricted practice were created. As these innovations in the medical education pathway did not occur at the same time in all countries, international and even regional differences within countries became apparent, with possible differences in outcomes [1, 2].

Additionally, countries and international regions have their own views on how best to educate doctors to serve the needs of their populations. Influential models arose from: the British model, influencing predominantly the Commonwealth countries; the North American model, influencing several emerging countries; and the continental European model. In Europe, all European Union (EU) countries must comply with EU rules regulating the internal market, including the mutual recognition of professional diplomas, based on rules that prescribe some features of medical training [3].

Despite increasing international communications about medical education through dedicated medical education journals, conferences, associations, a World Federation for Medical Education (WFME) [4], and organisations and initiatives devoted to or impacting international development of education such as FAIMER [5–7], the pathways to

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

medical practice remain distinctly different among countries. Information about these differences is important because of the growing mobility of students and graduates and the corresponding need to understand what level of performance and experience diplomas and qualifications signify [8–10].

To supply this much needed information, Wijnen-Meijer and colleagues carried out a qualitative questionnaire study among well-informed medical educators in several countries. This led to an overview of structures and terminologies in 40 countries, published in 2013 [11]. This chapter adds 10 more countries to the 2013 data set, for a total of 50 countries. Most questionnaire responses were collected by e-mail and supplemented with information obtained at international conferences. Well-informed respondents answered questions about the different stages of medical education in their country, the length of these stages, the point at which unrestricted practice is allowed, and any additional requirements such as exams.

Wijnen-Meijer and colleagues found six dominant pathways through medical education that they called 'routes' (see Figure 2.1). In most countries students enter medical school directly after finishing secondary school (Routes I though IV). Routes V and VI describe pathways for which a bachelor's degree is required. In many countries graduates can enter residency directly after finishing medical school (Route I and V), while in other countries graduates must first finish an internship or mandatory social service or both. Of note, the six pathways contain much variation within their general structures and within countries multiple routes may exist. For example, as shown in Figure 2.2, the length of postgraduate (residency) training varies among specialties within one country as well as within the same specialty across countries. Also, the requirements for unrestricted practice can range from attainment of the MD degree to one year of specialty training to completion of specialty training and fellowship.

Similar to structure, terminology differs from country to country and can pose challenges for translation of educational levels across borders or comparison of curricula, instruction, and outcomes internationally. Box 2.1 describes some of the commonly used terminology in medical education worldwide. These terms are used variably throughout the book, reflecting the international perspective of individual chapter authors. Box 2.2 identifies the degrees awarded in medical education.

Though appealing on many levels, attempts to harmonise medical education across countries have had limited success. For example, in 1999, the governments of all EU countries and some surrounding countries agreed to harmonise all of higher education in three phases: bachelor, master, and doctorate [12]. This Bologna Process was well accepted by all of higher education in 48 countries with the exception of medical education in all but 7 countries. Those seven countries now organise 'undergraduate' medical education in two phases (bachelor and master), while all of the others in the agreement do not. The attempt created more disparity than harmony [13, 14]. The WFME takes a different approach. Rather than attempting to harmonise the structure of medical education, the WFME provides a consensus-based set of 106 basic standards and 90 standards for quality improvement that provide 'a template for medical schools and other providers of medical education, and the agencies which accredit them to define institutional, national and regional standards,

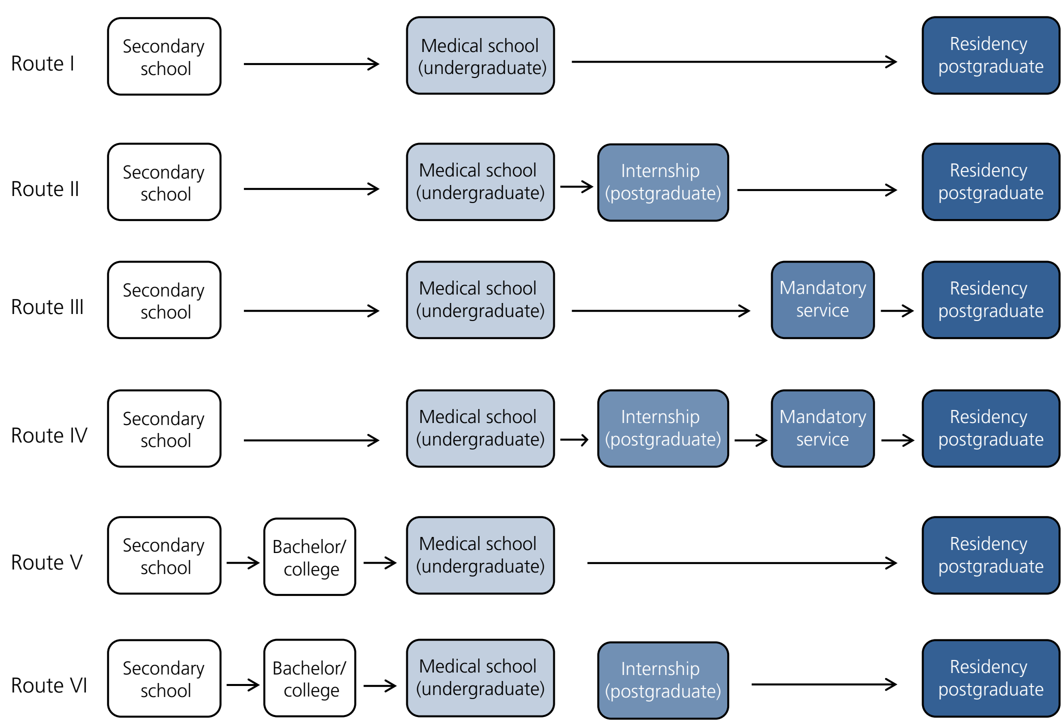


Figure 2.1 Six routes to unrestricted practice.

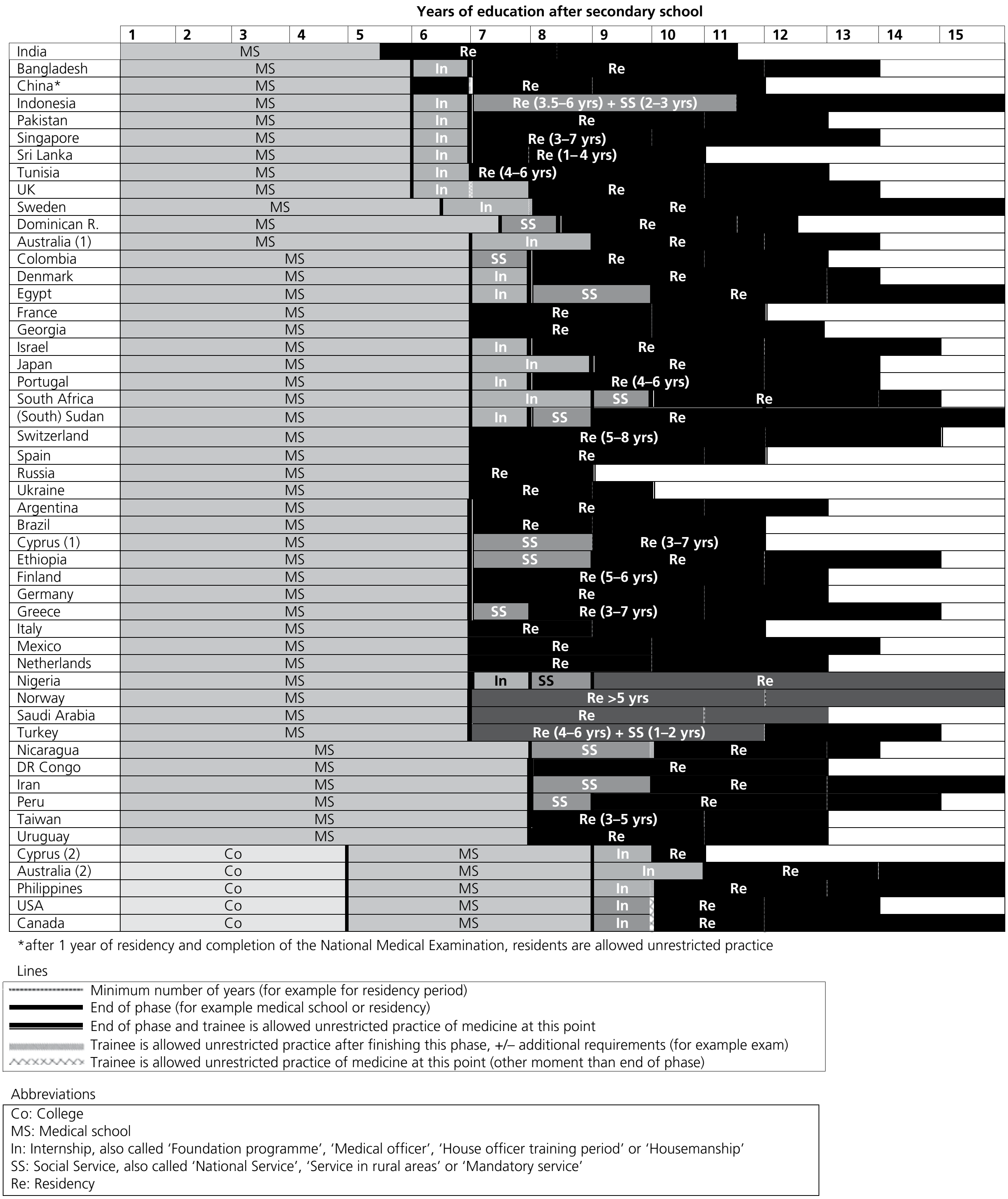


Figure 2.2 Medical education comparisons by country: phases and duration.

and to act as a lever for quality improvement' [15]. This approach aims to enable, or even foster, diversity so educational programmes across the continuum of medical education can accommodate economic, political, social, and cultural contexts while having an internationally recognised framework to guide curriculum development, learner assessment, faculty development, and programme evaluation.

There may not be a compelling reason or any chance of success in forcing countries to adopt similar structures or terminologies, if only because it cannot be determined

BOX 2.1 FOCUS ON: Common terminologies in medical education

Term	Description
Basic medical education	The portion of medical education that occurs in medical school; also called undergraduate medical education.
Chief resident	A selected senior resident with administrative and teaching responsibilities toward junior medical trainees.
Clerk	A medical student on a clinical rotation or in clerkship phase.
Clerkship	A period of one or more weeks of (clinical) experience in a medical specialty during medical school.
Consultant	Senior hospital-based physician who has completed residency.
Fellowship	A training period in a medical sub-specialty that occurs one or more years after completion of general specialty training.
Foundation doctor	A trainee in a Foundation Programme (UK).
Foundation programme	A two-year, clinical training programme after medical school and before postgraduate medical training in the UK.
Graduate medical education	Used in North America. Synonymous with postgraduate medical education.
House officer	Period of practice between medical school and full registration in several countries. Also called: medical officer or housemanship or a postgraduate medical trainee.
Intern	A trainee in a clinical training period directly after medical school, usually identical to the first year of residency training.
Medical bachelor	The first three years of medical school in countries that have signed the EU Bologna agreement and have included medical education in this structure.
Medical master	The second three years of medical school in countries that have signed the EU Bologna agreement and have included medical education in this structure.
Medical school	The institutional organisation that offers an undergraduate medical education programme, usually overlapping with the medical faculty of a university; sometimes used as undergraduate medical education phase.
Medical student	A person enrolled in an undergraduate medical education programme.
Physician	A graduate from a medical school who is formally licensed to practice medicine.
Placement	Synonymous with rotation.
Postgraduate medical education	Usually synonymous with residency training, but in Australia and New Zealand the phase after initial higher education.
Registrar	A medical trainee in a postgraduate education programme after registration as MBBS or MBChB.
Residency	A postgraduate training programme to become a medical specialist.
Resident	A medical trainee in a postgraduate education programme.
Rotation	A period or one or more weeks of experience with a medical specialty during medical school or residency.
Senior house officer	A year (or two) after house officer prior to specialist training.
Social service	A period of mandatory clinical service after medical school, usually as part of an agreement with the school or funding body, in a region in need of medical service (also called national service).
Specialist	Physician who has finished residency in a specific specialty of medicine.
Trainee	An individual who is in a formal educational or training programme at any level of medical education; often a term confined to the clinical phases of education.
Undergraduate education	Either initial higher education at bachelor level preceding undergraduate medical education, or medical school education.

which are better than others. But, as will become clear in the section on globalisation below, international interactions about medical education are naturally becoming much more intense. Schools and countries learn though publications, conferences, and student and faculty exchanges, and it may be expected that through natural processes of curriculum development, informed by what other countries do, that medical education will gradually converge to more similar models.

Purposes and Priorities in Medical Education

The pathways and terminologies described in the previous section reflect educational systems designed to meet societal needs for health care. These systems are steeped in cultural, historical, political, and economic contexts that have changed substantially since many of these systems were first established. Yet the basic systems of medical education

BOX 2.2 FOCUS ON: Degrees in medical education

Degrees	
BSc	Bachelor of Science in Medicine
MSc	Master of Science in Medicine, usually equal to MBBS/MBChB or MD
MD	Medical Doctor, the degree awarded after medical school; in Commonwealth countries MD is optional and requires additional doctoral training
MBBS	Medical Bachelor and Bachelor of Surgery, leading to a licence to practise (Commonwealth countries)
MBChB	Medical Bachelor and Surgical (chirurgia) Bachelor, leading to a licence to practise (Commonwealth countries)

remain largely the same, despite numerous local and national reform efforts [16]. Some would argue these systems are no longer 'fit-for-purpose' [6].

According to the mandate of social accountability, societal needs and priorities should drive medical education [17–19]. But this seemingly simple mandate is actually quite complex, as evidenced in a 2011 theme issue of *Medical Teacher*. Societal needs vary from the local communities served by the medical school and affiliated health care systems to national and international communities. Historically, local needs have taken priority, but in an increasingly international world we need to reconsider which of these takes priority and how an optimal balance among all three might be achieved [20]. Additionally, it takes students at least six years to complete medical training and enter practice. This lengthy process creates a lag between demand and supply that is difficult to correct midcourse. Curriculum planning faces a similar conundrum as content and processes try to keep pace with new discoveries, technologies, and epidemiological patterns. Can existing educational systems become more flexible and adaptable or do we need a major redesign that integrates multiple health professions? Furthermore, society consists of multiple stakeholders (e.g. patients, health professionals, government officials), each of whom may define societal needs and priorities differently. How are these to be reconciled?

Several national and international groups have attempted to establish a collective vision of the purpose of medical education [6, 21–25]. This vision can, in theory, provide the essential basis for accreditation standards, workforce and education policies, curriculum development, and required competencies for licensure or unsupervised practice. In practice, operationalising a global, collective vision of the purpose of medical education and enacting necessary structural and curricular reforms might be characterised as a 'wicked problem', one that lacks 'definitional clarity because multiple stakeholders in shifting social contexts have different interpretations and seek different outcomes' [26, p. 339, 27]. In the section that follows, we gain some insight into the issues that thought-leaders in medical education see as priorities in the sense that they are likely to impact the future of medical education over the next 10 years.

Glimpses of the Future

'*In the long run, we will neither need nor want professionals to work in the way that they did in the twentieth century and before*' (Susskind and Susskind) [28, p. 1].

To prepare a chapter discussing the future of medical education, we (the authors of this chapter) sought assistance from experts around the globe. In June 2017 we asked authors of chapters in this book, its editorial board members, and a group of thought-leaders from diverse geographic, disciplinary, and institutional perspectives to '*Identify at least 3 factors you think will impact the future of medical education in the next 10 years and describe why each of these factors will be so influential*'.

We contacted 91 individuals and 51 shared their ideas, from 18 countries across 6 continents (see Box 2.3). Respondents identified more than 150 factors likely to impact the future of medical education. We clustered these factors into five overarching themes. On balance, the responses might be characterised as 'cautiously optimistic', though as one respondent astutely noted, 'the answer depends on whether one takes an optimistic or pessimistic view of the future'.

Admittedly, predictions of social phenomena are often erroneous [29], but clairvoyance was not the goal of this endeavour. Rather, the point was a global snapshot to capture the current focus of attention as, perhaps, an important way of contextualising the content in the chapters that follow. We suggest readers consider the themes as commentary on the current state of affairs in health professions education and an opportunity for reflection as well as anticipation.

In writing up the themes from the responses, we attempted to capture the many thoughtful and insightful responses we received. That said, we acknowledge that the resulting picture cannot fully capture an accurate representation of the surveyed population. We integrated some of our own perspectives with those of respondents to the survey (noted in quotes) and referenced literature where we identified relevant connections. We realise that re-reading the chapter five years after its appearance may strike us with embarrassment [30], but if this chapter inspires readers to ponder a *possible* future, and guides readers in current educational and curricular decisions, then its purpose is fulfilled.

BOX 2.3 Location and number of respondents

Continent	Country (Number of respondents)
Africa	Ethiopia (1), Tanzania (1), South Africa (2)
Asia	Japan (1), Taiwan (1), Singapore (2)
Australia/New Zealand	Australia (4), New Zealand (1)
Europe	Denmark (1), Germany (1), The Netherlands (2), Spain (1), Sweden (1), United Kingdom (12)
North America	Canada (9), United States (7)
South America	Argentina (1), Venezuela (1)

We grouped the responses into five primary themes, each with several sub-themes:

a) Educational system factors that highlight developments in curricula across the continuum such as competency-based time-variable programmes, simulation, faculty development, and market aspects such as finances and selection procedures.

b) Globalisation, including attention to migration, sharing of educational tools and concepts, increasing international collaborations, and development of international standards.

c) Health care system factors, including greater attention to preventative medicine, the need for team-based approaches to care, and workforce shortages.

d) Cultural and societal factors, including further elaboration and clarification of core principles of professionalism, changing values and expectations among and toward patients, and changing values and expectations among and toward learners.

e) Technological factors, including technology-supported clinical reasoning, changing relationships with patients, information access and the role of knowledge acquisition, and methods of instruction in medical education.

Educational System Factors

An old saying, attributed to Harvard's past president Derek Bok, is that it is more difficult to change an undergraduate curriculum than to move a graveyard. Medical curricula, however, have changed over time, and national initiatives, such as the Flexner investigations a century ago [31], have significantly contributed. Yet, medical curricula do not change easily, given the considerable numbers of students, faculty members, departments, and external regulations and requirements [32]. Few higher education programmes train professionals with such clear societal and internationally agreed status as medical schools and residencies, despite the international disparities highlighted in the beginning of this chapter. Changes are therefore limited within the boundaries of societal expectations of what doctors and medical specialists are and should be. Yet, the second half of the twentieth century has shown significant innovations, well summarised in 1984 in the SPICES acronym [33]: Student-centred approaches, Problem-based methods, Community-based content, Electives, and Systematic clinical education; many of which still reflect the changes medical curricula undergo at the present day. However, the twenty-first century started showing renewed calls for reform [6, 34] and competency-based medical education dominated the renewal of postgraduate education, despite debates in the medical education community [35, 36]. Competency-based models have drawn attention to communication, collaboration, professionalism, advocacy, scholarship, and leadership as important attributes of doctors and objectives for training, but the definition of what a medical doctor is or should be has not become clearer [37]. Nevertheless, the rate of change in medical school curricula and postgraduate training programmes seems to have increased and few would now compare changing these programmes to moving a graveyard. Rather, medical school curricula in Western countries now seem to have a half-life of a decade. The desire for change is strong, as reflected in one survey respondent's remarks, 'If we had to design the education system from scratch we would never have designed it with the system we currently have inherited'. In fact, change and adaptability of educational programmes may become the constant in medical education, rather than an exception or rarity.

Respondents to our survey massively addressed educational systems changes that they foresee in the coming decade. We categorised these into five sub-themes.

Competency-based, Time-variable, Individualised Pathways across the Continuum

Fixed standards and flexible pathways, a recommendation from the 2010 Carnegie Report [24] aligns with the promises of competency-based education, in which not time but acquired competence should determine the licence to practise in health care [38]. Several respondents considered transparency and accountability key to competency-based education. While time-variable training poses substantial logistical challenges [39], several respondents predicted that future medical education models will focus more heavily on outcomes and will apply milestones and entrustable professional activities (EPAs) in a time-variable fashion [40]. A focus on EPAs, as units of professional practice – the things medical practitioners must be trusted to do – may ease the way to more individualised trajectories for learners. An individual, dynamic portfolio of EPAs, rather than a static general diploma, may define learners' licence to practise. Core EPAs may constitute a traditional specialty, while EPAs that are not practised may drop from the list of privileges and other, elective, EPAs may be added during or even after a formal training period. This approach would constitute true competency-based medical practice, but may be highly visionary. Narrowing the core and widening the elective components of curricula, as one respondent suggested, would lead to more individualised, contextualised, and diversified education that could be highly tailored to local needs. Another respondent envisioned 'a continuum of education with no longer separate stages of undergraduate, postgraduate and continuing education working in silos'.

Simulation

Simulation in medical education, first proposed by Barrows in 1964 [41], has slowly but very steadily matured and has now reached a level of sophistication that allows for standardising not only patients but technical procedures and even real-life clinical scenarios [42–44]. Several respondents suggested this would increase the quality of preparation for workplace-based training and improve patient safety. Meanwhile, others stressed the importance of reviving bedside teaching, not so much to deny the usefulness of simulation and the need for quality in sophisticated diagnostic procedures, but to focus on the core of patient-centred education [45–47] and restore the *raison d'être* of the physician [48].

Faculty Development and Education Careers

One respondent expressed worries that essential basic science education is at risk, as anatomy, physiology, and pharmacology do not offer attractive career prospects, which may subsequently threaten the teaching in these domains. Several respondents emphasised that academic careers for faculty members must include education as a core pathway for promotion if high-quality education is to be sustained. Translational scholarship should not only apply to the hard sciences, but also to educational science, another respondent suggested.

Funding of Education and Selection of Students and Residents

In countries with market-driven health care systems, the funding of medical education has become so problematic that changes seem inevitable. Exorbitant tuition fees, exorbitant debts after training, and exorbitant physician salaries seem to hold each other captive, at the cost of meritocratic admission of students and diversity in the health care workforce. Some respondents felt current systems of admission – for those who can afford medical school – with a strong focus on knowledge and academic achievement were inadequate, 'If we want reflective, considerate doctors who are good at team working, etc. then maybe we need to turn selection completely on its head, and select for personal attributes as the first hurdle'. Likewise, Aagaard and Abaza suggest that for residency in the US, the matching and selection process has become a source of frustration as it now consumes the energy of most of the final curricular year, and needs to change [49].

Curricular Content

Predicting which basic and clinical science topics and procedural skills will be most relevant to clinical practice of the future seems futile given rapid advances in knowledge, shifting epidemiological trends, and easy access to information. Instead, several respondents suggested curricula need to devote more attention to reflection, humanism, self-regulated and adaptive learning, communication, team working (especially across professions), ethical decision-making, effective and efficient use of technology, and leadership. Despite significant advances in the sciences (genetics, genomics, pharmacology, stem-cell therapies, personalised cancer care, and others), remarkably few respondents stressed these as impacting the future of medical education. There were several comments recognising the need for governance structures and 'mechanisms for enabling change of the curriculum when needed'. One other respondent noted 'curricula will need to be quickly responsive to global changes and not be expected to last 10 years!', which advocates for adaptive expertise not only in learners but also in educators and curriculum developers.

Globalisation

There is no doubt that globalisation will affect the future of health professions education. Our respondents discerned a number of specific topics that warrant special attention.

Migration Effects

Socio-economic differences between countries, population density differences, warfare, climate change, differences in workforce demands, and the ease of travel all stimulate migration. The significance for health professions education is that all doctors must be prepared to face the increasingly international population mix, both as patients and as colleagues. Learners must also be prepared to work with a growing number of foreign peers in medical schools and residency or postgraduate programmes. Humans have a natural tendency to develop and create a better life for themselves. This is no different for medical students and physicians than any other human being. In countries of lower socio-economic conditions, this may have a devastating effect, because the scarce resources used to educate doctors to serve the local health care needs often lead to graduates seeking a better life in a more developed country [50]. One of our respondents noted, 'in the Least Developed Countries like Tanzania in sub-Saharan Africa, one doctor serves 20 000 to 30 000 persons in the population, compared to one doctor for a few hundred persons in the developed world'. In the decade ahead, we must face the challenge to retain doctors where service is most needed. The FAIMER Institute, created to stimulate the improvement of local education and stop this brain drain, offers one example of an initiative [5].

Sharing Education Concepts Globally

One respondent eloquently wrote, 'No individual, community, nation or even continent can boast of good health when its neighbor's is wanting. That is truer now than ever in a world that has become a "global village". There is thus a need for the world to adopt a more global outlook towards medical education in the interest of health worldwide'.

Since the beginning of this century, stimulated by a rapid globalisation of information through the Internet, health professions education has become more globally oriented. Health professions educational ideas are shared in a large number of journals, increasingly open access, and as the number of international conferences increases, so does the number of participants and the opportunities to expand thoughts, approaches, and techniques beyond local or national borders. Concepts developed in one country are quickly adopted in other countries; examples include problem-based learning, curricular integration, competency-based education, simulation, and the objective structured clinical examination (OSCE) [33, 51–53]. Some respondents reminded us, however, of the costs involved in such adoptions for low-resourced countries.

Teaching methods constitute another area of exchange. Portable educational techniques using technology may help to support the training in countries with few faculty or patients. 'The training of large numbers of students stands face to face with the (relative) shortage of patients in poorly equipped hospitals. The deployment of clinical skills laboratories is one way of coping with this challenge. While this has budgetary implications, it is a very necessary area for development if acceptable standards of training are to be achieved and maintained.' Another respondent explained, 'We tend to think of health professional education as more of the same but better [suggesting identical development across countries]; however, this education is also taking place in less stable areas than our own and curricula need to adapt to local circumstances as well as global ones'.

International Collaborations Between Institutions

International courses and collaborations to develop medical education and its research are quickly expanding [54, 55]. 'I was impressed, for example, by students in Dundee learning about the cardiovascular system using a programme to which 14 medical schools had contributed. Some students were facilitated online by a cardiologist from Florida rather than a local cardiologist in Dundee.'

One example of a truly global enterprise to shape the future of medical education is the Initiative by Dr Hilliard Jason to establish a trust foundation to support medical education in developing countries through an 'adaptive medical education' model, guided from a school in London, UK, to be built in the coming years, serving educators and institutions worldwide. The adaptive education model aims to serve individual learner needs [56].

Towards International Standards with Local Applications

Objectives of education can lead to worldwide standards to globalise education, but some respondents voiced a nuanced view. 'We are now at the globalisation side, but there are some voices (including mine) beginning to speak in favour of going back to the local priority.' Why? 'There are the added difficulties of communication, professionalism and ethics – all of which differ fundamentally across different cultural groups.' '[There is] clearly a different view on what constitutes good medical education in the Asian countries and competitive Western countries.' This reflects a debate that is not new [57, 58] but that will become even more relevant with the upcoming economies that have cultures quite different from Western societies.

Health Care System Factors

Many of our respondents highlighted factors within the health care system that will impact its sustainability. These factors included rising health care costs, increasingly specialised/technical and siloed approaches to care, and the ageing population with multiple morbidities. The challenge moving forward is to find ways in which medical education can play a role in helping to mitigate these factors.

Solutions discussed, which are not new, included developing better interprofessional working relationships, increasing community-based care, and improving workforce planning. These can be categorised broadly as changing the context of medical education (from disease-based to preventative health care education), matching the context of training to that of care (away from hospital tertiary institutions toward community-based care models), and changing the ways physicians work with one another and across professions (workforce development, new roles in teams, and better interdisciplinary processes of care).

The blessing and curse of clinical education is that much of the learning occurs in practice, through delivery of care in existing systems. Correspondingly, physicians are well trained to 'provide medicine now and not in 10 years'. The challenge of preparing learners for an uncertain and rapidly changing practice environment is well recognised, but how best to do this remains unclear.

From Disease-based Education to Preventative Health Care Education

The growing proportion of the ageing population with co-morbidities has resulted in higher health care costs due to the demand for increasingly technical and complex care. Given this context, respondents were of the view that available health care resources (money/people/infrastructure) cannot keep pace. Accordingly, '… preventative medicine will be a cost-driven necessity. Doctors will be pressed to reduce the costs of infectious diseases, cardiovascular disease, smoking, obesity, drugs and mental disorders …' and 'Public health and Primary Care should be the curriculum drivers'.

At present, medical education focuses mainly on disease models, with limited attention given to public health and behavioural/psychosocial effects on health. This lack is exacerbated when graduates progress to postgraduate training in hospital-based systems. Some argue that shifting health care toward a preventive model will not save money unless over-testing and over-treatment are also addressed. Indeed, some medical school and postgraduate curricula are beginning to attend to these topics [59]. Whichever viewpoint one takes, achieving a cost-effective system that delivers quality care will require significant changes on multiple fronts, but, as one respondent noted, maintaining the status quo is not an option: 'Any future vision in which medical education does not embrace disease prevention and health promotion is doomed to be unsustainable and to produce doctors who are not fit for practice.'

From Hospital to Community-based Education

There is now quite wide consensus that health care needs to be organised into networks of care, with services integrated around the patient and based in the community as much as possible. This model of care requires physicians and other health professionals to partner with patients over the long term rather than providing single, unconnected 'episodes' of care [60]. Medical education will need to prepare physicians to coordinate and collaborate across these networks. As one respondent explained: 'We need to rethink our curriculum and offer greater balance in terms of the contexts in which training happens (hospital based training still dominates) … too often we train in disciplinary silos expecting our graduates to work in ways that are interdisciplinary.'

Longitudinal clerkships provide learners with continuity of patients and educators, and the opportunity to become identified members of the team [61]. Evidence suggests that learners who partake in these clerkships gain a better understanding of the impact of health and illness on the patient and the communities and develop a compassionate and caring approach.

As care shifts away from the hospital to outpatient and community-based settings, patients will need to be supported in self-care and self-management. One respondent speculated on the need to redefine the role of the doctor and how '…we educate our students to manage these issues in the face of change…'.

Team-based Care

There was a large consensus in responses that the increasingly technical and complex health care needs of the population require a team approach to care. Growing awareness that breakdowns in communication and poor teamwork are major contributors to many medical errors further bolsters the calls for effective team working and therefore interprofessional education (IPE) [62, 63].

For interprofessional team-based care to be effective there is a need for more training and experience working in IP teams; however, 'it seems we are still looking for effective ways in which professions can "learn from, about, and with each other"'. While many curricula include IP education, evidence of lasting impact on behaviour and communication among professionals in practice is sparse [64].

An aspect of teamworking in education that needs further exploration is that of the 'collective' competence. Above, we described patient care as a network; this network is of individuals working together in teams, in a complex system. However, we still generally educate and assess on individual competencies and not on how the individual affects the team. A body of work is building around 'team' performance and its translation into the undergraduate and postgraduate education of health care professionals is yet to be addressed [65].

Although practitioners already work in teams, most are not IP teams. Although such 'tribes' are good for moral and professional support they can have deleterious effects on patient care when 'tribes' defend their 'patch' [66, 67]. One physician respondent wondered, admittedly cynically, if 'the medical profession's defence of its own turf will become harder to sustain in the face of other professions rightly insisting they have as much actual or potential expertise as doctors, as the imperative for increasing teamworking will make doctors aware how many advanced skills their non-medical colleagues have, and make it apparent to other professions that there may be nothing special about the ones that physicians have.' Another respondent explained: 'As the technical opportunities for up-to-date medical care for patients will grow in numbers and complexity, there will be a more diversified team of professionals in health care.' These new 'team members' will add extra imperative to be able to work together effectively for patient care.

The Changing Workforce

'Without radical change to health care education we will not be preparing students for the future but delivering them to the past. Are we producing a flexible enough workforce prepared for the challenges of the future?'

The global economy is projected to create around 40 million new health sector jobs by 2030, mostly in middle- and high-income countries. Despite this growth, there will be a projected shortage of 18 million health workers needed to achieve the UN Sustainable Development Goals in low- and lower-middle-income countries [68]. As mentioned in the globalisation section of this chapter, the crisis in recruitment of health professionals means that countries all around the world will be 'fishing from the same [small] workforce pool'.

Many countries, particularly in the developing world, are looking at addressing projected workforce shortages by introducing new roles (such as physician's assistants – PA) or by role substitution (such as advanced clinical practitioners). The fundamental premise is that the training time and cost is less than for a doctor, and that graduate salaries are lower. This development has caused much angst, with some physicians calling PAs a 'poor man's' doctor and suggesting that these roles were ill thought out 'quick fixes'. However, the role of PAs (and others) seems likely to stay, probably with increasing scope of practice over the coming years [69, 70].

There was consensus among our respondents that the doctor of the future should be flexible and able to work in interprofessional teams to provide quality health care. Also, that the education of future doctors has to change, although the direction of that change is not as clear. Some declared that doctors should become broader based in their approach and community focused, others suggested that doctors should be even more sub-specialised, as other professionals could fulfil the community roles. The following two quotes illustrate these dichotomous views:

> 'We need more generalists rather than specialists, and health professionals who can work collaboratively rather than hierarchically.'

> 'Much of routine medicine can be carried out by health care professionals like Physician Associates, Specialist Nurses, Midwives, etc. Maybe medical education should prepare students for more specialised and complex medicine.'

Currently, a primary medical education qualification 'produces' an intern or resident who is a generalist, but with the least experience in the health care team. This intern, over many years of postgraduate training, focuses their skills to a specialism and often a sub-specialism. This process inherently puts forward the value proposition that being a 'generalist' is not worth as much as being a 'specialist', especially given that physicians are rewarded financially the more specialised they become. As a radical alternative, future medical school education may be envisioned to switch to producing narrow specialists with a competency-based model and postgraduate training could then prepare generalists. This would require a major paradigm shift in medical education.

Respondents also highlighted the need for alignment between training choices and workforce needs, which could result in more 'engineering' of career pathways [71]. This approach would reduce trainees' freedom to choose their specialty and location. One respondent suggested medical education leadership had to show the way: 'Over the next 10 years, medical education has to take a leadership role in producing graduates who take an ethical and responsible approach to health resource stewardship.'

Cultural and Societal Factors

Culture and society are part of the ubiquitous, multifaceted context in which medical education is situated [72, 73]. Both health care and medical education interact with broader social, cultural, political, legal, and economic forces – sometimes accommodating these forces, other times reacting against them. The social contract metaphor is often used to characterise the complex relationship between physicians (or the medical profession as a whole) and society [74]. The basis for this social contract rests in the power of physicians, as members of the medical profession, to self-regulate (to set and maintain standards for education and practice) in exchange for the provision of medical care that serves the needs of patients and society [73, 75]. Yet, increasing heterogeneity within the medical profession and throughout society has prompted important questions about the terms of this contract. In 2002 the American Board of Internal Medicine, the American College of Physicians Foundation, and the European Federation of Internal Medicine published the 'Physician Charter' in an effort to make the principles and professional responsibilities of physicians explicit. Altruism, honesty, respect, and trust were associated with three principles of professionalism, namely 'primacy of patient welfare, patient autonomy, and social justice'. These principles are presumed to provide ethical guidance to physicians in times when patients, organisations, governments, and markets place new, often competing, demands on physicians [74]. These new and competing demands may challenge the personal values held by individual or sub-groups of physicians, thus raising questions about the medical profession as a collective entity attempting to uphold a contract in a dynamic context with changing expectations [76–78].

Responses to our survey revealed many thoughts about how changes in cultural values and expectations among patients and health care systems might impact the future of medical practice and, correspondingly, medical education. Responses also highlighted changing values and expectation among learners in medicine [79], from those entering the profession to those engaging in continuing professional development and lifelong learning. Many suggested that physicians and educators will need to respond to and accommodate these changes.

Core Principles of Professionalism

As noted above, one of the core principles in the Physician Charter is 'the primacy of patient welfare' [74]. In line with this principle, one respondent wrote: 'The best doctors have an underpinning altruism. This finds expression in their attitudes towards their patients, the community, and the medical profession.' The respondent then identified several aspects of current medical training that may undermine altruism and emphasised the need for training experiences that reinforce altruism, such as working in teams over time and developing longitudinal clinical relationships. Other respondents anticipated a change in the social contract, with growing demand for 'more balanced lives' and 'less sense of self-sacrifice' among physicians which may require new ways of operationalising altruism [80].

Respondents also mentioned patients' need to trust physicians to provide safe, competent care, anticipating shifts toward greater social accountability and external regulation of competence and further reduction of professional self-regulation. Growing attention to physicians' roles in teams and systems and notions of competence as a group- or system-level construct as well as an individual-level construct appeared in few responses. One respondent noted: 'Future patients are going to ask for person-centered care, and must be able to trust that the team around the patient can deliver that – with safety.' Physicians of the future were expected to be more involved in systems improvement and to have a stronger orientation toward public health and holistic interventions.

Changing Values and Expectations Among Patients

Several respondents described patients as 'empowered', 'engaged', and increasingly 'involved in self-care' – largely made possible through technology and increasing access to information. They anticipated growing demand for not only person-centred care, but personalised medicine and immediate access. Respondents also noted that patients will interact with physicians and health care teams in new ways, perhaps with less personal contact and 'less satisfying relationships'. Others suggested that these changing relationships might require more empathic physicians and development of 'novel communication skills that both embrace this evolution of the doctor-patient relationship while still promoting a long-term relationship'.

Changing Values and Expectations Among Learners

Perhaps signifying overarching cultural changes, several respondents described parallels between patient and learner expectations. 'We are rapidly evolving from a provider-centered health care delivery and teacher-centered model of educational delivery to patient-centered and learner-centered models.' Another respondent suggested: 'We have personalised medicine – we should have personalised education.' Ideas about how this might emerge included online, mobile, self-paced learning modalities replacing in-person, campus-based instruction; growing use of simulation; new sources of motivation for learning (driven by perceived needs and interests rather than proscribed curriculum); increasing use of international collaboration; and new forms of assessment.

Respondents also described the 'democratisation' of education as relationships between teachers and learners become less formal and less hierarchical compared to just a few decades ago. As discussed in the next section, ever-increasing access to information and rapidly changing practices are likely to contribute to this trend as less and less knowledge becomes exclusive property of expert clinical teachers. Instead, teachers may be increasingly learning alongside, or just barely ahead of, the learners they are teaching.

Addressing Strenuous Work Environments and Enhancing Student Resilience

Learners' desire for 'work-life balance and supportive, well-functioning working and learning environments' was mentioned by several respondents. One respondent emphasised 'we need to pay a lot more attention to the nature of the working and learning environment if we do not want to lose a generation'. Given the alarming figures on the prevalence of burnout and depression among medical students, residents, and physicians [81, 82], education, coupled with systems changes, can address the ways learners are prepared for a difficult working environment. One respondent suggested that physicians will need 'coping and resilience skills to thrive in this new practice'.

Comments about learners' desire for a work-life balance echo conversations about professionalism in the context of duty hours regulations for physicians. Many feared the degradation of altruism and prioritisation of patient welfare with the rise of a 'shift work' mindset while others saw an opportunity to redesign health care systems that honour the need for self-care without compromising the primacy of patient welfare and overly relying on individual altruism [83]. A fact is that resident restrictions in duty hours in Europe differ vastly from those in North America, while there are no reports of differences in professionalism. There is no doubt, however, that these topics of physicians' psychological distress, burnout, and well-being [84–87] will continue to be important for many years to come in countries across the world.

Technological Factors

Throughout history, people have pointed to technology as a primary source of change in society. The rise and rapid development of computers and robots has prompted many to think about the implications for the way people will work in the future and the educational requirements for such work [88]. Medicine and health professions education are no exception [28, 89, 90], with articles on the use of computers as consultants appearing in the late 1960s and early 1970s [91]. This exciting prospect inspired decades of research by cognitive scientists, physicians, and computer scientists that aimed to understand clinical reasoning and expert decision-making processes so they could be replicated in computer programs [92]. While some might argue that these efforts have shown limited success based on practical application and use, recent improvements in data processing capacity, coupled with exponential increases in the volume of data available through the digitisation of health care systems and records, are rapidly changing the rate of implementation and uptake in clinical practice. As these changes occur, medical education must keep pace and incorporate technology into competency domains, learning objectives, and pedagogical techniques.

The core of discussions of technology, both in the literature and among our respondents, seems to be about information and how technology is used to collect, analyse, synthesise, and ultimately transform information into an 'intelligent' judgement, action, or solution of value to people. Terms such as 'artificial intelligence' and 'machine learning', all made possible by access to 'big data', came up repeatedly. Artificial intelligence generally refers to machines capable of performing complex cognitive activities at or beyond the level of a knowledgeable and skilled (i.e. 'smart') human. Machine learning denotes the ability of these machines to access and process data in ways that allow them to improve their performance – essentially to learn or get smarter [93]. Big data refers to the huge volume of information available digitally through databases, photos, videos, audio recordings, text, and biometrics. Machine learning employs analytics technology to search for patterns among all these information sources and, ideally, provide insights and predictions that prove valuable, or intelligent [93]. Technology also relies on information to support the development of tools that automate functional or physical tasks (e.g. robots that dispense medications) and that supports human access to information used to make decisions (e.g. smartphones and apps).

We identified four general themes among the responses citing technology as a key factor impacting the future of health professions education. Two themes focus on changes in the practice of medicine, with implications for the content of clinical training and two focus directly on changes in education, with implications for educational processes. Several respondents also emphasised significant disparities in the impact of technology. They predicted further exacerbation of disparities between patients, populations, communities, and countries with limited access to technological and other resources and those that are well resourced.

Changing Clinical Reasoning

Many respondents described changes in how physicians make diagnostic and management decisions. They wrote about the growing capacity for artificial intelligence (AI), or machine learning coupled with big data, to identify patterns and algorithms. While this would not necessarily eliminate the need for physicians to engage in clinical reasoning, such technologies were expected to provide decision support systems to aid and enhance physicians' decision-making. One respondent wrote: 'Physicians will need to rely less on their own memory to recall information and use expert systems to avoid treatment biases.' Some expected that AI would, in fact, eventually 'provide an alternative to diagnosis and treatment' which would satisfy the growing demand and expectation for instant advice and treatment. Others expected automation of routine tasks and replacement of procedural skills with robotic instruments. With these technologies in place, some respondents predicted that physicians' role would become that of an interpreter or translator of patients' descriptions and critical appraiser of information and evidence. One respondent described this as 'the rise of the coach, the routinisation of the technician'. This change suggests a shift in required competencies, with less emphasis on 'technical competencies' reliant on medical knowledge for diagnostic and treatment purposes and more on innovation and 'humane competencies' such as patient education and advocacy. Indeed, some have already described these trends in fields such as radiology and pathology [94]. While the capacity for expert systems and AI to outperform human experts was established in the 1970s (e.g. MYCIN) [95], such systems never became part of routine medical practice because they required manual data entry, took considerable time to produce solutions, and generated fear and mistrust among users concerned about being outperformed by a machine [96]. Today, these systems can be integrated into existing health information systems and information can be processed almost instantaneously. While concerns about technological displacement still exist, such systems have been recast as 'decision support systems' and guidelines that enhance performance of human experts rather than competing with it [97].

One respondent offered a prediction, summarising many of these ideas: 'Without a doubt, the implementation of artificial intelligence systems as part of the practice of medicine will have a major impact on what it means "to be a doctor" and what competencies we will expect of future physicians …The greater use of technologies will free physicians to focus on treating patients and innovating with more automated tasks being relegated to engines that accomplish the latter with a much higher level of precision.'

Changing Relationships with Patients

Respondents highlighted technology's role in providing new ways for patients and physicians to communicate, including telehealth, email, and social media. Some predicted face-to-face visits would become obsolete or at least significantly less common. Correspondingly, they saw a need for medical education to prepare physicians to interact with patients through various new technologies and to use technology as a tool to engage, monitor, treat, and educate patients.

Respondents also noted technology's role in providing patients access to information that previously belonged almost exclusively to physicians and was more or less inaccessible to non-physicians. This includes both information about one's own health (from digital records as well as from genomic testing) and information about diseases, diagnostics, and treatment available (through websites of variable repute). While some respondents saw this as a beneficial way of empowering patients and promoting preventative measures that might reduce reliance on physicians, others anticipated an even more significant role for physicians to 'interface between patient concerns and technology' as interpreters of patients' concerns and evaluators of the quality of evidence or soundness of reasoning guiding patients' decisions.

Changing the Focus and Content of Medical Education

Responses suggested two primary ways in which technology will impact the core curriculum in medical education. First, given rapidly changing and growing knowledge bases that learners can instantly access in real time, respondents anticipated significantly less need to teach for knowledge retention and much greater need to teach for finding, critically appraising, synthesising, and integrating information. This view raises intriguing questions about the relationship between knowledge and reasoning, many of which are not new [98, 99] but perhaps are relevant in new ways as we consider what types of knowledge and conceptual understanding physicians will need to evaluate the quality of decisions recommended by a 'smart machine' and/or of the information returned from a query. Yet, these views assume ubiquitous access to credible information with no barriers. For resource-limited countries, this will require open access to reputable journals for hospitals and educational institutions.

These shifts have implications for assessment, as one respondent noted: 'Every medical educator knows that it is impossible for practising clinicians to hold in their heads all of the knowledge that must be incorporated into their clinical reasoning and decision making. And yet we continue to assess our students on what factoids they can recall, rather than what they can find out. Assessments over the next 10 years will need to become much smarter not only at measuring what students understand about concepts, but how they go about finding the information they need to apply that understanding.'

Second, the analytic processes used to support and enhance AI in health care systems can feed information to educators and educational systems to align curricula better with practice. As one example of how this might work, Baker and colleagues described opportunities to use data from the US National Center for Health Statistics to guide the design and contextualisation of cases in a case-based learning curriculum. The authors described ways of using summary data about the most common diagnoses, tests, procedures, and medications associated with patients presenting to the emergency department with fever in a given geographic region [100]. Access to such information also raises questions about how to design current educational systems to support future practice that might look quite different. Should the focus be on common diagnoses or on challenging or uncommon diagnoses that are often missed?

Changing the Methods of Instruction and Assessment

The response, 'new technologies are changing not only our lifestyle but also our relation with students and the learning/teaching paradigm' captures the essence of themes about technology's impact on the methods of instruction and assessment in medical education. Several respondents anticipate changes in teachers' roles as technology increasingly mediates their interactions with learners. The current rise in online courses, many of which are freely available worldwide (provided you speak the language and have sufficient bandwidth and equipment), creates what one respondent called 'borderless education'. Teachers are also interacting with learners in virtual 'classrooms' or learning spaces, often asynchronously. In clinical settings, respondents expected the role of the teacher to change too, with teachers no longer functioning as the primary source of medical knowledge, but instead communicating and modelling expertise in clinical skills, attitudes, values, and professionalism. With growing technological resources available to enhance teaching and learning, educators will need to be more flexible and 'able to respond with innovative ways of matching learners' information technology skills with teaching methods' and 'channeling in a positive way' learners' willingness to challenge the status quo. Educators and learners alike will need to vigilantly monitor their use of technology to ensure that it enhances, rather than detracts from, learning [101–103].

Multiple respondents identified ways in which the combination of 'big data' with learning analytics and AI will change assessment processes and create opportunities for truly learner-centred, individualised approaches to learning. For example, they foresee these technologies allowing us to 'analyze student behavior', 'use digital traces of teacher and learner activities to generate feedback', 'rapidly and effectively screen assessment data to identify learners at risk or excelling', and improve selection processes by 'creating profiles on potential trainees … including not just formal educational performance (hand-offs on entrustment activities of students) but also information gathered through longitudinal portfolios and the Internet in general (e.g. social media, LinkedIn, videos or podcasts)'. Information and learning technologies were often mentioned in relation to programmatic assessment and competency-based medical education, topics discussed in greater depth under educational system factors. At present, few examples of these systems exist on a large scale, but design and development are underway [104, 105]. There are, of course, costs and risks associated with managing, securing, interpreting, and sharing this information. Respondents raised concerns about data security, privacy, and ethical use of information in the digital age.

Discussion

Opportunities to envision the future appeal to our human imagination. Perhaps we are drawn by the desire to plan better, avoid surprise, prove our hypotheses, or reflect on and learn from the past. Perhaps we hope to improve the future through such musings. Whatever the case may be, the respondents to our survey each had important messages to convey. And, while each drew on different contexts and structural models of medical education to formulate their visions for the next 10 years, we found considerable overlap in their perspectives. Some responses were truly forward looking, others reflected anxiety about current circumstances and urged recognition of 'festering problems for which a solution *must* be found'. Some predictions, particularly around technology development, resemble ones that could have been made 40 years ago when AI, diagnostic support systems, and computer-based education started to catch medical educators' attention.

The training of health professionals is the preparation for services that populations need or desire to improve their health, but cannot provide for themselves. Health and disease have been mysteries for many ages and remain so to some extent. The world has changed, however, and most educators would agree that in the twenty-first century these population demands have evolved. Many of the mysteries of disease have been solved, patients obtain information through many sources other than health professionals, many health care decisions are shared among health professionals and patients. All of these suggest less dependency in populations. On the other hand, many more health conditions can be improved than in the past, life expectancies have increased in many parts of the world, and chronic diseases are more prevalent. Just extrapolating these developments may shift the need for decision makers and curers to being supporters of patients in their navigation through complex health care systems and myriads of options. Education must prepare learners for these shifting roles. Akin to the loss of rote memory skills for long texts after the invention of print (think of Homer's *Iliad* and *Odyssey*), doctors in training may be better not spending most of their time on rote memorisation of facts and detailed physical examination skills, as these may be replaced by much more accurate diagnostic procedures. This is not merely a shift to different curricular content, but a change that may affect the identity of health professionals and their relationship to society.

This review illuminates the opportunities and concerns most salient to the leaders, educators, and researchers who responded to our inquiry. Their responses depict promising opportunities for education customised to individual capability and learning needs; integration of technologies that enhance precision, efficiency, and safety of patient care; redesign of working and learning environments to be more satisfying and sustainable for clinicians, learners, and patients; and international collaboration on and sharing of educational materials and resources (e.g. curricula, instructional techniques, assessment tools, and procedures). They also portray dark clouds on the horizon with the issues such as workforce shortages, resource disparities, inequities in access to education and patient care, growing demands on educators with insufficient support and recognition for their efforts, and challenging work environments. Clearly, many of these concerns will limit the array of opportunities if not addressed. For example, creating and implementing individualised learning pathways requires educators who are trained and invested in this approach. In environments where educators already feel stretched thin, support will be critically important as will open conversations about how such efforts will benefit (and change) their role as educators rather than replace it.

So now what to do with these glimpses of the potential future of medical education? The people who provided the responses are also the leaders in the delivery and development of medical education. These responses indicate the directions of their plans. and prompt a follow up question that asks what they are doing to address these issues because if they don't, then who will? The first section of this chapter shows us how medical education has previously responded to evolution in medical education by adapting or changing into different pathways/routes. Instead of convergence of the pathways, we may see more divergent models and individualised pathways as an adaption to new factors and drivers.

Rather than ending this chapter with recommendations for the future, we prefer to end with questions that educators and leaders can keep in mind when creating or revising curricula and educational programmes, advocating for educational policies, building partnerships within and across institutions, and drafting five or ten year plans.

1 What do we need to do now to prepare learners for future careers as physicians, where the work of physicians may be quite different from what it is now?

2 What is the unique value of a doctor to the health care workforce and what are the implications for education and training?

3 How does medical education and training reconcile individual aspiration with the social purpose of schools and programmes?

4 How do we create more satisfying and sustainable work and learning environments?

5 How do we move toward a more equitable distribution of health care resources around the world, both within countries and globally?

Acknowledgements

The authors would like to acknowledge the contributions of all who shared their insights about factors likely to impact the future of medical education: William Burdick, Angel Centeno, Fremen Chou, Jen Cleland, Debra Cohen, Richard Cruess, Sylvia Cruess, Andre de Champlain, Erik Driessen, Rachel Ellaway, Kevin Eva, Martin Fischer, Shiv Gaglani, Deborah Gill, Marja Govaerts, Janet Grant, Larry Gruppen, Ronald Harden, Ken Harris, Brian Jolly, Robert K. Kamei, Manuel Kassaye Sebhatu, David Kaufmann, Athol Kent, John Launer, Peter McCrorie, Judy McKimm, Charles Mkony, Lynn Monrouxe, Clare Morris, John Norcini, Hirotaka Onishi, José Peinado, Sari Ponzer, Pablo Pulida, Bill Reid, Richard Reznick, Charlotte Ringsted, Trudie Roberts, Dujeepa Samarasekera, Lambert Schuwirth, Jonathan Sherbino, Tim Swanwick, George Thibault, Jill Thistlethwaite, Aliki Thomas, Steve Trumble, Sebastian Uijtdehaage, Tim Wilkinson, Zareen Zaidi, and Gert van Zyl.

References

1 Jonker, G., Manders, L.A., Marty, A.P. et al. (2017). Variations in assessment and certification in postgraduate anaesthesia training: a European survey. *British Journal of Anaesthesia* 119 (5): 1009–1014.

2 Weller, J.M., Sullivan, M., and Boland, J. (2017). Does variable training lead to variable care? *British Journal of Anaesthesia* 119 (5): 866–869.

3 European Parliament (2013). Directive 2013/55/EU of the European Parliament and of the Council of 20 November 2013 amending Directive 2005/36/EC on the Recognition of Professional Qualifications and Regulations. Available from: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:354:0132:0170:en:PDF (accessed 9 December 2017).

4 Karle, H. (2006). Global standards and accreditation in medical education: a view from the WFME. *Academic Medicine* 81 (12 Suppl): S43–S48.

5 Burdick, W.P., Morahan, P.S., and Norcini, J.J. (2006). Slowing the brain drain: FAIMER education programs. *Medical Teacher* 28 (7): 631–634.

6 Frenk, J., Chen, L., Bhutta, Z. et al. (2010). Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. *Lancet* 376 (9756): 1923–1958.

7 Frank, J.R., Snell, L.S., and Sherbino, J.D. ed. (2015). *CanMEDS 2015 Physician Competency Framework*. Ottawa, Ontario, Canada: Royal College of Physicians and Surgeons of Canada.

8 Boulet, J.R., Norcini, J.J., Whelan, G.P. et al. (2006). The international medical graduate pipeline; recent trends in certification and residency training. *Health Affairs* 25: 469–477.

9 Hallock, J.A., McKinley, D.W., and Boulet, J.R. (2007). Migration of doctors for undergraduate medical education. *Medical Teacher* 29: 98–105.

10 Teichler, U. (2003). Mutual recognition and credit transfer in Europe: experiences and problems. *Journal of Studies in International Education* 7: 312.

11 Wijnen-Meijer, M., Burdick, W., Alofs, L. et al. (2013). Stages and transitions in medical education around the world: clarifying structures and terminology. *Medical Teacher* 35 (4): 301–307.

12 Bologna Working Group on Qualifications Frameworks (2005). Available from: http://ecahe.eu/w/images/7/76/A_Framework_for_Qualifications_for_the_European_Higher_Education_Area.pdf). *A Framework for Qualifications of the European Higher Education Area [Internet]*. Copenhagen, Denmark: Bologna Working Group on Qualifications Frameworks (accessed 9 December 2017).

13 Patrício, M., den Engelsen, C., Tseng, D., and Ten Cate, O. (2008). Implementation of the Bologna two-cycle system in medical education: where do we stand in 2007? – results of an AMEE-MEDINE survey. *Medical Teacher* 30(6):597–605.

14 Patrício, M., de Burbure, C., Costa, M.J. et al. (2012). Bologna in Medicine Anno 2012: experiences of European medical schools that implemented a Bologna two-cycle curriculum – an AMEE-MEDINE2 survey. *Medical Teacher* 34 (10): 821–832.

15 World Federation for Medical Education (WFME). http://wfme.org/standards/bme (accessed 9 December 2017).

16 Whitehead, C.R., Hodges, B.D., and Austin, Z. (2013). Captive on a carousel: discourses of 'new' in medical education 1910–2010. *Advances in Health Sciences Education* 18: 755–768.

17 Green-Thompson, L.P., McInerney, P., and Woollard, B. (2017). The social accountability of doctors: a relationship based framework for understanding emergent community concepts of caring. *BMC Health Services Research* 17: 269.

18 Lindgren, S. and Karle, H. (2011). Social accountability of medical education: aspects on global accreditation. *Medical Teacher* 33: 667–672.

19 World Health Organization (WHO) (1995). *Defining and Measuring the Social Accountability of Medical Schools*. Geneva: World Health Organization.

20 McKimm, J. and McLean, M. (2011). Developing a global health practitioner: time to act? *Medical Teacher* 33 (8): 626–631.

21 Global Consensus for Social Accountability of Medical Schools 2011. Available at www.healthsocialaccountability.org (accessed 3 December 2017).
22 Larkins, S.L., Preston, R., Matte, M.C., Lindemann, I.C., Samson, R., Tandinco, F.D., Buso, D., Ross, S.J., Pálsdóttir, B., Neusy, A.J. on behalf of the Training for Health Equity Network (THEnet). Measuring social accountability in health professional education: development and international pilot testing of an evaluation framework. *Medical Teacher* 2013;35(1): 32–45.
23 THEnet. Training for Health Equity Network. https://thenetcommunity.org/our-work (accessed 9 December 2017).
24 Cooke, M., Irby, D.M., and O'Brien, B.C. (2010). *Educating Physicians: A Call for Reform of Medical School and Residency*. San Francisco: Jossey-Bass.
25 Association of Faculties of Medicine in Canada (AFMC) (2010). The Future of Medical Education in Canada (FMEC): A Collective Vision for MD Education, Final project report. https://afmc.ca/future-of-medical-education-in-canada/medical-doctor-project/index.php (accessed 9 December 2017).
26 Hawick, L., Cleland, J., and Kitto, S. (2017). Getting off the carousel: exploring the wicked problem of curriculum reform. *Perspectives on Medical Education* 6: 337–343.
27 Carter, L. (2011). Gathering threads in the insensible global world: the wicked problem of globalization and science education. *Cultural Studies of Science Education* 6: 1–12.
28 Susskind, R. and Susskind, D. (2015). *The Future of the Professions: How Technology Will Transform the Work of Human Experts*. Oxford, UK: Oxford University Press.
29 Tetlock, P.E. and Gardner, D. (2015). *Superforecasting: The Art and Science of Prediction*. New York: Crown Publishers.
30 Eva, K. (2008). Covering up the crystal ball. *Medical Education* 42: 330–332.
31 Flexner, A. (1910). Medical Education in the United States and Canada. A Report to the Carnegie Foundation for the Advancement of Teaching.
32 Gale, R. and Grant, J. (1997). AMEE medical education guide no. 10: managing change in a medical context: guidelines for action. *Medical Teacher* 19 (4): 239–249.
33 Harden, R.M., Sowden, S., and Dunn, W. (1984). Educational strategies in curriculum development: the SPICES model. *Medical Education* 18: 284–297.
34 Irby, D.M., Cooke, M., and O'Brien, B.C. (2010). Calls for reform of medical education by The Carnegie Foundation for the Advancement of Teaching: 1910 and 2010. *Acad. Med.* 85: 220–227.
35 ten Cate, O. (2017). Competency-based medical education and its competency frameworks. In: *Competence-Based Vocational and Professional Education: Bridging the Worlds of Work and Education* (ed. M. Mulder), 903–929. Dordrecht, the Netherlands: Springer International Publishing Switzerland.
36 ten Cate, O., Hart, D., Ankel, F. et al. (2016). Entrustment decision making in clinical training. *Academic Medicine* 91 (2): 191–198.
37 ten Cate, O. (2014). What is a 21st-century doctor? Rethinking the significance of the medical degree. *Academic Medicine* 89 (7): 966–969.
38 Carraccio, C., Englander, R., Van Melle, E., ten Cate, O., Lockyer, J., Chan, M.-K. et al. Advancing competency-based medical education: a charter for clinician-educators. *Academic Medicine* 2016;91(5):645–9.
39 Kogan, J.R., Whelan, A.J., Gruppen, L.D. et al. (2018). What regulatory requirements and existing infrastructures must change if competency based time-variable training is introduced in the continuum of medical education in the United States? *Academic Medicine* 93 (3S): S27–S31.
40 Carraccio, C., Englander, R., Gilhooly, J. et al. (2017). Building a framework of entrustable professional activities, supported by competencies and milestones, to bridge the educational continuum. *Academic Medicine* 92 (3): 324–330.
41 Barrows, H. and Abrahamson, S. (1964). The programmed patient: a technique for appraising student performance in clinical neurology. *Journal of Medical Education* 39: 802–805.
42 Motola, I., Devine, L.A., Chung, H.S. et al. (2013). Simulation in healthcare education: a best evidence practical guide. AMEE Guide No. 82. *Medical Teacher* 35 (10): e1511–e1530.
43 Wijnen-Meijer, M., Van der Schaaf, M., Booij, M., Harendza, S., Boscardin, C., Wijngaarden, V. et al. An argument-based approach to the validation of UHTRUST: can we measure how recent graduates can be trusted with unfamiliar tasks? *Advances in Health Sciences Education* 2013; 18(5):1009–27.
44 Kalet, A., Zabar, S., Szyld, D. et al. (2017). A simulated "Night-onCall" to assess and address the readiness-for-internship of transitioning medical students. *Advances in Simulation* 2 (13): 1–9.
45 Peters, M. and ten Cate, O. (2014). Bedside teaching in medical education: a literature review. *Perspectives on Medical Education* 3 (3): 76–88.
46 Elder, A., Chi, J., Ozdalga, E. et al. (2013). The road back to the bedside. *Journal of the American Medical Association* 310 (8): 799–800.
47 Frenkel, J., Chen, H.C., and ten Cate, O. (2016). Patient-centered teaching in a technology-dominated era. *Southern Medical Journal* 109 (12): 743–746.
48 Rosenthal, D. and Verghese, A. (2016). Meaning and the nature of physician's work. *New England Journal of Medicine* 375 (19): 1813–1820.
49 Aagaard, E.M. and Abaza, M. (2016). The residency application process – burden and consequences. *New England Journal of Medicine* 374 (4): 301–303.
50 Burch, V.C. (2007). *Medical Education in South Africa – Assessment Practices in a Developing Country [doctoral dissertation]*, 1–266. Rotterdam: Erasmus University.
51 Barrows, H.S. (1983). Problem-based, self-directed learning. *Journal of the American Medical Association* 250 (22): 3077.
52 Harden, R. and Gleeson, F. (1979). Assessment of clinical competence using an objective structured clinical examination. *Medical Education* 13 (1): 41–54.
53 Frank, J.R., Snell, L.S., ten Cate, O. et al. (2010). Competency-based medical education: theory to practice. *BMC Health Serv. Res.* 32 (8): 638–645.
54 Armstrong, E.G., Doyle, J., and Bennett, N.L. (2003). Transformative professional development of physicians as educators: assessment of a model. *Academic Medicine* 78 (7): 702–708.
55 ten Cate, O., Mann, K., McCrorie, P. et al. (2014). Faculty development through international exchange: the IMEX initiative. *Medical Teacher* 36 (7): 591–595.
56 iMedtrust: a global solution for population health. http://imedtrust.org (accessed 9 December 2017).
57 ten Cate, O. (2002). Global standards in medical education – what are the objectives? *Medical Education* 36 (7): 602–604.
58 Karle, H. (2002). Global standards in medical education – an instrument in quality improvement. *Medical Education* 36 (7): 604–605.
59 Stammen, L.A., Stalmeijer, R.E., Paternotte, E. et al. (2015). Training physicians to provide high-value, cost-conscious care: a systematic review. *Journal of the American Medical Association* 314 (22): 2384–2400.
60 National Health Service (NHS) (2014). Five Year Forward View. https://www.england.nhs.uk/wp-content/uploads/2014/10/5yfv-web.pdf (accessed 9 December 2017).
61 Thistlethwaite, J.E., Bartle, E., Chong, A.A. et al. (2013). A review of longitudinal community and hospital placements in medical education: BEME Guide No. 26. *Medical Teacher* 35 (8): e1340–e1364.
62 Salas, E. and Rosen, M.A. (2013). Building high reliability teams: progress and some reflections on teamwork training. *BMJ Quality and Safety* 22 (5): 369–373.
63 Epstein, N.E. (2014). Multidisciplinary in-hospital teams improve patient outcomes: a review. *Surgical Neurology International* 5 (Suppl 7): S295–S303.
64 Reeves, S., Palaganas, J., and Zierler, B. (2017). An updated synthesis of review evidence of interprofessional education. *Journal of Allied Health* 46 (1): 56–61.

65 Lingard, L. (2012). Rethinking competence in the context of teamwork. In: *The Question of Competence: Reconsidering Medical Education in the Twenty-First Century* (ed. B.D. Hodges and L. Leingard), 42–70. Ithaca, NY: Cornell University Press.
66 Weller, J. (2012). Shedding new light on tribalism in health care. *Medical Education* 46: 134–136.
67 Weller, J., Boyd, M., and Cumin, D. (2014). Teams, tribes and patient safety: overcoming barriers to effective teamwork in healthcare. *Postgraduate Medical Journal* 90: 149–154.
68 World Health Organization (2016). Final Report of the Expert Group to the High-Level Commission on Health Employment and Economic Growth. Geneva: WHO. http://apps.who.int/iris/bitstream/10665/250040/1/9789241511285-eng.pdf?ua=1 (accessed 9 December 2017).
69 Ostler, J., Vassilas, C., and Parle, J. (2012). Physician Assistants: Friends or Foes to Doctors? *BMJ Careers*, 18 July 2012. http://careers.bmj.com/careers/advice/view-article.html?id=20008022 (accessed 9 December 2017).
70 de Graff, M. (2017). Britain is offering US physician assistants cash and long vacations in Europe if they relocate to fill desperate staff shortage. *Daily Mail*, 18 Sept 2017. www.dailymail.co.uk/health/article-4896782/Britain-begging-qualified-doctors-staff.html (accessed 9 December 2017).
71 Addicott, R., Maguire, D., Honeyman, M., and Jabbal, J. (2015). Available at: www.kingsfund.org.uk/sites/default/files/field/field_publication_file/Workforce-planning-NHS-Kings-Fund-Apr-15.pdf). *Workforce planning in the NHS*. London: The King's Fund (accessed 9 December 2017).
72 Bates, J. and Ellaway, R.H. (2016). Mapping the dark matter of context: a conceptual scoping review. *Medical EducationAnnals of Internal Medicine* 50: 807–816.
73 Cruess, S.R., Cruess, R.L., and Steinert, Y. (2010). Linking the teaching of professionalism to the social contract: a call for cultural humility. *Medical Teacher* 32 (5): 357–359.
74 ABIM Foundation 2002. Medical Professionalism in the New Millennium: A Physician Charter. http://abimfoundation.org/what-we-do/physician-charter (accessed 9 December 2017).
75 Cruess, S.R. and Cruess, R.L. (2004). Professionalism and medicine's social contract with society. *AMA Virtual Mentor* 6 (4): 1–4.
76 Hafferty, F.W. and Castellani, B. (2010). The increasing complexities of professionalism. *Academic Medicine* 85 (2): 288–301.
77 Hafferty, F.W. and Tilburt, J.C. (2015). Fear, regulations, and the fragile exoskeleton of medical professionalism. *Journal of Graduate Medical Education* 7 (3): 344–348.
78 Harris, J.M. (2017). It is time to cancel medicine's social contract metaphor. *Academic Medicine* 92: 1236–1240.
79 Twenge, J.M. (2009). Generational changes and their impact in the classroom: teaching generation me. *Medical Education* 43: 398–405.
80 Cruess, S.R., Cruess, R.L., and Steinert, Y. (2016). Entitlement in medical education: an ongoing discourse. *Canadian Medical Education Journal* 7 (2): e142–e143.
81 Mata, D.A., Ramos, M.A., Bansai, N. et al. (2015). Prevalence of depression and depressive symptoms among resident physicians: systematic review and meta-analysis. *Journal of the American Medical Association* 2115 (22): 2373–2383.
82 Rotenstein, L.S., Ramos, M.A., Torre, M. et al. (2016). Prevalence of depression, depressive symptoms, and suicidal ideation among medical students: a systematic review and meta-analysis. *Journal of the American Medical Association* 2115 (21): 2214–2236.
83 Coverdill, J.E., Alseidi, A., Borgstrom, D.C. et al. (2016). Professionalism in the twilight zone: a multicenter, mixed-methods study of shift transition dynamics in surgical residencies. *Academic Medicine* 91 (11): S31–S36.
84 Dyrbye, L.N., Trockel, M., Frank, E. et al. (2017). Development of an agenda to identify evidence-based strategies to improve physician wellness and reduce burnout. *Annals of Internal Medicine* 166 (10): 743–744.
85 Hayes, B., Prihodova, L., Walsh, G. et al. (2017). What's up doc? A national study of psychological wellbeing of hospital doctors in Ireland. *BMJ Open* 7 (10): e018023.
86 Salyers, M.P., Bonfils, K.A., Luther, L. et al. (2017). The relationship between professional burnout and quality and safety in healthcare: a meta-analysis. *Journal of General Internal Medicine* 32 (4): 475–482.
87 Shanafelt, T.D., Hasan, O., Dyrbye, L.N. et al. (2015). Changes in burnout and satisfaction with work-life balance in physicians and the general US working population between 2011 and 2014. *Clinic Proceedings* 90 (12): 1600–1613.
88 Zuboff, S. (1988). *In the Age of the Smart Machine: The Future of Work and Power*. New York, NY: Basic Books, Inc.
89 Hodges, B.D. (2018 (ePub ahead of print)). Learning from Dorothy Vaughan: artificial intelligence and the health professions. *Medical Education* doi: 10.1111/medu.13350.
90 Wachter, R. (2015). *The Digital Doctor: Hope, Hype and Harm at the Dawn of Medicine's Computer Age*. New York: McGraw Hill.
91 Schwartz, W.B. (1970). Medicine and the computer: the promise and problems of change. *New England Journal of Medicine* 283: 1257–1264.
92 Szolovits, P., Patil, R.S., and Schwartz, W.B. (1988). Artificial intelligence in medical diagnosis. *Annals of Internal Medicine* 108: 80–87.
93 Marr, B. (2016). What is the Difference Between Artificial Intelligence and Machine Learning. *Forbes*, December 6 2016. https://www.forbes.com/sites/bernardmarr/2016/12/06/what-is-the-difference-between-artificial-intelligence-and-machine-learning/#2129859c2742 (accessed 9 December 2017).
94 Jha, S. and Topol, E.J. (2016). Adapting to artificial intelligence: radiologists and pathologists as information specialists. *Journal of the American Medical Association* 316 (22): 2353–2354.
95 Shortliffe, E.H. (1976). *Computer-Based Medical Consultation: MYCIN*. New York, NY: American Elsevier.
96 Lillehaug, S.I. and Lajoie, S.P. (1998). AI in medical education – another grand challenge for medical informatics. *Artificial Intelligence in Medicine* 12: 197–225.
97 Mukherjee, S. (2017). AI versus MD: What Happens when Diagnosis is Automated. *The New Yorker*, April 3 2017. https://www.newyorker.com/magazine/2017/04/03/ai-versus-md?currentPage=all (accessed 9 December 2017).
98 Woods, N.N. (2007). Science is fundamental: the role of biomedical knowledge in clinical reasoning. *Medical Education* 41 (12): 1173–1177.
99 Woods, N.N., Brooks, L.R., and Norman, G.R. (2007). The role of biomedical knowledge in diagnosis of difficult clinical cases. *Advances in Health Sciences Education: Theory and Practice* 12 (4): 417–426.
100 Baker, A.J., Raymond, M.R., Haist, S.A., and Boulet, J. (2017). Using national health care databases and problem-based practice analysis to inform integrated curriculum development. *Academic Medicine* 92: 448–454.
101 Ellaway, R.H., Pusic, M., Yavner, S., and Kalet, A. (2014). Context matters: emergent variability in an effectiveness trial of online teaching modules. *Medical Education* 48 (4): 386–396.
102 Cook, D.A. and Triola, M.M. (2014). What is the role of e-learning? Looking past the hype. *Medical Education* 48 (9): 930–937.
103 Woodham, L.A., Ellaway, R.H., Round, J. et al. (2015). Medical student and tutor perceptions of video versus text in an interactive online virtual patient for problem-based learning: a pilot study. *Journal of Medical Internet Research* 17 (6): e151.
104 Chan, T. and Sherbino, J. (2015). The McMaster Modular Assessment Program (McMAP): a theoretically grounded work-based assessment system for an emergency medicine residency program. *Academic Medicine* 90: 900–905.
105 Schuwirth, L., van der Vleuten, C., and Durning, S.J. (2017). What programmatic assessment in medical education can learn from healthcare. *Perspectives in Medical Education* 6: 211–215.

3 The Science of Learning

Anique B. H. de Bruin[1], *Matthew Sibbald*[2,3], *and Sandra Monteiro*[4]

[1]School of Health Professions Education, Maastricht University, Maastricht, The Netherlands
[2]Department of Medicine, McMaster University, Hamilton, Ontario, Canada
[3]Centre for Simulation-Based Learning, McMaster University, Hamilton, Ontario, Canada
[4]Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, Ontario, Canada

KEY MESSAGES

- Learning is a long-term change in memory that results in the ability to transfer formal knowledge to new situations.
- Memory and attention processes interact in complex ways to facilitate learning.
- When faced with novice learners, educators should consider how to manage cognitive load, to create associations, and to facilitate deliberate practice.
- When faced with expert learners, educators should be conscious of the advantages and disadvantages of pattern recognition as well as the role for reflection in learning and reasoning.
- Applying the principles of learning and memory invariably involves invoking a learning theory. When applying a learning theory, consider carefully how the theory fits the learning content, learner expertise, and educational environment.
- The process or strategies that experts use to approach problems takes time and practice to develop. Novices must also invest time to develop their own strategies, rather than attempt to emulate the expert.

Introduction

What distinguishes us as humans from other animals is our infinite capacity to learn and the potential of our brain to adapt continuously. Whereas other animals are born with complex behaviour patterns – think of birds building nests or knowing how to fly south – we need extensive guided practice to acquire complex skills. But there are several upsides of this need, including the ability to develop very individualised skills, continue learning through the lifespan, and adapt to changing environments. Early researchers proposed that cognitive abilities are all formed prenatally and progressively decay from the age of 25. We now know that our cognitive development continues even into old age, so that an 80-year old can learn a new cognitive skill or acquire new knowledge [1]. The fascinating question that lies before us in this chapter is 'how do people learn?'

Ormrod defines learning as 'the long-term change in mental representations or associations as a result of experience' [2, p. 20]. Considering what we now know about the capacity to continue learning, however, this poses a challenge when attempting to define 'long-term'; a change or learning that occurs in medical school, may be modified later in life with new experiences. Using the science of learning, we will explore strategies that can support learning throughout the lifespan in order to build reliable, if not long-term, changes in mental representations.

The science of learning dates back to the late 1800s, when Wilhelm Wundt founded the first psychological laboratory. Introspection, that is people looking inside their heads to understand their thinking processes, was the common investigative method in those days, although this was later replaced by more objective behavioural methods (stimulus–response studies) and in-depth examination of mental phenomena. In the 1940s, humans' capacity to learn from observing (social learning) was uncovered. Around the same time, in Europe, the idea that stimulus–response observations were too limited to capture the complexity of the human mind gained support and cognitive psychology was born. More recent technological innovations have enabled the study of neural correlates of behavioural and cognitive processes, leading to the creation of a rapidly expanding field called 'cognitive neuroscience'.

In this chapter, we will first explore how learning is typically studied and measured in scientific research, before turning to the building blocks of human learning. Then, we will provide a state-of-the art synthesis of research as it relates to both novice and expert learners, concluding with an attempt to relate scientific insights to the practice of medical education.

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

Studying and Measuring Learning

As the science of learning ultimately aims at understanding, influencing, and predicting learning in real life, the measures used in scientific studies are very much related to those in educational practice. We describe measures of declarative knowledge (i.e. knowledge of facts and concepts), procedural knowledge (i.e. performance of skills), and transfer (i.e. applying knowledge in novel tasks).

Declarative knowledge is typically measured by tasks of recall or recognition. *Recall* tasks can be 'free', without guidance on what knowledge to reproduce (e.g. 'write down all you know about the circulatory system'), or 'cued', accompanied by a prompt or specific question to guide knowledge reproduction (e.g. 'what is the difference between the cardiovascular system and the lymphatic system?'). *Recognition* tasks require learners to determine whether they have specific knowledge accessible in memory, but without the need to reproduce it. This is the case in multiple-choice examination questions. The challenge for most students is to learn the content sufficiently to be able to recall or recognise. Instead, because of time pressure and inappropriate study habits, many students often achieve only a mere sense of familiarity; a vague sense that something has been experienced or seen before, which is ineffective for testing situations.

Learning non-declarative, procedural skills, or 'knowing how', is characterised by repeated practice with the skill or its component sub-skills until it can be performed automatically and without error. That this takes place (partly) implicitly was demonstrated by the now legendary amnesic patient Henry Molaison ('HM') who, without conscious awareness, was able to acquire novel skills such as mirror drawing – an eye-hand coordination task – without any recollection of having practised the skill [3]. Procedural skills are typically measured through standardised criterion tasks that are equal to or mimic real-life execution of the skill as closely as possible. Performance is then measured on sub-steps of the skills or on the complete skill execution (e.g. OSCEs or speed reading tests).

The goal of education is to acquire knowledge and skills that are usable in novel problems at a later time; learning is the ability to *transfer* formal knowledge to new situations. In research, a distinction between measures of *near* and *far transfer* is made. *Near transfer* measures have both surface and structural similarities to the learning tasks, whereas *far transfer* measures have structural similarities but different surface characteristics [4]. For example, learning about La Place's law as an explanation for neonatal respiratory distress may more easily be transferred to understanding difficulties in artificially ventilating an adult patient (near transfer) rather than applied to explain cardiac remodelling in aortic regurgitation (far transfer). The challenge for medical students is recognising the similarity in how principles apply because their learning occurred in only one context. Similarly, even experienced anaesthesiologists (say) may feel challenged to transfer their skill to a new procedure or operating room if the surgical team and equipment vary in some way; the requirement to perform the same procedure with different equipment may be a far transfer task in some cases.

Unfortunately, the concept of transfer is often overlooked when designing education. Clinical reasoning, for example, is often considered a skill [5]. The implications of this are that trainees should be able to learn clinical reasoning in discrete observable steps, much like they might learn the procedures involved in intubation. This is true to the extent that trainees can be taught an approach to taking a history, ordering tests, or documenting findings. Yet, without sufficient knowledge or experience, a clinician will struggle to make sense of even a perfectly recorded patient history and physical examination findings. Rather than focusing on improving the ability to transfer knowledge, educators are drawn in by the promise of generalisable, observable approaches to patient problems that are intended to help organise the features of the problem and make it easier to solve. The rationale for this approach is it will expedite the learning process and help trainees organise clinical information to facilitate interpretation. However, decades of research on clinical reasoning have demonstrated that one characteristic of expert clinicians is their ability to interpret information before organising it [5]. Acquiring aptitude in clinical reasoning requires more than learning approaches to history taking. While it is possible to observe a clinician 'reason' through a patient case, the underlying cognitive processes are inaccessible and unique to each clinician. Hence, clinical reasoning cannot be learned solely from watching others or using think aloud protocols. While it may be reasonable to provide trainees with the skills to interact with patients and record information, the most effective learning occurs later when they put their knowledge to use in practice.

Understanding that high-quality patient care involves the transfer of clinical knowledge means challenging students to apply their knowledge in novel situations [5, 6]. In this way, they will be engaging in learning-for-transfer. The focus then becomes less about perfecting the art of history taking and more about ensuring trainees have the opportunity to practice diagnosis and management for a variety of cases. In practice, distinguishing between strategies that support the transfer of knowledge and those that focus only on observable behaviours can be challenging. The distinction is often subtle and understanding which strategies truly improve transfer is increasingly difficult given the complex environment of medical education. Later in this chapter we provide guidance by linking the mechanisms of human learning and transfer with strategies proven effective, particularly for the context of medicine. We begin by describing some neural mechanisms that underpin learning before considering three key neural functions: memory, attention, and cognitive processing.

Neural Anatomy and Mechanisms

The adult human brain comprises about 1600 g of cells, sharing over 10^{24} connections, shaped by two hemispheres – divided into a number of anatomically and functionally distinct lobes – which in turn are connected by thick bundles of fibres, the largest of which is the corpus callosum [7]. See Figure 3.1. For the purpose of this chapter, we will

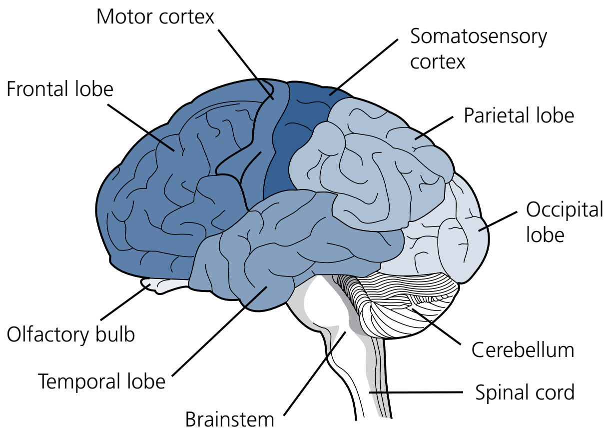


Figure 3.1 Gross anatomy of the human brain (https://commons.wikimedia.org/wiki/File:Figure_35_03_03.png).

consider the role of the hemispheres, the neuron (a brain cell which transmits and stores information as electrical activity), and neurotransmitters (chemicals in the brain that help regulate certain cognitive functions and affect) [7]. Many readers may be familiar with brain-imaging techniques such as functional magnetic resonance imaging (fMRI) which allow us to visualise and map locations in the two hemispheres that are activated for different tasks (see Box 3.1). Often neuroscience research that incorporates fMRI imaging will report a single representative image that describes neural activation patterns. A misconception that may result from this is that these patterns of activation are identical between individuals. Instead each person exhibits variations in activation areas and it is only by averaging across multiple images from different people that we can define a region of interest most often related to the performance of a task. As with individual variations in neural imaging, individuals understand information in different ways and will create different representations of the same information. Therefore, the same experience in a classroom will result in qualitatively different learning for each student. We return to this concept later when we discuss strategies for learning.

Conversely, findings that certain tasks appear lateralised – that they initiate activation in one hemisphere and not the other – have become associated with terms like 'left-brained' and 'right-brained' [8]. For example, it is often said that face processing is a right-brain task [9, 10]; that the left hemisphere is associated with logic and methodical approaches while the right hemisphere is associated with more creative and intuitive approaches [8]. There are also popular beliefs that individual differences exist that distinguish people into left- and right-brained personality types; logical or intuitive [8]. A similar misconception exists regarding our limited working use (less than 10% according to some reports) of the brain. Just as this is essentially a folktale, there is growing evidence that the lateralised brain hypothesis is incorrect. Studies have demonstrated widespread interconnectivity between brain regions even for tasks previously considered to be lateralised to one side of the brain [8, 11, 12].

Some neurons are specialised for processing certain types of input and some neurons support connections. For example, primary sensory and motor processing areas serve to receive incoming information, such as light or touch, and send signals to control muscle responses [7]. However,

BOX 3.1 FOCUS ON: Educational neuroscience

The field of educational neuroscience emerged in the late twentieth century with the aim of developing principles for education based on brain-based research. One scientific method often used in this field is functional magnetic resonance imaging (fMRI). Magnetic resonance imaging (MRI) is a broadly used brain-imaging technique, used both in research and clinical decision-making. It takes advantage of the magnetic properties of hydrogen atoms. In an MRI-scanner, participants lie in a strong magnetic field, which aligns the protons in the hydrogen atoms with the magnetic field. The alignment of those protons is then disturbed by radiofrequency (RF) pulses and the timing with which the RF pulses are emitted allows the construction of a 3D representation of the brain.

fMRI uses the same MRI apparatus, but with a different RF-pulse and measurement timing. This way, it can measure changes in blood flow and oxygenation in the brain (measuring the blood-oxygenation-level dependent [BOLD] signal), providing information about brain activity. fMRI can thus be used to measure task-dependent changes in activity, needed for investigating functional changes in the brain with increased motor expertise.

The original promise of educational neuroscience, that it would directly lead to improved teaching in the classroom, has proven overly ambitious [59]. Most studies entail replications of behavioural phenomena, but with less room for contextual variability due to methodological constraints, adding currently little to nothing to educational practice. But if we let go of the original claim, educational neuroscience can still prove beneficial by addressing neurobiological questions that emerge from other methodological approaches to educational research questions. This is possible with the following caveats [60]:

- Educational neuroscience methods are useful for studying robust phenomena, not for exploratory research questions. If we know little about a phenomenon at a behavioural level, neuroscience methods will have little to add. If no neuroscientifically appropriate measure of the phenomenon is available (yet), any attempts to design one will most likely be in vain. Prior extensive behavioural piloting of the measure can reduce risk.
- When cognitive processes are too fast to study behaviourally or when they are mostly implicit, educational neuroscience (fMRI, but also EEG) measures can prove beneficial. This requires, however, a clear neurobiological research question.

these areas make up only 25% of the human cortex, with the remaining dedicated to associating input and simple ideas to form more complex thoughts and behaviour [7]. For the purpose of understanding memory, it is useful to know that the primary function of a large proportion of neural anatomy is to find associations between sensory input [7]. These associations form patterns such that similar types of sensory input would initiate similar patterns of activation [7]. The transmission of information that occurs in these patterns of activation involves the exchange of neurotransmitters [7]. Some neurotransmitters like serotonin are implicated in the formation of memories while others are associated with motivation to learn. Fluctuations in levels of these neurotransmitters can impact both the desire to learn and the effectiveness of learning strategies. For example, dopamine has been associated with the need to seek out novel experiences. This intrinsic need may explain the drive of students to seek novelty in education, as the experience alone is rewarding, but notably novelty seeking is not itself a form of optimal learning [13]. Educators should be cautious at integrating novelty into instructional methods purely for the enjoyment of students as this may not support long-term learning.

Memories are represented by a series of specific neuronal activations that occur in parallel, or at the same time [7]. In other words, a memory is represented by a pattern of multiple neuronal cells activated simultaneously. Memories contain information about features of the environment, such as colour, shapes, even language and emotions, as well as information about relationships between features. The associative nature of human memory is quite powerful as we are able to recall important information that is cued by some other stimulus. As an analogy, a *cue* is like a search term that allows us to connect to (or 'retrieve') experiences that are related to the cue. We may consciously choose cues that help us remember, or cues in our environment may automatically activate associations to memories. For example, hearing a patient report 'chest pain' may automatically cue associations to the diagnosis of pneumonia, bronchitis, or heart failure. Additionally, a clinician may consciously search for cues by running tests to identify possible diagnoses.

Repeated exposure to cues and patterns of information in our environment, often accompanied by feedback and guidance, help us understand categories of objects that define our environment as well as concrete and abstract concepts [14–20]. The neural mechanisms that allow children to learn to distinguish between the concept of dogs and cats also support their learning of human behaviours and abstract concepts like emotion. And these same mechanisms also allow a medical resident to learn to distinguish between, say, the visually distinct features of pneumonia and pulmonary embolism (PE) in imaging, or between the auditory profiles of pulmonary crackles related to heart failure and wheezing related to asthma. The mechanism that supports the learning of these concepts is best described using terminology from research on categorisation. Using the above example, medical residents may first learn about pneumonia from a few *exemplars*, or concrete experiences with individual radiographs of patients with pneumonia. However, this is not sufficient to understand the variations present in the broad category. Eventually, with sufficient exposure, physicians not only learn how variable the category of pneumonia is, but also learn conceptual differences and similarities between the categories of pneumonia and PE. We explore specific strategies that support this form of learning later in this chapter.

Memory

Functionally, there are different types of memory: working memory, sensory memory, and long-term memory [7]. Earlier we classified the process of remembering qualitatively into recall, familiarity, and recognition. The act of forgetting has also been classified qualitatively into decay and interference [21–25]. Here we discuss remembering and forgetting in relation to cue-based retrieval. We will define the different types of memory first. See Figure 3.2.

Working memory is a more recent conceptualisation of short-term memory [5], which generally refers to the ability to maintain, manipulate, and rehearse information in an intentional or conscious manner [7]. Working memory may rely on sensory memory to retain input long enough to decide what to do with it [7]. Working memory may not always be engaged and is often reserved for times of stress, challenge, or uncertainty [7]. Essentially, sensory memory is the momentary conscious attempt to copy an input; the initial activation of neurons that experience an image, a sound, or a touch. Neuroscience researchers theorise a 'visuospatial sketchpad' (a momentary copy of a visual image) and 'phonologic loop' (a momentary copy of an auditory stimulus) that retains initial input for a very brief amount of time; roughly one second [7]. As the name implies, long-term memory refers to the massive store of experiences from our entire life [7]. Even though theoretically long-term memory should represent every experience or thought we have ever had, our ability to retrieve specific information will depend in part on how meaningful it is to us, how often we use it, and the strength of associations to the cue that may retrieve the memory.

Remembering

In some cases, we remember things in deliberate ways, such as recalling answers for a multiple-choice exam. If we

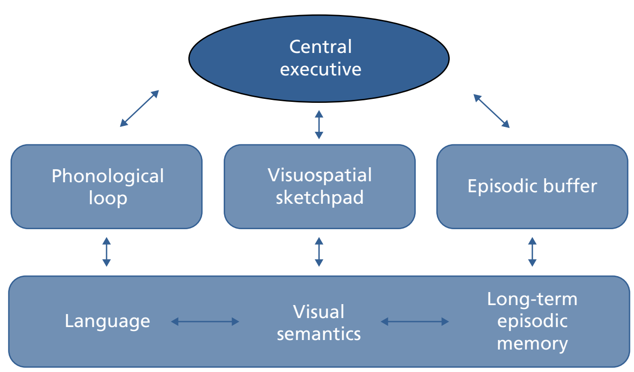


Figure 3.2 Baddeley's model of working memory.

are prepared, there will be many strong cues present in the exam itself and we may rely on cues in the questions to access the relevant information. In other cases, we remember things unintentionally, such as unwillingly recalling and reliving stressful experiences because cues in our environment automatically activate our memory. If we experienced a car accident, then any cue related to the event such as cars, roads, traffic signals, etc. may cue the memory of the accident at any time. When cues or associations are weak, we may experience remembering as a weak sense of familiarity [26]. When cues and associations are strong, we may experience remembering as explicit recall [14]. Recognition falls somewhere in between as we may have a strong sense we have seen something before, but no explicit recall of where or when [14].

Forgetting

The permanence of a memory is also determined by how many associations it has and how they might relate to cues. A single cue that is associated to multiple different memories can cause confusion about which memory is the correct or most relevant one. This experience may be interpreted as forgetting because access to the correct memory is ineffective when other memories are also activated. When there are multiple memories being accessed, which might conflict with each other, it is called *interference* [23]. Alternatively, if the association between a cue and a memory is weak, because it is rarely used, it will *decay* [22]. It is important for instructors and students to understand the mechanisms of forgetting as learning may be hindered by decay or interference. We highlight strategies that can combat interference and decay specifically.

Memory and Stress

During their medical training, students will almost continuously experience a wide range of emotions both positive (e.g. pride, gratitude, excitement, happiness) and negative (e.g. stress, fear, uncertainty) [27]. Apart from expressing a typical human ability, these emotions serve a role in training, providing, for example, warning signs to the student that action is needed to prevent harm or signposting positive development of knowledge and skill. But negative emotions and the accompanying physical responses potentially hinder learning and skill development. In particular, stress is known to affect working memory performance, especially when the stress is experienced as posing a threat (i.e. when both the sympathetic nervous system [SNS] and the hormone system respond) instead of as posing a challenge (i.e. when only the SNS responds). This happens especially when performance is also assessed under stressful circumstances. Interestingly, memory consolidation is improved by moderate levels of stress. When given a moderate dose of cortisol prior to learning words, word recognition two days later was improved [28]. In medical training, however, memories (e.g. about how to perform a surgical procedure) are only better consolidated if the stress is caused by the clinical case itself, not if peripheral factors (e.g. an intimidating senior member of staff) are the cause [29]. Finally, stress is known to have an impairing effect on memory retrieval. Skilled paramedics were hindered in their drug-dose calculations if they shortly before had gone through a number of highly stressful simulated scenarios [30]. In sum, it is not so much stress itself, but the level of stress and the learner's appraisal of stress that determine its effect on memory.

Attention and Selection

In this section we will explore what it means to selectively attend to important information, to ignore or inhibit irrelevant information, and how this fits in with memory [7]. Attention is not a unitary or concrete ability [31]. Attention is an end result of multiple cognitive processes being directed at objects or people that are relevant for a current task goal. Consequently, paying attention is often experienced as a conscious awareness that takes effort. To make things more complicated, conscious awareness or attention is not critical for learning [32, 33]. In some cases, learning can occur implicitly, without direct attention. Instructors may need to be aware of what students might learn implicitly (e.g. hidden curriculum). Attention has two potential mechanisms that support learning: selective attention and active inhibition.

Selective attention refers to the process of sharpening focus on the relevant information. Selective attention may be applied in a conscious or unconscious manner [33]. For example, salient features of a cancer lesion in a lung radiograph may capture attention automatically without effort from the radiologist. This pop out effect has been documented in many experimental and applied studies [31–38]. This can occur because the radiologist has seen multiple exemplars of cancer and her visual system and long-term memory automatically recognise the pattern. The relative salience of a pattern is therefore experience dependent; with increasing experience different patterns become more salient. Alternatively, if there are no salient features or patterns, and the radiologist is looking for evidence of cancer, she may choose one of two strategies. The radiologist may employ a serial search strategy, examining smaller sections of the image in sequence, much like examining a grid [39]. This process may incorporate the use of active inhibition; blocking the processing of extraneous information in order to highlight smaller sections. The extent to which a physician can engage in active inhibition to ignore distracting information depends on several factors, including working memory capacity, experience, and motivation [40]. The radiologist may also rely on pattern recognition in a different way by attempting to holistically process the image, in the hopes that a less salient pattern will emerge. In cognitive psychology, these strategies are not mutually exclusive, both can be effective, and have been demonstrated to interact with the demands of the task [32, 40–43].

The need for selective attention or active inhibition while learning may help strengthen associations, leading to improved memory. For example, the interaction between selective attention and memory may be responsible for the desirable difficulty effect [44, 45]. That is, the requirement for extra cognitive effort to apply selective attention may at times improve memory for the target [44, 45]. Consequently, students may learn more effectively from actively seeking out the solution to a problem. However, currently, the exact mechanisms responsible for the desirable difficulty effect are not well understood. Therefore, both instructors and learners may need to work together to determine if these principles apply to learning in their context.

Cognitive Processing

There are two main types of cognitive processing that relate to attention and memory – automatic and controlled. Automatic processes are relatively rapid and are recruited unconsciously, whereas controlled processes are relatively slower and are recruited consciously. Typically, controlled processes rely on working memory and automatic processes directly access long-term memory without involving working memory [46, 47]. Often, in the progression of learning in a new task, students start out relying on controlled processes and with time and practice rely more on automatic processes [48, 49]. Although, controlled and automatic processes are often contrasted with each other, they can occur in parallel and complement each other [46–49]. In medical education, these processes are often referred to as Type 1 and Type 2 (or System 1 and System 2). Type 1 reflects the automatic, default but powerful mode of processing that draws on a lifetime of experiences in long-term memory [48, 49]. Type 2 reflects the combination of resource intensive, executive functions including working memory and attention [48, 49].

A common misconception is that one process or system is better suited to undertake some tasks than the other. In medicine, for example, there is a prevailing belief that Type 2 is superior for the task of medical diagnosis, reflected in the push to rely on evidence-based medicine guidelines, algorithms, and debiasing [50–53]. However, instructors and students should understand that automatic and controlled processes are both important and recruited as needed to meet the demands of the task [54–58]. Box 3.1 provides a closer look at education neuroscience.

Strategies for the Novice Learner

Dunlosky and colleagues describe several popular learning strategies in higher education and extensively review research on their relative effectiveness [58]. Unfortunately, the most popular strategies, according to self-reports from college students, were not necessarily the most effective based on evidence from educational research. For example, a common study practice of students is to re-read material from textbooks or written notes. However, repeated exposure to the same material without additional attempts to construct meaningful associations may only lead to a sense of familiarity. Achieving only a sense of familiarity when re-reading will lead to a false sense of confidence with the content and a false belief that re-reading is effective. Students have been known to hold on to this belief even when presented with evidence that re-reading alone is ineffective. This suggests that learners are not typically aware of the relationship between their study habits and outcomes, indicating poor metacognitive awareness. Therefore, novices may require guided practice and more structured learning strategies that help them better detect patterns, construct relationships and so transfer their formal knowledge to various contexts [61, 62]. In this section we outline several strategies that have been shown to be successful in higher education settings. Many have been tested in health professions contexts as well. These strategies are not mutually exclusive and although they are most effective when led by the instructor they can also be implemented by trainees.

Managing Cognitive Load

A useful framework when instructing novices is cognitive load theory. This distinguishes between various aspects of the task that pose load on working memory, namely *intrinsic load*, or the load inherent in the difficulty of the learning task, *germane load*, or the load that accompanies the construction of schemas, and *extraneous load*, or the load that is generated as the result of how the learning task is designed [63, 64]. For example, when needing to manage a disruptive family member during the examination of an acutely sick child, the management of the family member may be extraneous to learning the skills of diagnosing and treating the child. In another context, however, managing the disruptive behaviour of the family member may be the critical skill being learned. For this reason, the goal of the learning task determines what is considered intrinsic and extraneous load. Instructors may need to explore various approaches to identify the most effective strategies for reducing cognitive load. In the context of multimedia presentations (e.g. PowerPoint slides), extraneous load may be more obvious. For example, the extensive use of graphics, technology, or overlay of text with animations and sounds can decrease learning considerably compared to simpler presentation techniques (i.e. plain text or static images) [63]. Another example of extraneous load is the use of augmented reality technology for instruction in anatomy. Students with lower working memory capacity may be disadvantaged when learning anatomy using rotating, computer controlled, augmented reality animations, compared to multiple, student controlled, still views [65].

Identifying components that reduce intrinsic load may be more challenging, but there are some strategies for managing it. For example, the use of worked examples is theorised to reduce the intrinsic load of learning the mechanics of problem solving in statistics [66]. A similar strategy may be appropriate for teaching medical students to calculate the total water deficit in a patient presenting with hypernatraemia to target fluid replacement therapy. Some research suggests that students do show improved performance following explanations using already solved or worked problems compared to working through the same problems on their own [67]. However, this benefit may be modulated depending on the content or level of the learner; the use of worked examples leads to decreased performance with more intermediate or expert learners [68, 69].

Creating Associations

The following strategies are theorised to build connections or associations between cues and prior knowledge: testing, integration, and blocked and interleaved practice.

Testing as a study strategy has been linked to test enhanced learning, i.e. improved performance on a subsequent test following an initial (formative) test [70, 71]. Testing may help strengthen links between simple cues, such as key clinical features contained in a question stem, and more complex diagnostic categories, contained in the question options [70, 71]. One mechanism that is proposed to support test-enhanced learning is retrieval practice or repeated exposure to cues that facilitate access to important information. However, testing is most effective with feedback on the correct answers [70]. Another important feature of testing that improves learning is varying the format of questions using different cues. For example, instructors might create multiple versions of the same question so the same content is applied to different contexts. In this way instructors can help facilitate near and far transfer.

Integrating instruction in basic science principles with clinical concepts may also strengthen associations between underlying biomedical principles and the manifestation of clinical symptoms [62, 72, 73]. This approach to teaching the medical sciences can be challenging as there is often a separation of basic science topics, such as cell development, from clinical concepts, such as cancer. However, evidence suggests that students develop a stronger understanding of the material when it is successfully integrated [62, 72, 73]. Successful integration requires more than pairing lectures from the content areas, but instead requires specific and explicit linking between concepts [73]. Baghdady and colleagues demonstrated improved performance on a diagnostic test for students who learned about dental disorders using images and text that integrated knowledge of the anatomy, basic science, and clinical presentation on a single slide compared to students who learned the same content separated into two slides [74]. A simple division of material over time was sufficient to disadvantage students. Instructors may need to collaborate to successfully integrate content from previously separate lectures.

Interleaving and *blocking* practice problems allows learners to practise identifying the nature of the problem, or categorising it, before applying a solution [74–77]. For example, medical students are often taught the principles of electrocardiography (ECG) in a lecture format and then, as individual clinical diagnoses are introduced, they are provided with a few exemplars. Instead, students require multiple exemplars of each diagnosis in order to recognise the relevant patterns and the different presentations [77]. This is referred to as blocked practice; the goal is not to identify the diagnosis, but rather to recognise similarities and differences of items within the same category. Even though there may be clear indicators of each diagnosis (e.g. ST elevation to suggest acute myocardial infarction) the manifestation of those will vary from patient to patient. Then, once sufficient mastery of the material has been achieved students should be challenged with interleaved practice, mixing exemplars from different categories, where the goal is to identify the diagnosis. Interleaved practice would allow students to learn to discriminate between similar looking ECGs that are actually from different categories; for example, comparing ST elevation due to acute myocardial infarction with ST elevation due to acute pericarditis. This form of study also matches the requirements of real practice and allows students to develop the skills required as independent practitioners.

Deliberate Practice

Although identified in expert learners, *deliberate practice* can also aid novice learners in their first steps of expertise development. It incorporates a series of planned practice sessions with targeted feedback and in-depth analysis [78]. It is typically relevant to learning procedural skills and is similar in approach to learning a musical instrument or a sport [79]. The key elements are a coach, feedback, assessment, and mastery learning. Independent of these elements, practice sessions may also be organised according to a schedule that maximises learning. The mechanisms that support this strategy include repeated and focused learning on correct movements through feedback and correction. However, in medicine there are many challenges to following a deliberate practice model. Trainees may not have access to the equipment, or the procedures may be rare, or they may not have access to skilled coaches. The following strategies may mitigate these challenges.

Trainees may apply *spaced* practice by revisiting a skill every month for brief amounts of time to ensure they are able to perform it consistently. Scheduling guided practice in a spaced manner has been shown to lead to better retention of the skill over time compared to scheduling practice only on one academic half-day [80]. These shorter practice sessions may be recorded and examined at a later date with a coach. *Peer to peer* feedback has also been shown to be effective in helping trainees advance in their skill level [81, 82]. Effective use of paired learning may reduce the burden on staff or senior physicians. During gaps in training or practice, trainees may also employ *mental rehearsal* to maintain some level of skill [83, 84]. Instructors may guide students through an exercise of imagining the steps of a procedure prior to actual practice to help compensate for the lack of practice [83, 84].

Finally, students may be placed in charge of their own learning by employing an approach of *mastery learning* [85]. For example, they may independently practise with simple aspects of suturing, using inexpensive resources to build up speed and knot-tying skills. A more experienced coach can then step in to help the trainee fine tune their movements with more advanced techniques. The use of progress tests in medical schools has also proved useful to help students self-identify their need for further coaching. For example, progress tests have been used at Maastricht and McMaster University medical schools [86]. Because the progress test is designed to assess knowledge at the level of the graduate, students can gauge their current level of knowledge in comparison to that of a graduating student. They can also compare themselves against their own class [87, 88]. This form of guided self-assessment has proven more successful in identifying students in need for remediation.

Strategies for the Expert Learner

Where the novice's learning is mostly characterised by juggling the limits of working memory capacity, the expert learner can reap the benefits of chunking, schema, and script formation [101]. While the various constructs of illness scripts, schemas, and chunks are often discussed independently of each other, they are all related to the associationist perspective of memory. That is, in a chunk, schema, or script separate pieces of information have become interrelated because of repeated, joint retrieval, and are accessed as a single unit in working memory, thereby reducing the cognitive load considerably and maximising processing capacity. A *chunk* [102] reflects linked information, when organised in a way that facilitates distinguishing between diagnoses it is called a *schema* [103, 104], and when it represents a prototype of a disease including its common features, clinical presentation, and response to management it is termed an *illness script* [105]. A family physician, for example, diagnosing a female patient who is complaining of a slowly progressive headache at the left side of the head will rapidly activate the 'migraine script' [106]. Note that prior knowledge and repeated exposure to the constellation of symptoms is needed to enable script formation and activation.

Pattern Recognition and Analytical Reasoning

The formation of chunks, schemas, and scripts underlines a defining characteristic of human cognition, namely the tendency to form and recognise patterns through association. Pattern recognition greatly improves processing speed and accuracy and is a necessary prerequisite for complex skill development. Only when sub-skills become routine and/or pattern recognition has emerged will development of more advanced skills (utilising the routine skills as building blocks) become possible. Some scholars suggest that pattern recognition, a form of 'Type 1' processing' [107, 108], carries a risk as this heuristic type of processing is prone to biases, such as premature closure [109]. Since patterns are recognised without effortful deliberation, but through rapid, automatic association, they are believed to work well in standardised situations and prototypical clinical cases. In clinical reasoning, however, 'standardised' and 'prototypical' are terms hardly encountered and clinicians' work is mostly characterised as being prepared to expect the unexpected. Type 2, or analytical, processing is thought to correct for errors and a lack of experience by compensating for missing information. However, general strategies to be more analytic have not proved beneficial; if physicians do not know the answer right away unguided reflection alone cannot target the error [110, 111].

Practical strategies that have been demonstrated to promote the appropriate application of either Type 1 or Type 2 processing are scarce [112]. Attempts to focus on debiasing strategies specifically have not been successful [112]. Instead, strategies that are successful tend to organise knowledge and re-evaluate clinical features in very specific ways [113]. For example, the use of clinical checklists offers some promise. In a number of studies, Sibbald and colleagues have demonstrated that the guided application of checklists can support the diagnostic process in ECG interpretation [113, 114]. This strategy requires exploration in other medical contexts.

Pattern recognition is still possible, when constellations of symptoms tell the physician something is the matter, even though the correct diagnosis is not immediately recognised. For example, several studies document variations of Type 1 processing that support accurate diagnosis. Woolley and Kostopoulou [115] refer to physician's gut feelings and insight as important factors in the diagnostic process and Van den Bruel et al. [116] note the benefit of following these gut feelings in preventing serious illness in children.

What does this mean for the expert learner in medicine? We will assume that the expert learner has developed sufficient skill to optimise knowledge and skill acquisition where needed (e.g. when a novel treatment is put in place, or new technology needs to be learned). Learning strategies described under the previous section are then effectively enacted to fill the knowledge or skill gap. However, the expert learner must guard against the limits of their own expertise and overuse of pattern recognition or Type 1 processing. Considerable debate and controversy exists on the role of each reasoning process in expert errors, and how these should be corrected. Some researchers argue that many expert errors stem from bias inherent in unbalanced use of Type 1 processing. If this is the case, then experts need to be aware of when Type 1 processing is in place, and recognise under what circumstances reasoning is jeopardised by biases, but also learn *how* to balance Type 1 for Type 2 processing.

Reflective Practice

More than at any other level of learning, the expert learner's actions are driven by prior reflection on where his learning stands and where effort should be exerted to further improve performance and expertise. Reflection is also the basis of deliberate practice; that is, only when an expert has accurately identified what competences or skills to work on, is effective deliberate (instead of rote) practicing of novel (sub)skills possible [78]. But even for expert learners, accurately diagnosing their level of learning and learning needs is not straightforward. This requires careful scrutiny of multiple feedback sources, with the help of a coach or mentor to interpret information [64]. Synthesis and interpretation of feedback will then provide informative cues as to where learning efforts are best directed [117]. Competency-based education speaks to this need, by providing students, early on, a framework to guide reflection and raise awareness of the areas of performance they are expected to develop [118]. It is generally assumed that a minimum level of knowledge and experience in a domain is prerequisite (but not sufficient) to accurate self-reflection.

To sum up, the expert learner utilises strategies developed earlier on in training to optimise knowledge and skill development. How and where knowledge and skill improvement is needed is determined through reflection, with or without guidance by a mentor. The expert learner may experience caveats in more advanced knowledge and skills such as when adapting to novel technology or sensing insecurity about specific skills. It almost goes without saying that these endeavours all take place within the context of a specific domain, and do not generally transfer to other domains without extensive training (see Boxes 3.2 and 3.3).

BOX 3.2 Focus on neural plasticity

The common phrase 'use it or lose it' summarises the nature of neural plasticity. It is the developmental and neurophysiological stages that support learning early in life that set the stage for continued learning later in life. The same mechanisms that allow infants to develop sensitivity for human faces [89], or native musical rhythms [90], also facilitate the development of expertise in any skill. Abilities, supported by neural connections or associations that are critical to the individual, are strengthened and those that are not used are lost. For example, newborn infants can distinguish between phonemes (i.e. speech sound fragments) from many languages, but once accustomed to their native language they lose sensitivity to detect phonemes not present in their native tongue [91]. This process of neural shaping and pruning is referred to as plasticity, and was previously considered terminated by late adolescence [92, 93].

Work on recovery from brain trauma indicates that neural plasticity is possible late in life; however, not much is known about how to successfully encourage plasticity or synaptogenesis (i.e. growth of new neural connections) under normal conditions. While changes caused by damage or ageing lead to observable behavioural changes, in some cases behavioural changes can lead to neural changes [94]. However, neural changes become increasingly more difficult with age, requiring extreme conditions such as loss of a sense to trigger plastic changes in the brain or behaviour [94]. This is because the cognitive system is primed to achieve stability. For learners this means that as they focus on skill development in one area, disuse of other skills makes those skills weaker; the system provides increased advantages for the new skill and may take resources away from other connections [93]. Achieving this stability in one skill may have negative consequences as physicians seek recertification or re-training in another specialty.

Some work with non-human animal research suggests that exposure to stimulating, enriched environments may be useful [95]. If newborn animals are deprived of necessary experiences early in life, such as exposure to light, the mechanisms of neural plasticity will prune away the unused connections, such as sensitivity to contrast. However, gradually introducing the animal to light again can initiate different mechanisms and re-growth of those lost connections [96, 97]. The equivalent in medicine is the notable loss of skill in other areas once a physician has specialised. A senior obstetrician, for example, might struggle to diagnose a male patient with hypertension, but easily recognise and treat pre-eclampsia in pregnant women. This has been identified as a problem for an increasingly complex health care system [98–100]. Additionally, migrating physicians or those choosing to switch specialties will experience extensive challenges with the transitions. The solution may be the use of enriched environments or high demand simulations to encourage new connections to form.

BOX 3.3 WHERE'S THE EVIDENCE: Novice–expert learning differences

This chapter has explored different strategies based on an assumption that instructional techniques have different relative effects based on learning expertise level. Indeed, the literature is full of examples of novice–expert interactions with individual learning techniques. For instance, Kayluga and colleagues describe the concept of expertise reversal [68], an idea that techniques which improve novice learning may be less effective and possibly harmful for expert learning. Many of the differential effects of instructional techniques have to do with the presentation of information. Experts tolerate and may even benefit from more context and complexity in contrast to the novice who benefits from structure and scaffolding. For example, the benefit of worked examples is attenuated and might even be reversed among experts, whereas freedom to discover or explore may selectively curtail novice learning [67].

Defining and identifying expertise to tailor instructional techniques and materials has subtleties. Many experimental designs create novice–expert contrast by comparing learners at different points in their careers (e.g. comparing medical students with practising physicians). Those learners who already function within a clinical environment have a tacit understanding of the workplace and their clinical role – a potential mediator of expertise reversal. However, expertise has many other relevant components, including past educational experiences, familiarity with the instructional technique, and ability at solving problems outside of routine practice (termed 'adaptive' expertise) [119]. For example, online multimedia training of senior faculty who have little experience in this instructional modality may respond best to novice instructional techniques around the learning process, even though they have substantive content expertise.

Expertise in the process of learning itself deserves special mention. In contrast to content experts who have command over a content domain, those who learn with ease in a variety of contexts are expert learners. Expert learners are believed to engage in activities and behaviours that maximise learning success [120]. These include (i) planning learning tasks with personal abilities and task requirements in mind, (ii) monitoring progress, motivation, and concentration, and (iii) reflecting and evaluating on the success of the approach [121]. Whereas novice learners can be easily overwhelmed if they have to split their attention between learning content and managing their learning process, expert learners may benefit from the latitude to manage their own learning process.

Challenges in Applying Neuroscience and Theory to Practice

We have provided in this chapter an overview of the science of learning and memory as it relates to medical education. Applying these principles of learning and memory requires the use of a learning theory, or way of understanding how the data fit together and can be interpreted in usable form. We allude to several learning theories including cognitive load theory, mastery learning, and deliberate practice. However, these theories occasionally overlap or even conflict with each other. For example, one common conflict exists between theories which advocate for minimal instructional guidance (e.g. constructivism, problem-based learning approaches, experiential learning and inquiry learning) and those providing more structured support (e.g. mastery learning, deliberate practice, cognitive load theory) [122]. These conflicts highlight the complex challenge of integrating and applying the literature data on the basic science of learning and memory in practice. Theories are useful, but fallible, offering unique but incomplete perspectives.

The utility of theories lies in their ability to help translate the experimental findings of the science of learning to the health professions context [123]. How does an educator minimise extraneous load for a medical resident training to be an internist working in a hospital environment? Translation is inherently subjective, and much has been written to help educators navigate the challenges [124]. Educators need to consider how these principles apply across different content domains, individual learning preferences, and in varied environments, which may be unstructured, inconsistent, and dynamic.

Content Factors

Many of the principles of learning and memory described in this chapter are robust, having been replicated across a variety of different content domains. However, how these effects translate across different learning task types is less well studied. For instance, it is less clear whether spacing or distributed practice is of equal value in cognitive learning tasks, skill-based learning tasks, and social learning. Whereas the existing literature on learning and memory focuses mostly on cognitive and skill-based tasks, social learning tasks, such as communication, collaboration, managing, and leadership skill development, are becoming more formalised components of emergent health professions curricula, as they align more responsively to workplace and practice environments.

Individual Factors

It is not unusual for a learner to claim that they are a 'visual' or 'auditory' learner. The idea that learners vary in their response to instructional techniques and format is a hypothesis that many advocates of learning styles find appealing. After all, who wouldn't want to diagnose a learner's learning style and individualise their instruction? While learners are often vocal about their learning styles, and countless correlations with learning and performance have been found in many areas of medical education, evidence that modifying instruction based on an individual's learning style will enhance their learning is scant to non-existent [124–126].

Environmental Factors

Most health professional training programmes begin with classroom-based learning and migrate to workplace contexts. Many of the studies in the basic science of learning and memory were conducted in classroom settings. Applying these findings to the workplace requires accommodation for the added pressures on trainees to negotiate their social integration into the workplace and balance their learning needs against their contribution to patient care. Transfer of learning from the classroom setting to the workplace is complex [127], with different dimensions of the workplace context to consider [128, 129]. However, the context for learning need not be an exact replica of the practice setting where the learning is to be applied. The functional similarity of the instruction to the workplace task may be more important than the physical similarity [130].

The impact of sleep deprivation on learning deserves special mention, as many health care trainees are exposed to interruptions in the sleep wake cycle during training. Sleep is believed important in the consolidation of experiences into long-term memory, and should have impact on learning [131]. Within the field of health professions education, the impact of sleep deprivation on performance and medical errors has been studied with mixed results [132]. However, learning outcomes are infrequent endpoints. One study found little impact on learning based on assigned journal reading under sleep deprivation conditions, but noted increased fatigue and decreased motivation [133–135]. While the data is still evolving, many health professions training programmes have reworked duty hours to reduce sleep deprivation in the hope of enhancing patient care and improving learning.

Areas for Future Research

While much is known about memory and learning, many of the tools to study learning are evolving and becoming more accessible, such as portable eye tracking and functional magnetic resonance. These technological advances provide fresh perspective on well-established theories and principles [136]. Furthermore, enhanced portability allows more study in the workplace to understand how the principles of learning and memory apply in the health care setting.

Conclusion

This chapter explored the science and theory of learning and memory to inform medical education practice. We began by defining learning as a long-term change in mental representation, and we showed how it can be measured. We highlighted the importance of transfer, or application, of formal knowledge to new situations. We argued that the key principles of memory, selection, and attention provide insights into why we remember or forget. Finally, we discussed how these principles apply across the spectrum of learner expertise. For novices, educators should be particularly sensitive to cognitive load demands, and should harness strategies for creating associations and promoting deliberate practice. For experts, educators should keep in mind the role of different reasoning strategies, and the importance of reflection. Bringing the science and theory of learning into practice is an art, requiring careful consideration of learners, content, and environment.

References

1 Nurke, S.N. and Barnes, C.A. (2006). Neural plasticity in the aging brain. *Nature Reviews Neuroscience* 7: 30–40.

2 Ormrod, J.E. (2016). *Human Learning*. Upper Saddle River, New Jersey: Prentice Hall.

3 Squire, L.R. (2009). The legacy of patient H.M. for neuroscience. *Neuron* 61: 6–9.

4 Ross, B.H. (1987). This is like that: the use of earlier problems and the separation of similar effects. *Cognitive Psychology* 13: 629–638.

5 Gruppen, L.D. (2017). Clinical reasoning: defining it, teaching it, assessing it, studying it. *The Western Journal of Emergency Medicine* 18 (1): 4.

6 Regehr, G. and Norman, G.R. (1996). Issues in cognitive psychology: implications for professional education. *Academic Medicine* 71: 988–1001.

7 Reisberg, D. (2010). *Cognition: Exploring the Science of the Mind*, 6e. New York: WW Norton & Company Incorporated.

8 Nielsen, J.A., Zielinski, B.A., Ferguson, M.A. et al. (2013). An evaluation of the left-brain vs. right-brain hypothesis with resting state functional connectivity magnetic resonance imaging. *PLoS One* 14 (8): e71275.

9 Uono, S., Sato, W., Kochiyama, T. et al. (2017). Time course of gamma-band oscillation associated with face processing in the inferior occipital gyrus and fusiform gyrus: a combined fMRI and MEG study. *Human Brain Mapping* 38 (4): 2067–2079.

10 Keller, C.J., Davidesco, I., Megevand, P. et al. (2017). Tuning face perception with electrical stimulation of the fusiform gyrus. *Human Brain Mapping* 38 (6): 2830–2842.

11 Frässle, S., Paulus, F.M., Krach, S. et al. (2016). Mechanisms of hemispheric lateralization: asymmetric interhemispheric recruitment in the face perception network. *NeuroImage* 124: 977–988.

12 Kolb, B. and Gibb, R. (2011). Brain plasticity and behaviour in the developing brain. *Journal of the Canadian Academy of Child and Adolescent Psychiatry* 20 (4): 265.

13 Costa, V.D., Tran, V.L., Turchi, J., and Averbeck, B.B. (2014). Dopamine modulates novelty seeking behavior during decision making. *Behavioral Neuroscience* 128 (5): 556.

14 Jacoby, L.L. (1973). Encoding processes, rehearsal, and recall requirements. *Journal of Verbal Learning and Verbal Behavior* 12: 302–310.

15 Jacoby, L.L. and Hollingshead, A. (1990). Toward a generate/recognize model of performance on direct and indirect tests of memory. *Journal of Memory and Language* 29: 433–454.

16 Jacoby, L.L. (1991). A process dissociation framework: separating automatic from intentional uses of memory. *Journal of Memory and Language* 30: 513–541.

17 Smith, J.D. and Minda, J.P. (1998). Prototypes in the mist: the early epochs of category learning. *Journal of Experimental Psychology, Learning, Memory, and Cognition* 24: 1411–1436.

18 Murphy, G.L. and Medin, D.L. (1985). The role of theories in conceptual coherence. *Psychological Review* 92: 2–316.

19 Medin, D.L. and Schaffer, M.M. (1978). Context theory of classification learning. *Psychological Review* 85 (3): 207–238.

20 Medin, D.L. (1989). Concepts and conceptual structure. *The American Psychologist* 44 (12): 14–81.

21 Shea, J.B. and Morgan, R.L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory* 5 (2): 179–187.

22 Sikström, S. (2002). Forgetting curves: implications for connectionist models. *Cognitive Psychology* 45 (1): 95–152.

23 Mensink, G.J. and Raaijmakers, J.G. (1988). A model for interference and forgetting. *Psychological Review* 95 (4): 434–455.

24 Oberauer, K., Farrell, S., Jarrold, C. et al. (2012). Interference between maintenance and processing in working memory: the effect of item–distractor similarity in complex span. *Journal of Experimental Psychology, Learning, Memory, and Cognition* 38 (3): 665–685.

25 Lewandowsky, S. and Oberauer, K. (2009). No evidence for temporal decay in working memory. *Journal of Experimental Psychology, Learning, Memory, and Cognition* 35 (6): 1545–1551.

26 Whittlesea, B.W., Jacoby, L.L., and Girard, K. (1990). Illusions of immediate memory: evidence of an attributional basis for feelings of familiarity and perceptual quality. *Journal of Memory and Language* 29 (6): 716–732.

27 McConnell, M.M. and Eva, K.W. (2012). The role of emotion in the learning and transfer of clinical knowledge and skills. *Academic Medicine* 87 (10): 1316–1322.

28 Elzinga, B.M. and Roelofs, K. (2005). Cortisol-induced impairments of working memory require acute sympathetic activation. *Behavioral Neuroscience* 119: 98–103.

29 Christianson, S.-A. (1992). Emotional stress and eyewitness memory: a critical review. *Psychological Bulletin* 112: 284–309.

30 LeBlanc, V.R., McArthur, B., King, K. et al. (2005). Paramedic performance in calculating drug dosages following stressful scenarios in a human patient simulator. *Prehospital Emergency Care* 9: 439–444.

31 Treisman, A.M. and Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology* 12 (1): 97–136.

32 Merikle, P.M., Smilek, D., and Eastwood, J.D. (2001). Perception without awareness: perspectives from cognitive psychology. *Cognition* 79 (1): 115–134.

33 Merikle, P.M. and Joordens, S. (1997). Parallels between perception without attention and perception without awareness. *Consciousness and Cognition* 6 (2): 219–236.

34 Buschman, T.J. and Miller, E.K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science* 315 (5820): 1860–1862.

35 Drew, T., Evans, K., Võ, M.L. et al. (2013). Informatics in radiology: what can you see in a single glance and how might this guide visual search in medical images? *Radiographics* 33 (1): 263–274.

36 Evans, K.K., Birdwell, R.L., and Wolfe, J.M. (2013). If you don't find it often, you often don't find it: why some cancers are missed in breast cancer screening. *PLoS One* 8 (5): e64366.

37 Evans, K.K., Georgian-Smith, D., Tambouret, R. et al. (2013). The gist of the abnormal: above chance medical decision making in the blink of an eye. *Psychonomic Bulletin & Review* 20 (6): 1170–1175.

38 Evans, K.K., Cohen, M.A., Tambouret, R. et al. (2011). Does visual expertise improve visual recognition memory? *Attention, Perception, & Psychophysics* 73 (1): 30–35.

39 Drew, T., Võ, M.L., Olwal, A. et al. (2013). Scanners and drillers: characterizing expert visual search through volumetric images. *Journal of Vision* 13 (10): 3.

40 Kane, M.J. and Engle, R.W. (2003). Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General* 132 (1): 47–70.

41 Engle, R.W., Kane, M.J., and Tuholski, S.W. (2004). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. *Psychological Bulletin* 130 (4): 553–573.

42 Kramer, A.F., Humphrey, D.G., Larish, J.F., and Logan, G.D. (1994). Aging and inhibition: beyond a unitary view of inhibitory processing in attention. *Psychology and Aging* 9 (4): 491–512.
43 Kane, M.J. and Engle, R.W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: an individual-differences perspective. *Psychonomic Bulletin & Review* 9 (4): 637–671.
44 Bjork, E.L. and Bjork, R.A. (2014). Making things hard on yourself, but in a good way: creating desirable difficulties to enhance learning. In: *Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society*, 2e (ed. M.A. Gernsbacher and P. Pomerantz), 59–68. New York: Worth.
45 Bjork, E.L., Little, J.L., and Storm, B.C. (2014). Multiple-choice testing as a desirable difficulty in the classroom. *Journal of Applied Research in Memory and Cognition* 3 (3): 165–170.
46 Shiffrin, R.M. and Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review* 84 (2): 127–190.
47 Schneider, W. and Shiffrin, R.M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review* 84 (1): 1–66.
48 Evans, J.S. (2003). In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences* 7 (10): 454–498.
49 Evans, J.S. and Stanovich, K.E. (2013). Dual-process theories of higher cognition: advancing the debate. *Perspectives on Psychological Science.* 8 (3): 223–241.
50 Graber, M.L., Franklin, N., and Gordon, R. (2005). Diagnostic error in internal medicine. *Archives of Internal Medicine* 165 (13): 1493–1499.
51 Graber, M.L., Kissam, S., Payne, V.L. et al. (2012). Cognitive interventions to reduce diagnostic error: a narrative review. *BMJ Quality and Safety* 21 (7): 535–557.
52 Mamede, S., Schmidt, H.G., and Penaforte, J.C. (2008). Effects of reflective practice on the accuracy of medical diagnoses. *Medical Education* 42 (5): 468–475.
53 Mamede, S., van Gog, T., van den Berge, K. et al. (2010). Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. *Journal of the American Medical Association* 304 (11): 1198–1203.
54 Klein, G.A. and Hoffman, R.R. (1993). Seeing the invisible: perceptual-cognitive aspects of expertise. In: *Cognitive Science Foundations of Instruction* (ed. M. Rabinowitz), 203–226. Hillsdale, NJ: Lawrence Erlbaum Associates.
55 Phillips, J.K., Klein, G., and Sieck, W.R. (2004). Expertise in judgment and decision making: a case for training intuitive decision skills. In: *Blackwell Handbook of Judgment and Decision Making* (ed. D.J. Koehler and N. Harvey), 297–315. Oxford: Blackwell.
56 Gigerenzer, G. and Goldstein, D.G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological Review* 103 (4): 650.
57 Shanteau, J. (1988). Psychological characteristics and strategies of expert decision makers. *Acta Psychologica* 68 (1): 203–215.
58 Dunlosky, J., Rawson, K.A., Marsh, E.J. et al. (2013). Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest* 14 (1): 4–58.
59 Gabrieli, J.D.E. (2016). The promise of educational neuroscience: comment on Bowers (2016). *Psychological Review* 123: 613–619.
60 De Bruin, A.B.H. (2016). The potential of neuroscience of health professions education: towards convergence of evidence and resisting seductive allure. *Advances in Health Sciences Education* 21: 983–90.
61 Brown, P.C., Roediger, H.L. III, and McDaniel, M.A. (2014). *Make It Stick: The Science of Successful Learning*. Harvard, Cambridge, MA: Harvard University Press.
62 Norman, G. (2009). Teaching basic science to optimize transfer. *Medical Teacher* 31 (9): 807–811.
63 Van Merriënboer, J.J., Kirschner, P.A. and Kester, L. (2003). Taking the load off a learner's mind: instructional design for complex learning. *Educational Psychologist* 38(1):5–13.
64 Paas, F., Renkl, A., and Sweller, J. (2004). Cognitive load theory: instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science* 32 (1): 1–8.
65 Khot, Z., Quinlan, K., Norman, G.R., and Wainman, B. (2013). The relative effectiveness of computer-based and traditional resources for education in anatomy. *Anatomical Sciences Education* 6 (4): 211–215.
66 Paas, F.G. and Van Merrienboer, J.J. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: a cognitive-load approach. *Journal of Educational Psychology* 86 (1): 122.
67 Kalyuga, S., Chandler, P., Tuovinen, J., and Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology* 93 (3): 579–588.
68 Kalyuga, S., Ayres, P., Chandler, P., and Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist* 38 (1): 23–31.
69 Kalyuga, S., Chandler, P., and Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors* 40 (1): 1–7.
70 Roediger, H.L. III and Karpicke, J.D. (2006). The power of testing memory: basic research and implications for educational practice. *Perspectives on Psychological Science* 1 (3): 181–210.
71 Butler, A.C. and Roediger, H.L. III (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology* 19 (4–5): 514–527.
72 Baghdady, M.T., Carnahan, H., Lam, E.W., and Woods, N.N. (2013). Integration of basic sciences and clinical sciences in oral radiology education for dental students. *Journal of Dental Education* 77 (6): 757–763.
73 Kulasegaram, K.M., Martimianakis, M.A., Mylopoulos, M. et al. (2013). Cognition before curriculum: rethinking the integration of basic science and clinical learning. *Academic Medicine* 88 (10): 1578–1585.
74 Rohrer, D. (2012). Interleaving helps students distinguish among similar concepts. *Educational Psychology Review* 24 (3): 355–367.
75 Carvalho, P.F. and Goldstone, R.L. (2014). Putting category learning in order: category structure and temporal arrangement affect the benefit of interleaved over blocked study. *Memory & Cognition* 42 (3): 481.
76 Carvalho, P.F. and Goldstone, R.L. (2015). The benefits of interleaved and blocked study: different tasks benefit from different schedules of study. *Psychonomic Bulletin & Review* 22 (1): 281.
77 Birnbaum, M.S., Kornell, N., Bjork, E.L., and Bjork, R.A. (2013). Why interleaving enhances inductive learning: the roles of discrimination and retrieval. *Memory & Cognition* 41 (3): 392–402.
78 Ericsson, K.A., Krampe, R.T., and Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review* 100 (3): 363–406.
79 Watling, C. (2014). Resident teachers and feedback: time to raise the bar. *Journal of Graduate Medical Education* 6 (4): 781–782.
80 Moulton, C.A., Dubrowski, A., MacRae, H. et al. (2006). Teaching surgical skills: what kind of practice makes perfect?: a randomized, controlled trial. *Annals of Surgery* 244 (3): 400.
81 Ten Cate, O. and Durning, S. (2007). Peer teaching in medical education: twelve reasons to move from theory to practice. *Medical Teacher* 29 (6): 591–9.
82 Field, M., Burke, J.M., McAllister, D., and Lloyd, D.M. (2007). Peer-assisted learning: a novel approach to clinical skills learning for medical students. *Medical Education* 41 (4): 411–418.
83 Sanders, C.W., Sadoski, M., van Walsum, K. et al. (2008). Learning basic surgical skills with mental imagery: using the simulation centre in the mind. *Medical Education* 42 (6): 607–612.
84 Sanders, C.W., Sadoski, M., Bramson, R. et al. (2004). Comparing the effects of physical practice and mental imagery rehearsal on learn-

ing basic surgical skills by medical students. *American Journal of Obstetrics and Gynecology* 191 (5): 1811–1814.

85 Dreyfus, S.E. and Dreyfus, H.L. (1980). A Five-stage Model of the Mental Activities Involved in Directed Skill Acquisition. University of California at Berkeley, Operations Research Center. http://www.dtic.mil/dtic/tr/fulltext/u2/a084551.pdf (accessed 31 July 2017).

86 Vleuten, C.V., Verwijnen, G.M., and Wijnen, W.H. (1996). Fifteen years of experience with progress testing in a problem-based learning curriculum. *Medical Teacher* 18 (2): 103–109.

87 Blake, J.M., Norman, G.R., Keane, D.R. et al. (1996). Introducing progress testing in McMaster University's problem-based medical curriculum: psychometric properties and effect on learning. *Academic Medicine* 71 (9): 1002–1007.

88 Van der Vleuten, C.P., Schuwirth, L.W., Muijtjens, A.M. et al. (2004). Cross institutional collaboration in assessment: a case on progress testing. *Medical Teacher* 26 (8): 719–25.

89 Pascalis, O., Scott, L.S., Kelly, D.J. et al. (2005). Plasticity of face processing in infancy. *Proceedings of the National Academy of Sciences of the United States of America* 102 (14): 5297–5300.

90 Lewkowicz, D.J. and Ghazanfar, A.A. (2009). The emergence of multisensory systems through perceptual narrowing. *Trends in Cognitive Sciences* 13 (11): 470–478.

91 Werker, J.F. and Tees, R.C. (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development* 7 (1): 49–63.

92 Kolb, B., Forgie, M., Gibb, R. et al. (1998). Age, experience and the changing brain. *Neuroscience & Biobehavioral Reviews* 22 (2): 143–159.

93 Kolb, B. and Gibb, R. (2011). Brain plasticity and behaviour in the developing brain. *Journal of the Canadian Academy of Child and Adolescent Psychiatry* 20 (4): 265.

94 Vida, M.D., Vingilis-Jaremko, L., Butler, B.E. et al. (2012). The reorganized brain: how treatment strategies for stroke and amblyopia can inform our knowledge of plasticity throughout the lifespan. *Developmental Psychobiology* 54 (3): 357–368.

95 Merabet, L.B., Hamilton, R., Schlaug, G. et al. (2008). Rapid and reversible recruitment of early visual cortex for touch. *PLoS One* 3 (8): e3046.

96 Grand, R.L., Mondloch, C.J., Maurer, D., and Brent, H.P. (2004). Impairment in holistic face processing following early visual deprivation. *Psychological Science* 15 (11): 762–768.

97 Lewis, T.L. and Maurer, D. (2009). Effects of early pattern deprivation on visual development. *Optometry and Vision Science* 86 (6): 640–646.

98 Mylopoulos, M. and Regehr, G. (2009). How student models of expertise and innovation impact the development of adaptive expertise in medicine. *Medical Education* 43 (2): 127–132.

99 Mylopoulos, M. and Regehr, G. (2007). Cognitive metaphors of expertise and knowledge: prospects and limitations for medical education. *Medical Education* 41 (12): 1159–1165.

100 Lingard, L. (2009). What we see and don't see when we look at 'competence': notes on a god term. *Advances in Health Sciences Education* 14 (5): 625–628.

101 Tulving, I. and Craik, F.I.M. (2000). *The Oxford Handbook of Memory*. Oxford: Oxford University Press.

102 Simon, H.A. (1974). How big is a chunk? *Science* 183 (4124): 482–488.

103 Rikers, R., Winkel, W.T., Loyens, S., and Schmidt, H. (2003). Clinical case processing by medical experts and subexperts. *The Journal of Psychology* 137: 213–223.

104 Schmidt, H.G. and Rikers, R.M. (2007). How expertise develops in medicine: knowledge encapsulation and illness script formation. *Medical Education* 41: 1133–1139.

105 Charlin, B., Boshuizen, H., Custers, E.J., and Feltovich, P.J. (2007). Scripts and clinical reasoning. *Medical Education* 41: 1178–1184.

106 Lubarsky, S., Dory, V., Audétat, M.C. et al. (2015). Using script theory to cultivate illness script formation and clinical reasoning in health professions education. *Canadian Medical Education Journal* 6: e61–e70.

107 Evans, J.S.B. and Frankish, K. (2009). *In Two Minds: Dual Processes and Beyond*. Oxford: Oxford University Press.

108 Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

109 Croskerry, P., Singhal, G., and Mamede, S. (2013). Cognitive debiasing 1: origins of bias and theory of debiasing. *BMJ Quality and Safety* 22: ii58–ii64.

110 Sibbald, M., McKinney, J., Cavalcanti, R.B. et al. (2013). Cardiac examination and the effect of dual-processing instruction in a cardiopulmonary simulator. *Advances in Health Sciences Education* 18: 497–508.

111 Monteiro, S.D., Sherbino, J., Patel, A. et al. (2015). Reflecting on diagnostic errors: taking a second look is not enough. *Journal of General Internal Medicine* 30: 1270–1274.

112 Norman, G.R., Monteiro, S.D., Sherbino, J. et al. (2017). The causes of errors in clinical reasoning: cognitive biases, knowledge deficits, and dual process thinking. *Academic Medicine* 92: 23–30.

113 Sibbald, M., de Bruin, A.B., and van Merriënboer, J.J. (2013). Checklists improve experts' diagnostic decisions. *Medical Education* 47: 301–308.

114 Sibbald, M., de Bruin, A.B., Cavalcanti, R.B., and van Merriënboer, J.J. (2013). Do you have to re-examine to reconsider your diagnosis? Checklists and cardiac exam. *BMJ Quality and Safety* 22: 333–338.

115 Woolley, A. and Kostopoulou, O. (2013). Clinical intuition in family medicine: more than first impressions. *Annals of Family Medicine* 11: 60–66.

116 Van den Bruel, A., Thompson, M., Buntinx, F. and Mant, D. (2012). Clinicians' gut feeling about serious infections in children: observational study. *BMJ* 25: e6144.

117 De Bruin, A.B.H., Dunlosky, J. and Cavalcanti, R. (2017). Monitoring and regulation of learning in medical education: the need for predictive cues. *Medical Education* 51: 575–584.

118 Burke, J. (1989). *Competency-Based Education and Training*. Abingdon: Routledge.

119 Durning, S.J. and Artino, A.R. (2011). Situativity theory: a perspective on how participants and the environment can interact: AMEE guide no. 52. *Medical Teacher* 33 (3): 188–199.

120 Hodges, B.D. and Kuper, A. (2012). Theory and practice in the design and conduct of graduate medical education. *Academic Medicine* 87 (1): 25–33.

121 Schön, D.A. (1983). *The Reflective Practitioner: How Professionals Think in Action*. New York: Basic Books.

122 Bordage, G. (2009). Conceptual frameworks to illuminate and magnify. *Medical Education* 43 (4): 312–319.

123 Van Merrienboer, J.J.G. and Sweller, J. (2010). Cognitive load theory in health professional education: design principles and strategies. *Medical Education* 44 (1): 85–93.

124 Cutting, M.F. and Saks, N.S. (2012). Twelve tips for utilizing principles of learning to support medical education. *Medical Teacher* 34 (1): 20–24.

125 Gooding, H.C., Mann, K., and Armstrong, E. (2017). Twelve tips for applying the science of learning to health professions education. *Medical Teacher* 39 (1): 26–31.

126 Young, J.Q., Van Merriënboer, J., Durning, S. and Ten Cate, O. (2014). Cognitive load theory: implications for medical education: AMEE guide no. 86. *Medical Teacher* 36 (5): 371–84.

127 Boshuizen, H.P.A. (1996). The shock of practice: Effects on clinical reasoning. Annual Meeting of the American Educational Research Association, New York, 8–14.

128 Koens, F., Mann, K.V., Custers, E.J.M. and Ten Cate, O. (2005). Analysing the concept of context in medical education. *Medical Education* 39 (12): 1243–49.

129 Koens, F., Ten Cate, O. and Custers, E.J.M. (2003). Context-dependent memory in a meaningful environment for medical education: in the classroom and at the bedside. *Advances in Health Sciences Education*. 8 (2): 155–65.
130 Norman, G., Dore, K., and Grierson, L. (2012). The minimal relationship between simulation fidelity and transfer of learning. *Medical Education* 46 (7): 636–647.
131 Lindner, R.W. and Harris, B. (1993). Self-regulated learning: its assessment and instructional implications. *Educational Research Quarterly* 16 (2): 29–37.
132 Ertmer, P.A. and Newby, T.J. (1996). The expert learner: strategic, self-regulated, and reflective. *Instructional Science* 24 (1): 1–24.
133 Ferguson, E., James, D., and Madeley, L. (2002). Factors associated with success in medical school: systematic review of the literature. *BMJ* 324 (7343): 952–957.
134 Gurpinar, E., Alimoglu, M.K., Mamakli, S., and Aktekin, M. (2010). Can learning style predict student satisfaction with different instruction methods and academic achievement in medical education? *Advances in Physiology Educaton* 34 (4): 192–196.
135 Newble, D.I. and Entwistle, N.J. (1986). Learning styles and approaches: implications for medical education. *Medical Education* 20 (3): 162–175.
136 Pashler, H., McDaniel, M., Rohrer, D., and Bjork, R. (2008). Learning styles concepts and evidence. *Psychological Science in the Public Interest* 9 (3): 105–119.

Further Reading

Brown, P.C., Roediger, H.L. III, and McDaniel, M.A. (2014). *Make it Stick: The Science of Successful Learning*. Harvard, Cambridge, MA: Harvard University Press.

Regehr, G. and Norman, G.R. (1996). Issues in cognitive psychology: implications for professional education. *Academic Medicine* 71: 988–1001.

Kulasegaram, K.M., Martimianakis, M.A., Mylopoulos, M. et al. (2013). Cognition before curriculum: rethinking the integration of basic science and clinical learning. *Academic Medicine* 88 (10): 1578–1585.

Norman, G.R., Monteiro, S.D., Sherbino, J. et al. (2017). The causes of errors in clinical reasoning: cognitive biases, knowledge deficits, and dual process thinking. *Academic Medicine* 92 (1): 23–30.

Van Merrienboer, J.J.G. and Sweller, J. (2010). Cognitive load theory in health professional education: design principles and strategies. *Medical Education* 44 : 85–93.

4 Teaching and Learning in Medical Education: How Theory can Inform Practice

David M. Kaufman
Faculty of Education, Simon Fraser University, Burnaby, British Columbia, Canada
Faculty of Health Sciences and Department of Gerontology, Simon Fraser University, Burnaby, British Columbia, Canada

KEY MESSAGES

- Understanding educational theory can enhance curriculum development, teaching, and learning.
- Learners are active contributors in the educational process; they interact with curricula, patients, and teachers in a complex, changing environment.
- The entire context of learning is important, rather than any single variable, and includes interactions of all the variables.
- Values, attitudes, and the culture of the profession are often learned implicitly and without explicit teaching or awareness of learning.
- Learning is enhanced when it is relevant, particularly to the solution and understanding of real-life problems and practice.
- Past experience and knowledge play a critical part in how people learn.
- Learning has a significant emotional aspect to it that is often under-recognised.
- Learners are capable of self-regulation, that is, setting goals, planning strategies, and monitoring their progress.
- The ability to reflect on one's practice (performance) is critical to lifelong, self-directed learning.
- Learning occurs collectively as well as individually as learners construct shared knowledge and understanding through their work together.
- Knowledge creation and dissemination require interaction among many different human and material entities – for example, tools, instruments, technologies, texts, and images.

Introduction

The frequently identified gap between theory and practice has led practitioners in many professions to conclude that theory belongs in an ivory tower, neither useful nor relevant to those in practice. Education is no exception [1]. However, as the processes that underpin educational practice are better understood, it is clear that theory has the potential both to inform practice and to be informed by it. Theories help us move beyond merely understanding a phenomenon to also critiquing and improving it [2, 3]. Bordage [4] reported that in a recent study of the quality of reporting experimental studies in medical education, barely half the articles examined contained an explicit statement of the conceptual framework used. He described conceptual frameworks as ways of thinking about problems or ways of representing how complex things work, which can come from theories, models, or best practices. Bordage explained that conceptual frameworks help educators to understand and illuminate problems, but different conceptual frameworks emphasise different aspects of a situation. Several conceptual frameworks may be relevant to a given situation, and any one of these, or combinations of them, could lead to alternative solutions or approaches. It is interesting to note that many theories of learning overlap, offering different ways of conceptualising and studying the same phenomena, but with different epistemologies, assumptions, and methodological approaches.

Mann [5] has argued that learning theories arising from behaviourist, cognitivist, humanist, and social learning traditions have guided improvements in curriculum design and instruction, in understanding of memory, expertise, and clinical decision-making, and in self-directed learning approaches. She asserts that although these remain useful, additional perspectives are needed that recognise the complexity of education and effectively foster the development of knowledge, skills, and professional identity.

Feldman and Orlikowski [6] have expanded our view of theory by proposing the concept of 'practice theory'. They explain that theoretical generalisations produced through the use of practice theory are not predictions in the conventional sense but may be better understood as principles that can explain and guide action. Particular relationships or enactments (e.g. technologies in practice, resources in use) offer insights for understanding situations while being historically and contextually grounded. Although each context

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

is different, the dynamics and relations that have been identified and theorised can be useful in understanding other contexts. In this way, theoretical generalisations are powerful because they travel [6].

Medical education is a field informed by many theoretical perspectives and disciplines. Cognitive psychology, social psychology, sociology, anthropology, ethics, and economics have been particularly important. Medical and health professions education (HPE) are increasingly recognised as distinct fields influenced by the literature in general education and the social sciences. For example, a US committee recommended that medical students be provided with an integrated behavioural and social science curriculum that extends throughout the four years of medical school [7]. The committee came to several conclusions, including the following one relevant to this chapter:

> Human health and illness are influenced by multiple interacting biological, psychological, social, cultural, behavioral, and economic factors. The behavioral and social sciences have contributed a great deal of research-based knowledge in each of these areas that can inform physicians' approaches to prevention, diagnosis, and patient care [7, p. 8].

Finally, there have been fairly strong influences from the clinical and basic sciences (particularly in terms of the research approach and what counts as evidence), given that medical education is situated in that context. Many excellent masters programmes in medical education are now available and they address educational theory contextualised in medical practice. Until 1996, there were only seven masters-level programmes in HPE; in 2012, there were 76 such programmes [8], and by 2017 the number had increased to 125 [9].

My purpose in this chapter is to describe 10 selected theoretical approaches to education, exploring their implications for the practice of medical education. I use the term 'theory' in a general sense, that is, as a set of assumptions and ideas that help to explain a phenomenon. Knowles put this succinctly more than 40 years ago, defining a theory as: 'a comprehensive, coherent, and internally consistent system of ideas about a set of phenomena' [10, p. 5].

The 10 theoretical approaches discussed in this chapter are:

- social cognitive theory [11]
- reflective practice [12]
- transformative learning [13]
- self-directed learning [14]
- experiential learning [15]
- situated learning [16]
- communities of practice [17]
- constructivism [18]
- sociomateriality [19]
- adult learning principles [20]

I selected these 10 theories because I believe them to be particularly useful in the context of the issues facing medical education today. I will describe each one, highlighting its major constructs, and present implications of the theory for educational practice. I will conclude by considering the connections and commonalities among these theories, so that readers may make these connections within their own practice.

Social Cognitive Theory

Social cognitive theory belongs to the family of social learning theories, which acknowledge that our learning is social in nature: we learn from and in interaction with others and with our environment. Social cognitive theory [11], formerly social learning theory, was developed by Bandura [21] and unites two approaches to our understanding of learning. These are the behaviourist approach, which emphasises the influence of the environment on our actions, and the cognitive approach, which emphasises the importance of cognition in mediating our learning and functioning.

These two approaches are united in a basic tenet of social cognitive theory, which posits that our actions, learning, and functioning are the result of a continuous, dynamic, reciprocal interaction among three sets of determinants: personal, environmental (situational), and behavioural conditions. Personal factors include the individual's attitudes, perceptions, values, goals, knowledge, and all previous experience. Environmental factors encompass all influences that may enable or hinder actions and the achievement of goals. Bandura states explicitly that: 'Personal and environmental influences do not function as independent determinants; rather, they affect each other. People create, alter, and destroy environments. The changes they produce in environmental conditions, in turn, affect them personally' [11, p. 25]. Bandura further states that behaviour, rather than being a 'detached by-product' of persons and situations, is itself an interacting determinant in the process. Figure 4.1 shows these interactions schematically and how they might apply to medical education.

Bandura asserts that the relative influences exerted by each of the three sets of factors will vary for different activities, different individuals, and different circumstances. For example, when environmental conditions exert a powerful influence, they will prevail. In a medical education example, when trainees are thrust into the busy environment of a clinical ward, they will do what is required to 'get the job done' and to meet expectations. In other cases, the behaviour and its feedback will be a major influence. For instance, when students are learning and practising a new skill, the feedback they receive will have a strong influence. Finally, in those instances where situational influences are relatively weak, personal factors will exert the strongest regulatory influence. To complete our example, when not pressed by powerful environmental forces students may choose to learn a new skill or to learn more about talking with patients. These choices will be affected by the student's own values, perceived needs, and individual goals. There may also be interaction within each factor (for example, conflicting values within an individual). The simple example provided here is not intended to convey lack of complexity; rather, it is to emphasise the ongoing, dynamic nature of our interaction with our environment.

Environmental influences can affect people in ways other than their behaviour, as when thoughts and feelings are modified through observing others' behaviour (modelling), or through teaching or social persuasion. Our thoughts do

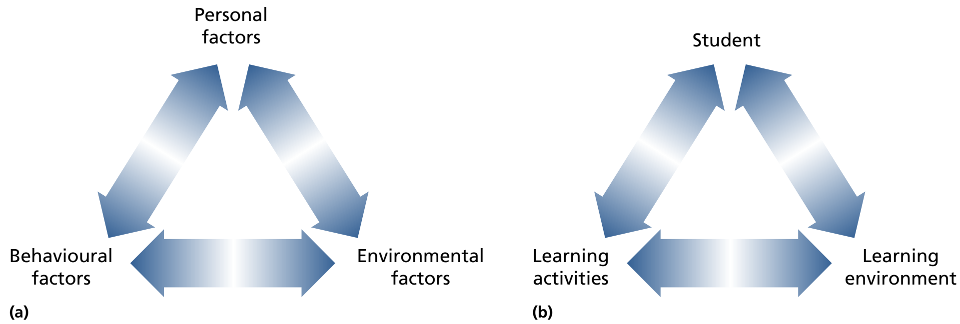


Figure 4.1 Diagrammatic representation of (a) reciprocal interaction among personal, situational, and behavioural factors; (b) the same factors using a medical education example.

not arise in a vacuum. Individual perceptions and understandings are developed and verified through both direct and vicarious experience, through judgements of others, and by inference from what is already known [11, p. 29].

Basic Human Capabilities

Bandura views humans as possessing five basic capabilities that underpin our learning and functioning in all situations. These capabilities are particularly important when we consider the processes of learning in medical and health professional education.

Symbolising Capability

Almost every aspect of our lives is touched by our remarkable ability to use symbols to transform our experience into a form that can be internalised and serve as a guide to future actions. This ability enables us, when confronted with a new problem, to test possible solutions symbolically, rather than laboriously trying out each alternative.

Forethought Capability

Most of our behaviour is regulated by thought. We anticipate the likely outcomes of our actions and plan goals and courses of action to maximise the likelihood of obtaining them. Also, as noted, images of desirable future events, such as achieving our goals, can become motivators of our current behaviour.

Vicarious Capability

If learning occurred only through performing actions and experiencing their effects, learning and development would be slow, tedious, and enormously inefficient. Fortunately, much learning that can be acquired through direct experience can also be acquired or facilitated vicariously through observation of other people's actions and their consequences. This applies to social development, especially where, in some situations, new behaviours can only be conveyed effectively by modelling. Even if learning can occur in other ways, the ability to learn vicariously distinctly shortens the process.

Self-regulatory Capability

In social cognitive theory, the capability for self-regulation is central. Much of our behaviour is regulated primarily by our internal standards and our evaluative reactions to our own actions. Any discrepancies between our actions and those standards activate a self-evaluation, which will influence our subsequent behaviour. Self-evaluation is our personal guidance system for action. We exercise self-regulation or self-directedness by arranging facilitative environmental conditions for ourselves, using our images of future events as guides and creating incentives for our efforts.

Self-reflective Capability

Perhaps the most distinctive capability is self-reflection, whereby we can look critically at our experiences and think about our thought processes. Cognitive theorists refer to this as metacognitive capability. Through self-reflection we gain understanding of ourselves, our behaviour, and the world around us. (Reflection and reflective practice will be addressed later in the chapter.)

Self-efficacy

A central concept in social cognitive theory is *self-efficacy*; the individual's judgement about his or her ability to carry out a specific task or activity and to produce certain attainments. It is not a global perception, rather it is specific to a domain of activity. Self-efficacy beliefs influence the courses of action we pursue, the goals we set and our commitment to them, the level and difficulty of these goals, the effort we invest and how long we persist in the face of obstacles, our resilience in the face of adversity, the life choices we make, and what we can achieve [22]. In 2006, Bandura noted that self-efficacy beliefs affect not just our behaviour, but our goals and aspirations; they also determine what barriers and opportunities we see in the environment [23] (see Box 4.1).

Implications for Educational Practice

Understanding the concepts of ongoing dynamic interactions, basic human capabilities, and how people form perceptions of their abilities allows us to plan a learning environment that is most conducive to maximising each individual's development. We will consider some implications of this theory for effective teaching and learning. In particular, five learning processes that build on basic capabilities can be brought to bear in medical education:

BOX 4.1 FOCUS ON: Self-efficacy [22]

According to Bandura [11], a central type of thought that affects action is people's judgements of their capabilities to deal with different realities, or their self-efficacy. This judgement influences what people choose to do, how much effort they invest in activities, how long they persist in the face of disappointment, and whether tasks are approached anxiously or assuredly. Judgements about our personal efficacy, whether accurate or faulty, arise from four main information sources:

- *Performance attainments* – our own performance is the most influential source of efficacy because it is based on authentic experience of mastery. Successes raise our efficacy appraisals; failures generally have a lowering effect, especially if they occur early in the learning and do not reflect lack of effort or difficult situations. Once strong positive efficacy perceptions are developed, occasional failures do not have a marked effect. Feelings of capability are generally task-specific, though they can generalise to other, similar tasks.
- *Vicarious experience* – observing other similar people perform successfully can raise our own beliefs that we can perform similar tasks. This source of information is particularly effective when people encounter new tasks and have little experience on which to base their perceptions. Learning from role models is an excellent example of vicarious learning.
- *Verbal persuasion* – we have all had the experience of trying to convince people that they possess capabilities that will enable them to achieve what they seek. If the heightened efficacy that the persuasion is attempting to achieve is realistic, it can be influential, particularly in affecting the amount of effort individuals put into a task.
- *Physiological state* – people often judge their capabilities based on their physiological states. We frequently interpret arousal in taxing situations as an ominous sign of vulnerability, and we tend to expect more success when we are not tense and aroused.

- formulation of a clear objective, goal, or desired outcome
- modelling or demonstration
- provision of task-relevant knowledge
- guided practice and feedback
- opportunities for learners to reflect on their learning.

A *clear objective*, goal, or image of the desired outcome enhances learning. It builds on our capability for forethought, providing a guidepost for monitoring and directing our progress appropriately. Awareness of the goal also increases the energy and effort expended and stimulates the development of strategies to reach the goal. Encouraging learners to set their own goals builds on this basic capability.

Modelling or demonstration of the desired process or skill facilitates vicarious learning through observation. This opportunity not only shortens the learning process; it is often essential when new skills are being acquired. Demonstration can help students to form an image of the desired skill or behaviour that can be used as a guide for action and as a standard of performance against which to monitor their personal progress. Finally, a learner's perception of efficacy is increased by observing someone else perform successfully.

Learners require *task-relevant knowledge*. Learners must have the basic building blocks to use as a foundation for newly acquired knowledge and skills. New knowledge, whether related to content or to process, must be relevant to the individual's prior knowledge and skills and to the current learning goal. Further, learners may need stimulation and assistance to activate prior knowledge and to relate it to the new learning. Prior knowledge promotes students' views of themselves as capable of the task. Otherwise, their perceptions of their efficacy are likely to be low, which will affect both developing efficacy perceptions and their future performance.

Guided practice of a new skill with feedback allows learners to develop positive efficacy perceptions about the task and to experience successes rather than failures in the crucial early learning period. Practice promotes the internalisation of personal standards, which can then be used in self-regulation and self-evaluation. Corrective, formative feedback is integral to effective learning. Without feedback, the level of performance achieved is lower. Similarly, feedback is less effective in improving performance when it is not related to a goal or desired level of achievement [23]. A large literature exists about feedback and factors that influence its provision, its acceptance, and its incorporation and use for improvement. Feedback is central to effective self-direction, setting of goals and internal standards, and self-assessment.

Finally, and arguably most critically, learners require *opportunities to reflect* on their learning, to consider their strategies, to determine whether new approaches are required to achieve their goals and to draw lessons for future learning. Reflection also allows the integration of new experiences into existing experience and knowledge. Finally, it allows the learner to build accurate and positive perceptions of efficacy based on their experience.

Understanding that learning occurs through observation (i.e. 'vicariously') has important connections to, and implications for, our practice. This is particularly so when we consider ourselves as role models. The literature continues to support role-modelling as a pervasive means of teaching and a powerful means of learning. Teachers model knowledge, attitudes, behaviours, approaches to problems, applications of knowledge and skill, and interactions with colleagues, learners, other health professionals, patients, and families. Modelling occurs both when we are aware of it and when we are not. Further, the meaning and intent of what learners observe may not always be clear to them. This suggests the importance of being willing to reflect openly in appropriate situations to allow the meaning to be understood. It is this process of 'making the implicit explicit' that promotes learners to reflect on what they have learned and integrate it into their growing knowledge, skill, and developing professional identity. Recognising our roles as models and reflecting on the ways in which we teach through this method can raise our awareness and allow us to be more mindful of ourselves as models.

In summary, social cognitive theory provides us with several important constructs that may inform our educational practice. They include the central concept that learners are constantly interacting with their environment and their actions have consequences. Many of the characteristics that we seek in our learners are present as basic capabilities common to all. Rather than creating these characteristics, learning opportunities can be created to develop and build on them. Finally, we can have some confidence that people are inherently self-directed. Given the appropriate conditions and support, they will set goals, develop strategies to attain them, and monitor their progress regularly.

Reflection and Reflective Practice

The concepts of *reflection* and the *reflective practitioner* are at the centre of the epistemology of professional practice. They borrow from and link three previously well-established epistemologies or world views about the nature of knowledge and how we can know and understand our world: positivism, interpretive theory, and critical theory [12]. The positivistic view of science assumes that theory is a scholarly pursuit that may be unrelated to practice. It is the predictive value of theory that is of practical value. Reflection in professional practice extends this view by proposing that theory and practice inform each other. It is a basic premise of reflection that we can learn from our experience in an ongoing iterative process. As knowledge is embedded in practice, practitioners are positioned to test and revise theories through practice. They do so by reflection and action. The reflective process, as such, serves as a bridge in the theory–practice relationship. Reflection is also related to the interpretivist model, which proposes that theory is interpreted in light of personal current and past experiences. Theory guides or enlightens action and understanding. Lastly, the concept of reflective practice shares with critical theory the observation that theory is intimately linked to practice through a process of critical thinking and examination. This process permits professionals to break free from established paradigms and reformulate the ways in which practice, problems, and problem solving are viewed. This reframing is part of learning and change. It is how practice helps organise theory [12, 24, 25]. Reflective practice then becomes a vehicle for learning effectively.

Several definitions and approaches to reflection and reflective practice are found in the educational literature. Boud, Keogh, and Walker define reflection as 'a generic term for those intellectual and affective activities in which individuals engage to explore their experiences in order to lead to a new understanding and appreciation' [26, p. 19]. With respect to clinical education for medical students, Branch and Paranjape describe reflection as 'consideration of the larger context, the meaning and the implications of an experience or action' [27, p. 1185]. Lastly, Moon describes reflection as 'a basic mental process with a purpose, an outcome, or both, applied in situations in which material is unstructured or uncertain and where there is no obvious solution' [28, p. 10].

Models of Reflective Practice

Donald Schön has arguably been the most influential thinker in understanding reflective practice among professionals. Schön [12, 24] summarises the need for a new scholarship that recognises knowing-in-action, on-the-spot experimentation (reflection-in-action), and action research (reflection-on-action). Schön's writings about reflective practice [12, 24, 25] are based on the study of a range of professions. He argued that formal theoretical knowledge, such as that acquired in the course of professional preparation, is often not useful to the solution of the 'messy, indeterminate' problems of real-life practice. Central to his premise is the need for professional scholarship and the recognition of an epistemology of professional practice. The reflective practitioner incorporates these principles by relating professional knowledge to practical competence and professional activity. By linking theory to practice, each can inform the other.

Professionals develop zones of mastery around areas of competence. They practise within these areas almost automatically. Schön terms this a professional's 'knowing-in-action'. Indeed, practising one's profession has been likened to riding a bicycle. Occasionally the bicycle skids. This occurs in response to a surprise or to the unexpected. Two types of reflection are triggered at this time: 'reflection-in-action' and 'reflection-on-action' [12].

Reflection-in-action occurs during the course of an experience and involves three activities: (i) reframing and reworking the problem from different perspectives; (ii) establishing where the problem fits into learned schema (i.e. already existing knowledge and expertise), and (iii) understanding the elements and implications present in the problem, its solution, and consequences. *Reflection-on-action*, which occurs after an event, is a process of thinking back on what has happened in the situation to determine what may have contributed to the unexpected and how what has been learned from this situation may affect future practice. Both are iterative processes whereby insights and learning from one experience may be incorporated into future 'knowing-in-action' [12, 24].

Other approaches to reflection and learning from experience have also been influential [26–28]. Boud et al. [26] also outline an iterative process comprising three main phases, beginning with the *experience*. The second phase involves *returning to the experience* and, through reflective processes, dealing with both negative and positive feelings about it and re-evaluating it. Boud et al. labelled the last aspect of the process *outcomes*, in which new perspectives on experience can lead to a change in behaviour and a readiness for application and commitment to action. These authors view reflection as the key to learning effectively. They also emphasise the importance of recognising the emotional aspects of experience that accompany effective learning from experience.

Moon [28] views reflection as the catalyst that moves surface learning to deep learning. Deep learning can be integrated with current experience and knowledge, resulting in rich cognitive networks that the individual can draw on in practice. Reflection has been described as a multifactorial approach that can bring a more systematic method to understanding situations and problems of practice [29, 30].

A number of models of reflection have similar characteristics, as follows:

- reflection is described as an iterative process
- levels of reflection are defined, from the superficial to the deep
- deeper reflective levels are generally regarded as more difficult to achieve, although they hold greater potential for learning and growth.

There also appears to be a dynamic relationship between reflective practice and self-assessment, both explicitly and implicitly. The ability to self-assess depends on the ability to reflect accurately on one's practice, and the ability to reflect effectively relies heavily on accurate self-assessment [30, 31].

In the workplace, professionals are known for their ability for on-the-spot experimentation and improvisation, their commitment to ongoing practice-based learning, and their self-directed reflective learning skills. It is these collective skills that permit professionals to continually and subtly learn from practice, adapt to change, and maintain their competence. Professionals' core capabilities are tied to a number of essential skills. Professionals recognise and value the traditional form of knowledge that is gained in school or in study, as well as experiential knowledge that is gained through experience and practice. In the context of their practice, professionals use both of these forms of knowledge to continually reshape their approach to problems, solutions, actions, and outcomes. This creative process, sometimes called wisdom or artistry, occurs in response to new meanings, insights, and perspectives gained through reflection on current and past experiences. It leads to continued learning and ongoing competence within a profession [32].

Reflection has frequently been viewed as an individual professional activity. In some cases, reflecting inadequately or inaccurately on one's performance can result in circular, 'single-loop', learning, which can lead to confirmation of current behaviours rather than to questioning and identifying areas for learning [32]. For this reason, reflection is increasingly suggested as a collective activity whereby individuals can share individual insights and reflections, and increase their collective and individual learning [33, 34]. Collective reflection is also proposed as a vehicle for developing collective norms and values [32]. The growing evidence surrounding reflective practice is summarised in Box 4.2.

BOX 4.2 WHERE'S THE EVIDENCE: Reflective practice

While reflection has been described in several different fields and much has been written about it in the respective literatures, the research literature is relatively early in development. A review [31] of the research across medicine, nursing, and other health professions suggests the following:

- Reflective thinking is seen in practising professionals and in students across a variety of health professions, including nursing, dentistry, medicine, and health sciences.
- Reflection appears to serve a number of purposes. In medicine, it appears to occur most naturally in response to complex and new problems [35]. However, it is also demonstrated in anticipation of challenging situations [36].
- The phenomenon of reflection is not unitary. Several elements and aspects of reflection have been demonstrated. The tendency and ability to reflect vary across individuals and across situations.
- Attempts to measure and classify reflective thinking have resulted in validated instruments, which demonstrate that differences exist and are measurable. Generally, it seems that deeper levels of reflection are achieved less often and are more difficult to achieve.
- It appears that reflective ability can be developed. Strategies associated with reported changes in reflective ability used small-group resources and activities such as portfolio and journal-keeping.
- Several factors appear to constantly influence reflection, both negatively and positively. These include environment, time, maturity, effective guidance and supervision, and the organisational culture.
- Reflective practice appears to be linked to learning, particularly to deep learning, the development of self-regulated learning, and the development of professional identity [28].

Implications for Educational Practice

Reflection and reflective practice have become expected capabilities of practising professionals. This expectation is stated explicitly in goal statements and definitions of competence. For example, Epstein and Hundert define competence as 'the habitual and judicious use of communication, knowledge, technical skills, clinical reasoning, emotions, values and reflection in daily practice for the benefit of the individual and community being served' [37, p. 226]. In their definition, reflection becomes a 'habit of mind'.

Reflective practitioners are able to assess a situation from the perspectives of both theoretical background and practical experience. They must be able to successfully bridge the theory–practice gap and apply both aspects of learning, while examining a situation from all perspectives. They must also be able to use their reflective skills to review their practice critically and to inform their self-assessment based on the feedback they receive. Reflecting in practice is a learned skill of critical thinking and situation analysis. Deliberate critical reflection on practice may stimulate a new way of thinking about one's practice and lead to the development of adaptive expertise [38].

Some individuals may be more oriented to reflection than others. Nonetheless, reflection, practice-based learning, and taking appropriate action are all skills to be learned and applied, and opportunities to acquire them must be made available. The skills required for reflection can be developed in professional courses within undergraduate, graduate, and continuing medical education. Initially, the mentor or teacher models, shares, and demonstrates the skills. He or she facilitates learners' abilities to perceive options and alternatives and to frame and reframe problems. He or she also assists learners in reflecting on the actions and options they have chosen and on the knowledge and values that may have influenced their choices. Finally, teachers help learners to consider critically what they have learned and integrate it into their existing knowledge.

Once a learner has gained sufficient experience and insight into the profession, the teacher's role becomes one of facilitating systematic experiential learning, on-the-spot experimentation, and reflection. Teachers observe, provide feedback, and help to make explicit those situations in which the learner's reframing has occurred [39]. This helps the learner to become consciously aware of the reflection process.

However, reflection and how it can enhance learning may not be clear to all learners. Modelling the process becomes very important. This is a challenge for clinicians, as reflection may be a tacit process for many practitioners. Teaching reflection requires making the implicit process explicit. Faculty development programmes developed to enhance the teaching of humanistic skills report success in helping faculty use reflection on their own experience as a source of learning [33, 40]. Also, studies of distinguished clinician teachers reveal that they use reflection deliberately, both to improve their own practice and to foster it in their learners [33].

Several authors have explored how reflection may be taught and incorporated into practice. Slotnick [34] linked Schön's work to how physicians learn in practice. He emphasised the importance of thinking while solving problems (reflection-in-action) and thinking after problem solving (reflection-on-action). These two activities are required for clinicians to gain new insights and perspectives around practice-based problems, problem solving, and practice itself. Slotnick [34] also outlined related principles and implications for learners and teachers in practice. Shapiro and Talbot [41] applied the reflective practice model to family medicine. They proposed that open learning environments encourage a continual reshaping of practice-based learning along with the development of continuous competence. Lockyer et al. [33] explored how reflection could be used in both classrooms and practice to enhance the integration of knowledge and its translation into professional practice.

Other authors have addressed reflective practice in specific areas of teaching, learning, and curriculum. Clift et al. [42] analysed issues and programmes that encourage reflective practice in education. Palmer et al. [36] addressed curriculum design issues specific to professional education and reflection in nursing and described roles for lecturer-practitioners, mentors (coaches), and mentees. Atkins and Murphy [30] conducted a review and identified five skills as essential for reflection: self-awareness, description, critical analysis, synthesis, and evaluation. Through interviews with clinical teachers, Crandall [35] demonstrated that stages of Schön's model occur during effective clinical learning events and offered strategies for using the model to implement reflective practice across the medical education continuum.

Westberg and Jason [43] offer practical approaches for fostering reflection in medical education, before, during, and following experience. They emphasise the importance of the learning environment for effectively supporting reflection. Lastly, Moon [44] proposes a process of reflection to promote transfer of new learning to practice.

Several authors [43–48] have linked reflective practice to adult learning theory, deep approaches to learning, professional identity development, and self-directed learning. It appears that reflection may be most useful when it is seen as a strategy to enhance learning. Reflection can help learners to integrate new learning into their existing experience from the beginning of their professional education and throughout their practice. However, learners may require a structure to support them as they acquire reflective skills. To foster the development of these skills, learners may need feedback on both the content and process of their reflection. Guidance and supervision are critical to this process. The literature documents many approaches to incorporating reflection and reflective learning into professional curricula. These include various reflective exercises, reflective writing, and portfolio keeping.

There are challenges in the assessment of learning from reflection. Validated scales have been developed to measure and assess learners' reflection. Two that have been validated with medical students are the Self-Reflection and Insight Scale (SRIS) [49] and the Reflection in Learning Scale [50]. These scales can be useful for both learners and teachers in understanding students' readiness for and use of reflection in their learning and its development over time. Palmer et al. [36] provided specific guidance for assessing reflective learning and Wald et al. [51] developed and validated a rubric for fostering and evaluating reflective capacity in medical learners. Assessment of reflection raises the tension between public and private reflection, which students perceive as a challenge [51].

With growing evidence to support the importance of reflection, there has been a notable effort to incorporate more reflective activities into all levels of medical education [52, 53]. However, this challenges us to select strategies that will both facilitate active development of reflective capacity and be relevant to learning and practice. A further challenge in the professional context involves helping learners to appreciate the relevance of these activities to their development as competent professionals. A supportive learning environment is essential in order to value and support critical reflection.

Transformative Learning

Mezirow's concept of transformative learning has developed over 30 years into a comprehensive and complex theory [13, 45, 48]. Transformative learning theory defines learning as the social process of constructing and internalising a new or revised interpretation of the meaning of one's experience as a guide to action. In other words, transformative learning involves elaborating, creating, and transforming meaning schemes (beliefs, feelings, interpretations, and decisions) through reflection on their content, the process by which the schemes were learned, and their premises (social context, history, and consequences) [48]. Transformative learning can be contrasted with conventional learning, which simply elaborates the learner's existing paradigm, systems of thinking, feeling, or doing, relative to the topic. Although learning is increased, the learner's fundamental structure is maintained. Transformative learning changes the learner's paradigm so radically that, although it may retain the old perspective, it is actually a new creation. Critical reflection and rational discourse are the primary processes used in learning. The core of transformative learning in Mezirow's [48] view is the uncovering of distorted assumptions or errors in learning.

Learner empowerment is both a goal and a condition for transformative learning. An empowered learner is able to participate fully and freely in critical discourse and the resulting action. This requires freedom and equality, as well as the ability to assess evidence and engage in critical reflection [45]. Reflection is a key concept in transformative learning theory. Mezirow [48] defines reflection as the process of critically assessing the content, process, or premises of our efforts to interpret and give meaning to an experience. He distinguishes among three types of reflection:

- *content* reflection – examination of the content or description of a problem,
- *process* reflection – examination of the problem-solving strategies being used, and
- *premise* reflection – questioning the problem itself, which may lead to a transformation of belief systems.

Perspective transformation may be the result of a major event in one's life, or the cumulative result of related transformations in concepts, beliefs, judgements, or feelings. The most significant learning involves critical reflection around premises about oneself. This kind of learning is triggered by a disorienting dilemma that invokes self-examination and a critical assessment of assumptions. Through a process of exploring options for new roles, relationships, and actions, new knowledge and skills are acquired. This leads to planning and implementing a new course of action, provisionally trying new roles, renegotiating relationships and forming new ones, and building competence and self-confidence.

Mezirow [13] explains that discourse is a crucial process, referring to a special kind of dialogue in which the focus is on content and on attempting to justify beliefs by giving and defending reasons and by examining the evidence for and against competing viewpoints.

Transformative learning is a complicated, emotional process requiring significant knowledge and skill to implement effectively [54]. A new paradigm emerges only after the old one becomes dysfunctional, and it is the task of the transformative educator to challenge the learner's current perspective. A paradigm shift will occur only if the learner perceives the existing paradigm to be significantly inadequate in explaining his or her experience. However, the new paradigm appears only after a period of disorientation during which no clear paradigm remains. It is typical for the learner to resist letting go of the old paradigm and beginning the transition to the new one. During this process, the teacher–learner relationship may intensify enormously because the learner may begin to resent the teacher or feel anger towards him or her. Often learners feel a complex love–hate for the teacher who intentionally assists in the collapse of their existing paradigm.

Successful transformative learning questions assumptions (this is a key to the process), provides support from others in a safe environment, presents challenge, examines alternative perspectives, and provides feedback. New assumptions are tested in the authentic settings or in discussion with others.

Alternative conceptions of transformative learning theory refer to similar ideas and address factors often overlooked in the dominant theory of transformation (Mezirow's), such as the role of spirituality, positionality, emancipatory learning, and neurobiology. The exciting part of this diversity of theoretical perspectives is that it has the potential to offer a more diverse interpretation of transformative learning and to have significant implications for practice [55].

One new perspective is a distinctive neurobiologically based pathway to transformative learning. From this perspective, learning is seen as 'volitional, curiosity-based, discovery-driven, and mentor-assisted' and most effective at higher cognitive levels [56, p. 144]. Furthermore, a neurobiological approach suggests that transformative learning: (i) requires discomfort prior to discovery; (ii) is rooted in students' experiences, needs, and interests; (iii) is strengthened by emotive, sensory, and kinaesthetic experiences; (iv) appreciates differences in learning between males and females; and (v) demands that educators acquire an understanding of a unique discourse and knowledge base of neurobiological systems. Other perspectives have been described by Taylor [57] and are appropriate for application in a variety of contexts.

Transformative learning theory continues to be a growing area of study in adult learning and has important implications for the practice of teaching adults. This growth is so significant that it seems to have replaced andragogy as the dominant educational philosophy of adult education, offering teaching practices grounded in empirical research. Taylor [55, 57] reminds educators that the body of research and alternative perspectives implies that fostering transformative learning is much more than implementing a series of instructional strategies with adult learners. Transformative learning is first and foremost about educating from a particular world view, i.e. a particular educational philosophy. It is not an easy way to teach. It means asking oneself, 'Am I willing to transform in the process of helping my students transform?' Without developing a deeper awareness of our own frames of reference and how they shape practice, there is little likelihood that we can foster change in others.

Patricia Cranton, author of several books on the application of transformative learning, argues that transformative learning requires that students hear and question alternative viewpoints; critically analyse their own assumptions, beliefs, and values; and as a consequence shift their own perspectives. Readings, videos, field experiences, simulations, and role plays can help. Critical reflection and questioning are key, and the educator must create an environment in which these are possible, encouraged, and supported [54].

Implications for Educational Practice

How can educators promote and support transformative learning? First, as educators, we need to take a reformist perspective, rather than a subject-centred or consumer-oriented perspective [58]. In a subject-centred perspective, the educator is the expert authority figure and designer of instruction. In a learner- or consumer-oriented perspective, the educator is a facilitator and resource person. In a reformist perspective, essential to transformative learning, the educator is a co-learner and provocateur working to challenge, stimulate, and provoke critical thinking [54].

Cranton [54] provides the following guidelines for transformative educators:

- Promote rational discourse, a fundamental component of transformative learning and part of the process of empowering learners.
- Promote equal participation in discourse by stimulating discussion through a provocative incident or controversial statement.
- Develop discourse procedures (e.g. stay on topic, summarise) and avoid using one's own position to make dismissive statements.

Box 4.3 illustrates the stages that learners pass through during the transformative process. A learner begins in their comfort zone with freedom to decide whether they wish to participate. After choosing to participate, the learner begins to question their assumptions and these are challenged through dialogue with others. This process leads to increased consciousness by the learner about their beliefs. With the support of the educator and the peer group, the learner revises their assumptions and may take action (if required) based on their new world view. This empowers the learner and leads to increased autonomy and the willingness and ability to do critical self-reflection.

There are a number of actions that the educator can take to support the transformative learning process. To be successful here, a climate of openness and supportiveness needs to be established. The educator should:

- Develop group facilitation skills (e.g. handle dominant or silent participants).
- Encourage learners' engagement in decision-making by making the process open and explicit.
- Encourage critical self-reflection by challenging learners, asking critical questions, and proposing discrepancies between learners' experiences and new or conflicting information.

BOX 4.3 FOCUS ON: Stages of change in transformative learning [54]

Stage of change	What happens to the learner	Example
Initial learner development	Freedom to participate. Learner is in comfort zone	Medical student is doing a routine physical exam on a heavy smoker
	Learner makes a decision to confront a belief	Student decides to confront the patient at the next appointment about the health risks of smoking
Learner critical self-reflection	Questioning assumptions	Student asks herself whether she can convince the patient to stop smoking
	Consciousness-raising	Student reads about change models and speaks to her mentor and others in her team about applying the model
	Challenging assumptions	Medical student suspends current beliefs and learns how to apply the model
Transformative learning	Revision of assumptions	Student learns a new process for creating change with high-risk patients
	Support from the educator	Preceptor and peers provide guidance
	Learner networks created	Student presents her case at seminars or Grand Rounds. Student asks for feedback and invites others to learn to apply the model.
	Action taken (if appropriate)	Student applies the model effectively and learns that it can work
Increased empowerment	Critical self-reflection	Student keeps a journal that describes what did and did not work
	Transformative learning	Student continues applying the model to different situations with support from her preceptor
	Development of autonomy	Student is able to apply the model autonomously in various situations

- Consider individual differences among learners. Learners should be assisted in becoming more aware of their own learning styles and preferences. The educator needs to develop a strong awareness of how learners vary in the way they think, act, feel, and see possibilities.
- Employ various teaching/learning strategies. Many strategies are effective, for example: role playing (with skilful debriefing), simulations and games, life histories or biographies, exposure to new knowledge, journal writing (with feedback from self or others), and critical incidents arising in the practice setting.

Self-directed Learning

Self-directed, lifelong learning (SDL) is increasingly essential in the development and maintenance of professional competence. It is integral to the process of self-regulation. Those responsible for professional education, including that of physicians, are challenged to create curricula that ensure the development of these skills as well as the assessment methods needed to ascertain their achievement.

The literature on SDL has developed along two overlapping pathways. The first has framed self-direction as a goal towards which individuals strive, reflecting a humanistic orientation such as that described by Maslow [59] and Brockett and Hiemstra [60]. These models imply achievement of a level of self-actualisation, along with the acceptance of personal responsibility for learning, personal autonomy, and individual choice.

The second line of development has framed SDL as a method of organising learning and instruction, with the tasks of learning left primarily in the learner's control. Early development included linear models, where learners moved through a series of steps to reach their learning goals (e.g. Knowles [61]). Later models have described the self-directed learning process as more interactive, involving opportunities in the environment, the personal characteristics of learners, cognitive processes, the context of learning, and opportunities to validate and confirm self-directed learning collaboratively. Examples of this are seen in several models clearly described by Merriam and Caffarella [62]. This line of development also includes models of instruction such as those of Grow [63] and Hammond and Collins [64] that present frameworks for integrating self-directed learning into formal educational settings.

Candy [14] clarified the field of SDL significantly, bringing educators closer to understanding the specific characteristics to be identified, developed, and evaluated in the self-directed learner. He identified approximately 100 traits associated with self-direction, clustered around four dimensions:

- personal autonomy
- self-management in learning
- learner control of instruction
- the independent pursuit of learning.

Although these characteristics were identified in 1991, they are still relevant today.

Self-directed learning is an integral aspect of several theoretical approaches to learning, including cognitive, social, humanist, and constructivist. As noted earlier in the chapter, the social learning approach views individuals as inherently self-regulating, with self-direction as a natural activity. The humanist approach views self-direction as providing evidence of higher levels of individual development. The cognitive perspective recognises the need to build rich, interconnecting knowledge structures, based on existing knowledge, which allow continuing incorporation of new learning. The constructivist perspective recognises the unique personal and social construction of knowledge that occurs in different learners. Self-directed learning elements can also be seen in the ability to learn from experience through critical reflection, which allows learners to identify their personal learning needs and to be aware of, monitor, and direct the growth of their knowledge, skills, and expertise [61].

Generally, self-direction is a natural human process that can occur both within and/or outside of formal settings. SDL does not exclude formal activities such as lectures or courses. The learner's choice of activities to meet and manage a particular learning goal denotes self-direction.

A number of factors in the learner and in the environment will affect the learner's ability to be self-directing:

- The learner's view of him- or herself as a learner is an influencing factor. Learners who view themselves as competent, with the skills to learn in a variety of situations, are more likely to be self-directed and independent.
- Sometimes the demands of the learning situation influence the capacity for self-direction. Where the situation demands that certain (particular) knowledge and skills are non-negotiable, or where the situation requires the learner to reproduce exactly what has been taught, the capacity for self-direction may be obscured.
- Self-direction is, to some degree, a function of subject matter mastery. As the learner builds a base of relevant knowledge and skills, the capacity to be self-directed is enhanced. This basic knowledge is held by some to be essential for effective SDL. Others who promote learning based on activation of prior knowledge tell us that there are few learning situations where the learner is completely lacking in relevant knowledge to engage a learning task. Part of enhancing self-direction is helping learners to identify their relevant knowledge and experience.
- Much of professional learning is situated learning; that is, the learning is inseparable from the situation in which the knowledge is used. Similarly, professional knowledge and acumen become embedded in practice and form part of the professional's 'knowing-in-action' [47]. Learners may require help in understanding the way knowledge is structured and used, in order to understand fully the range of learning opportunities available to them. They also benefit from opportunities to participate in their community of practice and the knowledge embedded in it [16, 17, 33].
- Knowledge is also socially constructed, in that it is built from mutually understood perceptions and assumptions.

Learners' participation in the social construction of knowledge through discussion and participation provides a cultural basis for their self-direction.
- Knowledge is dependent on context for its meaning, its structure in memory, and its availability. Understanding and experience of a broad range of discipline-relevant contexts encourage self-direction in transferring knowledge to other appropriate contexts.

There are few comprehensive measures of self-directedness [64]. Three scales have been used sufficiently to provide validity evidence. The Self-Directed Learning Readiness Scale (SDLRS) was developed by Gugliemino [65] as a tool to assess the degree to which people perceive themselves as possessing the skills and attitudes conventionally associated with SDL. The Oddi [66] Continuing Learning Inventory purports to identify clusters of personality characteristics that relate to initiative and persistence in learning over time through a variety of learning modes. The SRIS, developed by Roberts and Stark [67], explores reflection as an activity that is basic for making self-directed change, thus uniting these two important elements of self-regulation.

The ability to self-assess is critical to effective self-directed learning. To properly direct one's ongoing learning and to assess where and what learning is required, the individual must be able to assess his or her current practice with reasonable accuracy. A recent review of the self-assessment literature suggests that our current understanding of self-assessment is insufficient and that our ability to assess our own performance accurately is limited. Eva and Regehr [68] suggest that accurate self-assessment requires knowledge of what constitutes appropriate performance and knowledge of the criteria by which to judge it. They further suggest that several sources of information may be necessary for accurate self-assessment, including feedback from others about one's performance. It is also important to better understand the cognitive, affective, and psychomotor bases of self-assessment to effectively promote the development of self-assessment capacity. Several authors have explored self-assessment further and the processes and conditions that influence it [69].

Self-regulation

More recently, there has been interest in the related concept of self-regulation in medical education [70]. Self-regulated learning (SRL) has been defined as: 'self-generated thoughts, feelings and actions that are planned and cyclically adapted to the attainment of personal goals' [70, p. 72]. Self-regulation theory could be considered as a modern extension of self-directed learning theory and appears to be a promising approach for medical educators. In their AMEE Guide (Association for Medical Education in Europe), Sandars and Cleary [71] focus on three important characteristics of self-regulated learners that are shared across the various theories: (i) goal-directed behaviour, (ii) use of specific strategies to attain goals, and (iii) the adaptation and modification of one's behaviours or strategies to optimise learning. They assert that most theorists conceive of self-regulation as a superordinate process characterised by various sub-processes, such as goal-setting, planning, strategy use, self-control, self-monitoring, and self-reflection. The authors propose a cyclical approach that bears a resemblance to the cyclical process of self-directed learning discussed in this chapter. The three stages they propose are: before (forethought), during (performance), and after (self-reflection).

Sandars and Cleary [71] assert that medical education could be enhanced by infusing self-regulation principles into curriculum delivery. They suggest that specific training to develop self-regulation processes can improve complex psychomotor-skilled performance. They make several suggestions for practice of self-regulation. First, educators could act as role models and verbalise their use of key self-regulation processes during their academic or clinical performance. Second, educators could provide strategic and process feedback to learners with a focus on strategy use and self-regulation processes. Feedback could be generated from context-specific assessment techniques, such as think-aloud or self-regulation microanalysis. The authors also propose supporting learners by implementing separate 'learn to learn' courses or using a peer mentoring approach, such as reciprocal teaching, in which learners mentor each other using a structured approach to understanding text.

Sandars and Cleary raise a key issue that needs to be addressed in self-directed or self-regulated learning in clinical education [71]. Research has consistently shown that many individuals, including medical students and doctors, are inaccurate in their own judgements about their knowledge, skills, and performance [72, 73]. Most learners tend to overestimate their skills, which unfortunately can have a negative effect on their selection and use of strategies to achieve a task. Learners who struggle the most are often those who have the greatest discrepancy between perceived competence and their actual performance. In addition, these learners often do not take corrective measures on the occasions when they do correctly self-assess poor performance. To engage in effective adaptation, these learners either need to be able to generate informative internal feedback or be given external feedback by supervisors or teachers. In short, feedback is an essential and integral component of the self-regulation of learning, since it provides an opportunity for the learner to make adaptive changes to their use of key self-regulation processes [74]. External feedback, such as comments made by others about the skills or task performance of a student, can be a powerful reality check that highlights the discrepancy between perceived and actual performance. This feedback can be obtained from teachers, tutors, and peers. When provided effectively, external feedback can be a very important component of self-regulation because it directs a learner's attention to the essential requirements of a task or the behaviours or processes needed to adapt or correct mistakes.

Motivation

The term 'motivation' appears throughout this chapter, as it is a core concept for learning.

Cook and Artino [75] recently argued that previous reviews of motivation in HPE focused on practical implications, broad overviews, or only one theory. They summarised five contemporary theories about motivation to learn, articulated key intersections and distinctions among theories, and proposed important considerations for future research. Their article articulates the complexity of the concept of motivation. They argue that motivation is far from a unitary concept and that more research and theoretical work still needs to be done to capitalise on the potential benefits of motivation for learning in HPE. Despite these conclusions, the authors have provided some helpful guidelines.

On a more practical level, Keller [76, 77] proposed a model to enhance motivation that he called the 'ARC model'. This model proposes four factors: Attention, Relevance, Confidence and Satisfaction (ARCS). Keller proposes that attention can be gained in two ways: (i) by using novel, surprising, incongruous, and uncertain events to gain interest; or (ii) by stimulating curiosity through presenting challenging questions or problems to be solved. For relevance, he suggests the use of concrete language and examples with which the learners are familiar. In order to enhance confidence, Keller suggests that teachers should help learners understand their likelihood for success. If learners believe they are not able to attain their objectives or that the time or effort required is too high, their motivation will decrease. Ensuring that learners are aware of performance requirements and evaluative criteria can help support their motivation. Finally, learning must be satisfying in some way, whether it is from a sense of achievement (intrinsic motivation), reward or praise from a teacher (extrinsic motivation), or simply being entertaining or enjoyable. The learner should be shown that the skills or knowledge to be learned are useful by being provided with opportunities to use them in a practice setting.

Implications for Educational Practice

Self-directed learning, self-regulated learning, and motivation theories and research have a number of important implications for curricula, teaching, and learning in medical education, all of which are facilitated by the creation of a supportive learning environment where learners feel that it is safe to ask questions and admit to not understanding. Learners must have the opportunity to develop and practice skills that directly enhance effective SDL. These include competency at asking questions, seeking relevant information, and critically appraising new information.

Learners also need to acquire multiple approaches to learning, along with the ability to decide when each is appropriate. For ongoing SDL, however, *deep learning skills* [78], which involve understanding principles and concepts and elaborating the relationships among them, are most likely to support self-direction. Making use of learners' existing knowledge structures, and assisting them to add to and enrich those structures and understand similarities and dissimilarities, will encourage individuals to understand their knowledge base and identify gaps. A fundamental skill in self-direction is that of critical reflection on one's own learning and experience. Learners must practise and develop skills at reflecting on all aspects of their learning to determine additional learning needs and to set goals accordingly. In an illustration, Miflin et al. [79] describe an unsuccessful attempt to introduce SDL into graduate medical education in a university in Australia in which lack of clarity among teachers and learners of what constitutes 'self-direction' forced a reconsideration of the curriculum.

Critical to the achievement of both explicit and implicit curriculum goals are congruence and alignment among the goals, the educational strategies, and the assessment methods [80, 81]. Assessment will invariably drive learning and give the strongest messages to learners about the real goals of the curriculum. Although there are genuine attempts to do otherwise, too frequently assessment methods reward teacher-directed, fact-oriented learning and do not reward or evaluate the learner's achievement of self-directed learning.

An elegant model for the use of self-directed learning by medical practitioners was developed in Canada more than 20 years ago and used by the Royal College of Physicians and Surgeons of Canada [82]. This was the first formal application of self-directed learning for re-certification and maintenance of competence by physicians. In this programme, physicians who used the new PCDiary software reported that it helped them to review and appraise their learning activities. Inspired by this project, a number of medical specialties since then have used self-directed learning as a method for continuing professional development.

Smith et al. [83] report on a curriculum that used self-directed learning plans based on clinical questions arising from internal medicine residents' practice. To teach residents self-directed learning skills, they implemented an ongoing curriculum integrated with their clinical practice. Residents recorded one clinical question monthly that formed the basis of a structured exercise. They documented the patient encounter triggering the question, described the resources used and the answer found, and reflected on its effect on patient management. Residents discussed their self-directed learning plans monthly at a pre-clinic conference where only their postgraduate year cohorts were present.

Another project in internal medicine developed and implemented a successful four-week curriculum to teach physicians self-directed learning skills during inpatient ward rotations [84]. The educational methods employed to teach these self-directed learning skills included individual study (e.g. physicians read individually on topics related to patients on the service), group study (e.g. the team performed the learning-resource exercise together), and the use of attending physicians as role models. In addition to these methods, the curriculum provided an organisational structure for the month's experience on the ward, as well as administrative tools that facilitated review and evaluation of learning experiences (e.g. the learning plan). A learning diary served as a record of all patients admitted during the month, eliminating the need for other redundant record-keeping methods that had historically been used.

Many forms of self-regulated learning have been suggested [70]. Examples include facilitating, prompting, modelling, and explaining through various resources, such as text, video, online modules, peers, and instructors. Since the learner is not expected to learn alone, self-regulation can be enhanced through the use of various levels and types of support. Research shows that trainees need help with learning how to enter a new 'culture' and how to understand the rules of engagement in all facets of medical education. To that end, supports can be built into important transition points in the medical education continuum, e.g. when PBL starts, in clinical clerkships, and when transitioning from fellow to staff member. Box 4.4 describes a process for doing self-directed learning, which can be viewed as a dynamic wheel (Figure 4.2).

Experiential Learning

Kolb's experiential learning theory [15] is derived from the work of Kurt Lewin [85], John Dewey [86], and Jean Piaget [87]. Lewin's [85] work in social psychology, group dynamics, and action research concluded that learning is best achieved in an environment that considers both concrete experiences and conceptual models. Dewey [86] constructed guidelines for programmes of experiential learning in higher education. He noted the necessity of integrating the processes of actual experience and education in learning. Piaget's [87] research regarding cognitive development processes constituted the theory of how experience is used to model intelligence. Abstract thinking, including the use of symbols, is closely linked to learners' adaptation to their environment. Fenwick [88] offered a summary of five contemporary perspectives on experiential learning – constructivist, psychoanalytic, situative, critical/cultural, and enactivist – that have emerged in recent scholarly writing addressing experiential learning and cognition. She compared these five currents along the following eight

BOX 4.4 HOW TO: Self-directed learning

For learners: A seven-stage self-directed learning process

Self-directed learning can be described as a process that involves seven overlapping and interlocking stages. Working through the stages, you need to:

1. **Create a vision.** What is your vision of success? What is the most important activity that you could undertake to move you towards your ultimate goal?
2. **Set a clear goal and objectives.** What is your goal? Is it reasonable, feasible? What are the specific objectives that you will need to achieve on the way to your goal?
3. **Identify resources and supports.** What are your strengths? What will help you to achieve your goal? Who can help you? (Success often depends on who and what resources are available to you, and how well you are able to engage the needed resources. In fact, interdependence is often the key to success, rather than independence.)
4. **Develop a step-by-step plan.** What steps are you going to take to achieve your goal? Describe what you are going to do and outline a timeframe.
5. **Implement your plan.** Most people have wonderful ideas about what they could, or might, or would like to do. But ideas are no more than dreams until you write them down and act.
6. **Evaluate.** What would success be for you? What result would be satisfactory and what would be excellent? How will you know? Who will give you feedback?
7. **Celebrate.** Give yourself a reward. Share your success with others. We all need recognition – we don't always get the recognition from others that we want or deserve. So, we may need to take time to celebrate a successful achievement.

Now that you've successfully achieved your learning goals, you may find that you want to go deeper into the subject, explore a related topic, or embark on learning something completely different. As our seven-step model indicates, you are back at the start of the process and can tackle new learning goals from your foundation of self-directed learning experience and success.

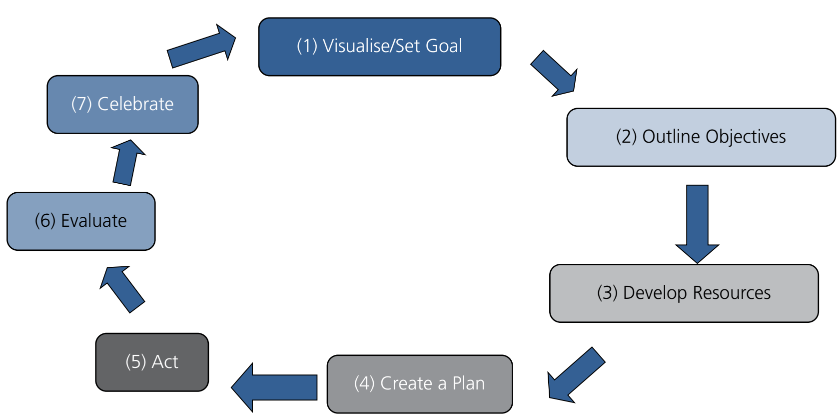


Figure 4.2 Seven-step self-directed learning model.

dimensions: focus, basic explanatory schemata, view of knowledge, view of relation of knower to object and situation of knowing, view of learning process, view of learning goals and outcomes, view of the nature of power in experience and knowing, and view of the educator's role, if any, in learning.

Kolb's experiential learning theory is useful as a model of learning from an applied perspective. It can be used as a framework for interpreting and diagnosing individual learners, as well as for designing learning environments [89]. Kolb's four learning environments are:
- affectively oriented (feeling)
- symbolically oriented (thinking)
- perceptually oriented (watching)
- behaviourally oriented (doing) [90].

Within these environments, grasping and transforming experiences are the two constituent activities of learning tasks [91]. Grasping the phenomena has two components: concrete experience, which filters directly through the senses, and abstract conceptualisation, which is indirect and symbolic. The transforming experience also consists of two processes: reflection and action. One or a combination of the four activities (concrete experience, abstract conceptualisation, reflection, and action) may be used in learning [15]. Learning is enhanced if students are encouraged to use all four components (see Figure 4.3).

Implications for Educational Practice

This section explores Kolb's learning environments in more depth by presenting practical implications for planners of educational programmes, teachers, and learners. Educational formats for delivering experiential learning activities are also included.

Programme Planning

Kolb provides three major guideposts for directing experiential instructional activities [92]. First, experiential learning methods and procedures are bridges connecting a learner's existing level of understanding, philosophies, affective characteristics, and experiences with a new set of knowledge, abilities, beliefs, and values. Second, in experiential learning the learner adopts a more assertive role in assuming responsibility for his or her own learning. This leads to a shift in the power structure from the traditional relationship between teacher and learner. Last, experiential learning involves the transfer of learning from an academic mode to one that involves more practical content.

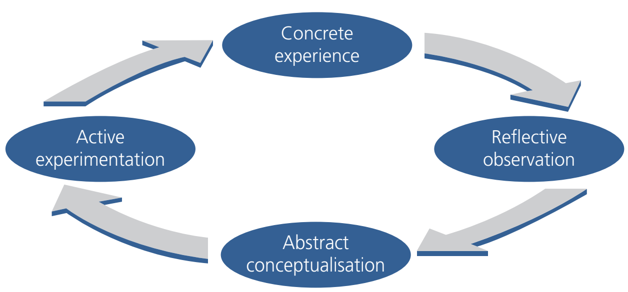


Figure 4.3 Kolb learning cycle.

More specifically:

In the affectively oriented environment, learners experience activities as though they were professional practitioners [15, 91]. The learner's existing values and experience generate information.

In the symbolically oriented environment, the learner uses experiences to develop skills or concepts that can provide the right answer or the best solution to a problem [15, 91]. The source of information is primarily conceptual.

In the perceptually oriented environment, the learner views concepts and relationships from different perspectives such as watching, thinking, and feeling [15].

Behaviourally oriented activities focus on specific problems or practices to which learners apply their competencies [15, 92].

Teaching

The roles and actions of teachers depend on the particular learning context [15, 93].

In the affectively oriented environment, teachers act as role models and relate to the learner as friendly advisors. They deliver information quickly and tailor it to the needs and objectives of individual learners. Teachers monitor progress by encouraging ongoing discussion and critique without constricting guidelines to inhibit students.

In the symbolically oriented environment, teachers act as content experts and facilitators in order for the learners to reach a solution or a goal [15, 93]. Success is compared against the correct or best solution by objective criteria. Teachers provide guidelines regarding terminology and rules.

In the perceptually oriented environment, teachers act as process facilitators, emphasising process rather than solution. They also direct and outline connections between discussions. Learners evaluate answers and define concepts individually. Performance is not measured against rigid criteria but by how well learners use predetermined professional criteria.

In the behaviourally oriented environment, teachers act as mentors and reflect on their background when giving counsel. There are few guidelines. Learners manage their own time and focus on 'doing' [15, 93]. It is essential that the learner complete the task using professional standards.

Learning

In the affectively oriented environment, the learner must work with people, be receptive to encompassing values and feelings, and become engaged in a learning group that is engaged in a concrete experience.

In the symbolically oriented environment, learners study quantitative data to test their theories and postulations [91, 93]. Using unique ideas and action plans, learners develop and conceptualise their experiences and models. This relates to the experience of abstract conceptualisation.

The perceptually oriented environment encourages the learner to analyse and manage data with an open mind [91, 93]. The learner must learn to see things with a broad point of view, compose complete plans of action, and conjecture about the implications of ambiguous circumstances. The learner undergoes the transformative experience of reflective observation by openly approaching the learning activity.

In the behaviourally oriented environment, learners must make their own choices in order to locate and exploit potential opportunities, committing themselves to meet predetermined goals and objectives. They are encouraged to adapt to uncertainty and shifting circumstances, and to guide others. This relates to the transforming experience of active experimentation.

Caffarella [92] describes a number of formats for experiential learning activities in medical education in a variety of settings, from practical clinical environments to strictly academic arenas. Depending on the format, the teacher may form a strict regimental relationship with the learner or may foster a caring bond.

Learners along the medical educational continuum use various experiential learning methods. These may include apprenticeship, internship or practicum, mentoring, clinical supervision, on-the-job training, clinics, and case study research. For maximum benefits, it is important that they continue to cycle through the four learning environments described in Kolb's model. Box 4.5 gives some basic suggestions for successful experiential learning in clinical rotations.

Situated Learning

Most medical educators accept that clinical educators should be trained to teach, but faculty development for clinicians has been shown to be undermined by weak attendance and poor learning transfer. Therefore, there has recently been interest in situating teacher development initiatives in clinical workplaces [94].

Situated learning belongs to those theories of learning that have a sociocultural basis, which view learning and development as occurring via transformation through participation in collaborative activities with other professionals. Learners transform their understanding, roles, and responsibilities as they participate [16, 95, 96].

BOX 4.5 HOW TO: Successful experiential learning in clinical rotations

1 CONCRETE EXPERIENCE: Gain practical experience in a number of prescribed intervention areas. 'Interventions' are actions that must be completed during the rotation.

2 These prescribed intervention areas are listed and provided at the beginning of rotations.

3 ACTIVE EXPERIMENTATION: Develop the skills required within the prescribed intervention areas. These skills are listed by the rotation Director.

4 ABSTRACT CONCEPTUALISATION: Produce 'evidence' in a portfolio to show experience in prescribed intervention areas and development of the required skills.

5 REFLECTIVE OBSERVATION: Produce several reflective reports (to be included within the portfolio) that demonstrate the learner's development in the prescribed 'key' intervention areas.

Sfard [96] described two metaphors for learning: *acquisition* and *participation*. In the *acquisition* metaphor, learning is seen as the acquisition of knowledge, skills, and attributes that belong to or are 'owned' by the individual. This is a very prominent metaphor, especially in older writings. However, most of the theories discussed to date in this chapter could be seen as fitting that metaphor. In contrast, the *participation* metaphor sees learning as a process of constructing meaning and identity by participating in the activities of that community and becoming a member of the community.

Situated learning is about participation. Learning occurs through collaboration with members of a community of practice (CoP) (e.g. other learners and more senior, or more experienced, individuals) who share responsibility for carrying out certain activities. The purposes of these activities and collaborative processes are connected explicitly with the history and current practices of the community [97]. New learners initially enter and contribute to the community of practice by observing and performing essential but low complexity tasks – a process Lave and Wenger called 'legitimate peripheral participation'. As they take on more responsibility in the community, learners become a more integral part of the community's social and work practices. Through increased participation, they come to understand the particular knowledge that distinguishes the community from others.

A central tenet of situated learning is that learning occurs through social interaction. It emphasises that learning is a process of co-construction through the interaction with and bonds between members of the community. Learners both acquire knowledge (as learners) and share knowledge (as teachers) through all aspects of their participation in the community. A powerful source of learning is the 'discourse' or 'talk' of the community [98]. This is the primary process used for co-constructing new knowledge and practices.

Steven et al. [99] have discussed CoPs as an effective approach to clinical teaching. The results of their study strongly suggest that medical students learn effectively from real patients by participating in patient care within an educational practice. They assert that learning will be affected by clinicians' willingness to engage in supportive dialogue and that an informal, inclusive dialogue on workplace learning might enhance clerkship education. Others [94] advocate use of CoPs as one way to implement teacher development in relation to clinical communities and institutions. Cantillon et al. found that:

> Two types of CoP occupied the horizontal plane of accountability, i.e. clinical teams (Firms) and communities of junior doctors (Fraternities). Participants reproduced teacher identities and practice that were congruent with CoPs' regimes of competence in order to gain recognition and legitimacy. Participants also constructed their teacher identities in relation to institutions in the vertical plane of accountability (i.e. hospitals and medical schools). Institutions that valued teaching supported the development of teacher identities along institutionally defined lines. Where teaching was less valued, clinicians adapted their teacher identities and practices to suit institutional norms.

> Becoming a clinical educator entails continually negotiating one's identity and practice between two potentially conflicting planes of accountability [94, p. 991].

Through discourse, learners begin to participate in a community. Discourse may be thought of as the way we talk about our work and other aspects of our world. The discourse or 'talk' both reflects the way we see our world and frames the way we view it. The community offers a wide variety of relationships and exemplars from whom to learn, including masters, more advanced apprentices, and peers. Learners learn how more senior members of the community walk, talk, and conduct their lives; they observe what other learners are doing and what is needed to become part of the community. Through this observation and participation they learn about the values and shared knowledge and practices of the community. They learn how people in the community 'collaborate, collude and collide, and what they enjoy, dislike, respect and admire' [16, p. 95].

For Lave and Wenger, who introduced the notion of situated learning in communities of practice, the opportunity to learn through relationships with other apprentices and to observe the masters' (senior practitioners') practice creates the curriculum in the broadest sense. Learners can develop a view of what the whole enterprise is about, and what there is to learn. 'Engaging in practice, rather than its object, may be a condition for the effectiveness of learning' [16, p. 93].

It is useful to consider the relationship of situated learning to other learning theories. Situated learning allows a broad view of learning that relates to several other conceptions of learning, both long-standing and more recent.

Situated learning shares with social cognitive theory [11] the view that learning occurs in a dynamic interaction between the learner and the environment. Situated learning suggests that learning is not separate from social influences. The context in which teaching and learning occur is critical to learning itself, and learning is culturally and contextually specific [100]. Learning occurs within social relations and the practices that occur there.

Situated learning also holds that some knowledge related to a task is only present in the context or location of the task. Brown et al. [101] described *situated cognition* and emphasised the idea of cognitive apprenticeship. Cognitive apprenticeship supports learning in a domain by enabling students to acquire, develop, and use cognitive tools in authentic domain activity. This happens in practice as teachers guide learners through processes of framing problems and applying disciplinary knowledge to their solution. In the process, teachers provide a scaffold for the learner's development, which can be withdrawn gradually as the learner gains more knowledge and experience.

Situated learning as described by Lave and Wenger [16] extends beyond the acquisition of concepts and structures by the individual and includes all of the learning in the learning environment. It views the community and learning opportunities as a way of structuring learning resources, with pedagogical activity (teaching) as only one resource among many.

Situated learning theory was originally a means for studying the learning that occurs through apprenticeship [16]. Traditionally, apprenticeship has been viewed as a relationship between a master or senior practitioner and a novice or learner. Through apprenticeship, the learner comes to understand the content and process of professional practice. Situated learning provides a way of understanding the process whereby apprentices acquire knowledge and skills through following and attempting to be like the teacher or expert practitioner. In the situated learning model, the learner's apprenticeship is actually to the whole community, and much of the learning occurs in the relationships between people, rather than solely as mental activity for the individual learner.

Situated learning also relates closely to our growing understanding of informal learning. According to Eraut [102], informal learning is a significant dimension of the learning that occurs in the course of our work. He suggests that it is implicit, unintended, opportunistic, and unstructured, and often occurs in the absence of a teacher. Learning about how things are done, exposure to a variety of different approaches, and practical approaches to problems occur daily. There is still much to understand about it; however, the evidence is convincing that informal learning and learning from others occur in the workplace. This image of learning contrasts with that of the independent learner – an image that is embedded in much of formal medical education. Informal learning has as a corollary implicit learning – 'the acquisition of knowledge independent of conscious attempts to learn and in the absence of explicit knowledge about what was learned' [103]. Implicit learning results in tacit knowledge – 'that which we know but cannot tell' [104].

Situated learning also relates to experiential learning, or learning by doing. Experiential learning has as its goal the integration of conceptual models and concrete experience [15], and of authentic experience and formal education. In medical education, situated learning extends the concept to include experiential learning occurring within a clinical context. It also extends the idea of experiential learning beyond the individual learner, as it views the learner as contributing to and participating in the shared experience of the community.

In addition to all the above theoretical relationships, situated learning is entirely in keeping with constructivism (described later in this chapter). Constructivism views learning as a process of active participation in problem-solving and critical thinking. Through these processes, learners construct their own knowledge and understanding of the world based on their previous knowledge and experience. Knowledge is integrated into previously existing concepts and schemata, which gradually become richer and more connected.

Post-modern constructivist approaches do not view the locus of knowledge as within the individual. Rather they view learning as a social constructivist process. Learning and understanding are social; cultural activities and tools are essential to conceptual development that will allow learners to develop the skills and standards that are valued by the community [100]. In the context of situated learning, knowledge may be constructed not only individually, but jointly by communities and the individuals who are members of those communities.

Implications for Educational Practice

Situated learning is relevant to medical education in many ways and at all levels of the continuum of education. Apprenticeship remains a pervasive teaching and learning method in physician learning. Learners in undergraduate and postgraduate medical education programmes are assigned to various clinical and community sites where they are immersed, to a greater or lesser degree, in the work of the community, performing minor tasks, and striving to learn from the more advanced learners and mentors in the community. Authentic activity is important for learners because it is the only way they gain access to the standpoint that enables practitioners to act meaningfully and purposefully. It is activity that shapes or sharpens their tools [101]. However, there is another important aspect of situated learning, namely socialisation.

Increasingly, medical and health professional education are recognised as a process of professional socialisation. In that process, learners are developing their professional identity. Their experiences, knowledge, interactions, and informal and formal learning all contribute to the professional identity that each individual constructs. The recent Carnegie Foundation report on physician education [105] suggests that a focus on learners' development of professional identity is one of four fundamental principles for reforming physician education.

Hafferty and Franks [106] articulated the notion of three levels of curriculum as including formal, informal, and hidden. These may be helpful in thinking about the environment or community in which our learners are placed. The *formal curriculum* represents the stated goals, explicit content, and planned educational materials or resources provided for learners. The *informal curriculum* includes both explicit and serendipitous goals and is found in the interaction between learners and teachers, clinical environments, other students, personal interests, and goals. Part of the informal curriculum may also be what Hafferty and Franks termed the *hidden curriculum*, which is seen in the practices and routines of the community, particularly in relation to how its members cope and thrive. The hidden curriculum often teaches values and moral judgements and is found especially in an institution's policies, language, assessment strategies, and allocation of resources. Clearly, these curricula all exist and are enacted in the context of situated learning in medicine. Importantly, not all messages from the hidden curriculum are negative. Both negative and positive aspects have been described. Often these are unintentionally imparted through actions, discussions, and relationships among members of the community. This relates the notion of situated learning closely to role modelling, as the senior members of the community enact through their behaviours, both tacitly and explicitly, how problems of the discipline are approached, how colleagues are regarded, and how knowledge is built and used.

When learners are involved in clinical placements, participation in the actual daily round of activities is important in enhancing the effectiveness of their learning. Clearly, the longer the learner's engagement in a community, the greater is the opportunity to participate meaningfully. Where attachments are short, learners may remain at the periphery and experience little feeling of participation in the community. Bates et al. [107] showed that clerkship students who studied for extended periods in one environment with one preceptor perceived assessment and feedback as authentic. They were authentic because the students were embedded in daily patient care, useful because the assessment and feedback were developmental and longitudinal, and constructive because they occurred in the context of a supportive learning environment and relationship.

Faculty (teachers) enact several roles concurrently. As in the perspective of social learning theory, they are modelling skills, knowledge, values, and attitudes that learners observe, along with how those actions are received in the community. Beyond role modelling, faculty are also demonstrating how knowledge is built and understood and how practices evolve. This aspect of observed practice offers both challenges and advantages. Learners who participate in practice and listen to the talk of the community are able to learn in a contextualised way. However, the nature and content of the talk become important considerations. As teachers, being mindful of our talk and being open to reflecting on practice with learners is important. Learning through observation is also vulnerable to misunderstanding, as learners will interpret what they observe in light of their current experience and understanding [108]. It is important to find opportunities and demonstrate willingness to discuss and reflect on experience with learners [109].

Brown et al. emphasise the idea of cognitive apprenticeship: 'Cognitive apprenticeship supports learning in a domain by enabling students to acquire, develop and use cognitive tools in authentic domain activity' [101, p. 39]. Cognitive apprenticeship means that the learner observes the thinking process, and not just the actions, of experts and other participants in the community. Box 4.6 outlines one way to use a cognitive apprenticeship approach with learners in a clinical setting.

Participation in the work of the clinical site or community is a key to this understanding of learning. Situated learning suggests that all members of the community are involved. In the case of medical education this means that more senior learners and other health professionals can all enhance the newer learners' participation and sometimes can also learn from the newcomer.

Different fields of medicine have distinct knowledge and skill bases. However, there will still be some aspects that are common to all, including communication with patients, ethical approaches and grounding of actions, basic clinical skills, learning in interprofessional teams, etc. in which learners can participate across their fields of experience. As teachers, we can think carefully about how we can promote this participation among learners.

Building on the advantage of situated learning, we have the opportunity to rethink our learners' experience and consider all the ways we have available to promote their learning. However, this involves thinking of learners as contributing members of our learning environment, rather than as temporary adjuncts to it.

BOX 4.6 HOW TO: Steps for conducting cognitive apprenticeship

Steps	Actions
Modelling	The expert performs the skill and the learner observes and builds a cognitive model of the steps involved.
Coaching	The expert observes the learner and offers tailored feedback and perhaps more modelling.
Scaffolding	The expert takes into account the learner's current skill level and provides appropriate activities to support the learner's progression.
	The expert 'fades' into the background and diminishes support until the learner can perform the skill alone.
Articulation	The expert assists the learner in articulating their knowledge and/or reasoning by questioning or explaining what they are doing and why.
Reflection	The expert encourages and supports the learner to assess their own performance and to compare it with that of an expert, another learner, and ultimately, their own internal cognitive model.
Exploration	The expert moves the learner into problem solving on their own and supports the learner in facing new problems in practice.

Source: Adapted from Collins et al. [110, p. 455].

Communities of Practice

In the context of situated learning, we discussed the idea of community as a place where participants are socialised and develop a professional identity. In this section on communities of practice, we expand this concept to include knowledge transmission, construction, and translation. Lave and Wenger [16] first proposed the term *community of practice* (CoP) to capture the importance of activity in integrating individuals within a community, and of a community in legitimising individual practices. Within this context, they described a trajectory in which learners move from legitimate peripheral participants to full participation in the CoP. The concept of legitimate peripheral participation means that access to a CoP, with its resources and activities, provides a means for newcomers to learn through observation and gradually deepen their relationship to the CoP. Barab et al. [111, p. 495] defined a CoP as 'a persistent, sustained social network of individuals who share and develop an overlapping knowledge base, set of beliefs, values, history, and experiences focused on a common practice and/or mutual enterprise'. Wenger and Wenger-Trayner defined CoPs more simply as: 'Communities of practice are groups of people who share a concern or a passion for something they do and learn how to do it better as they interact regularly' [112].

Wenger proposed three constituent parts of a CoP: *mutual engagement*, *joint enterprise*, and a *shared repertoire*. Mutual engagement involves both work-related and sociocultural activities, achieved by interaction, shared tasks, and opportunities for peripheral participation. Joint enterprise refers to the need for the group to respond to a mandate for itself rather than simply as an external mandate. Finally, a shared repertoire involves the 'routines, words, tools, ways of doing things, stories, gestures, symbols, genres, actions or concepts that the community has adopted in the course of its existence' [16]. Wenger summarised his conceptual framework for a social theory of learning comprising four components that are 'deeply interconnected and mutually defining' [16]. All of these should be present in a true CoP. Framework components include the following:

- *Meaning* – learning as experience. Members talk about their experience and create shared meaning.
- *Practice* – learning as doing. Members talk about the shared ideas and resources that can sustain action.
- *Community* – learning as belonging. Members talk about the community process and how they are learning and developing competence.
- *Identity* – learning as becoming. Members talk about how learning changes who they are.

Therefore, we can see that the concept of a CoP is complex and multidimensional, serving multiple purposes both for individuals and for the sub-communities that participate in the full community. Wenger and Wenger-Trayner [112] provide a description of typical activities in CoPs, shown in Box 4.7 with examples that reflect medical settings.

The primary purpose of CoPs in this conception is knowledge translation. Knowledge translation has been defined as 'the exchange, synthesis and ethically sound application of knowledge – within a complex system of interactions among researchers and users – to accelerate the capture of the benefits of research … through improved health, more effective services and products, and a strengthened health care system' [112].

More recently, other terms have been proposed for essentially the same broad concept. These terms include knowledge mobilisation [113], knowledge utilisation [114], knowledge exchange [115], knowledge management [116], and knowledge brokering [117], all of which involve an active exchange of information among various stakeholders, such as researchers, health care providers, policy makers, administrators, private sector organisations, patient groups, and the general public. Partnerships are at the heart of all knowledge translation activity [118], and effective

BOX 4.7 FOCUS ON: Typical activities in communities of practice

Activity	Example from medical practice
Problem solving	'Can we discuss this patient and brainstorm some ideas? I'm stuck.'
Requests for information	'Does anyone know a good website about treatment of Chagas' disease?'
Seeking experience	'Has anyone dealt with a patient in this situation?'
Reusing assets	'I have a proposal for a new clinic that I wrote for our hospital last year. I can send it to you and you can easily tweak it for your situation.'
Coordination and synergy	'Can we combine our purchases to achieve bulk discounts?'
Discussing developments	'What do you think of the new patient information system? Does it really help?'
Documentation projects	'We have faced this problem five times now. Let us write it down once and for all.'
Visits	'Can we come and see your clinic? We need to establish one in our city.'
Mapping knowledge and identifying gaps	'Who knows what are we missing? What other groups should we connect with?'

Source: Adapted from Wenger and Wenger-Trayner [112].

knowledge translation is dependent on meaningful exchanges among network members for the purpose of using the most timely and relevant evidence-based, or experience-based, information for practice or decision-making.

In the field of continuing medical education, the limitations of traditional workshop/presentation models are apparent [119]. It is now recognised that there is a need for continuous learning to occur in the context of the workplace and for reflection-in-practice and reflection-on-practice to be supported [12]. Knowledge translation is essential to shortening the path from evidence to application of that evidence in practice, and CoPs provide an opportunity to embed learning within a clinical context. A highly effective way to learn about complex issues is through experience, application, and discussion with mentors and peers in the same or similar contexts. Relevant learning occurs when the participants in the CoP raise questions or perceive a need for new knowledge. Using Internet technology enables these discussions to occur in a timely manner, and records of these can be archived for later review or by those who missed the discussion.

There are a number of key factors that influence the development, functioning, and maintenance of CoPs [120]. The legitimacy of the initial membership is important. Commitment to the desired goals of the CoP, relevance to members, and enthusiasm about the potential of the CoP to have an impact on practice are also key. On the practical side, a strong infrastructure and resources, such as good information technology, useful library resources, databases, and human support, are essential attributes. Ensuring that these key factors are in place requires strong, committed, and flexible leaders who can help guide the natural evolution of the CoP. If professional learning is to flourish, it is critical that a blame-free culture be established in which community members can learn from positive and negative experiences [121].

The benefit for being involved in a CoP is increased expertise and skills. Intrinsic motivation for participating in a CoP can include [122]:

- *Anticipated reciprocity*. A member is motivated to contribute to the community in the expectation that he or she will receive useful help and/or information in return.
- *Increased recognition*. The desire for prestige is a key motivation for an individual's contribution to a learning community.
- *Sense of efficacy*. The act of contribution results in a sense that the individual has had some effect on the community.

Endsley et al. [123] outline key questions to address in establishing a CoP:

- Why are we forming?
- Who will participate?
- What will we share?
- How will we interact?
- What will we impact?
- How will new knowledge be found and used?
- How will the community evolve to meet new choices and challenges in practice?

These authors also identify typical questions addressed in CoPs and examine the myths surrounding CoPs [123]. They describe CoPs as dynamic entities that need to be designed for adaptability and large growth. CoPs should combine the perspectives of both insider members and outsider participants, and all members should be valued, regardless of their level of participation. Both public and private spaces are necessary and need to be related. A critical principle is that the CoP must provide value to its members, otherwise participation will be minimal or absent. Although familiarity is important, challenge and excitement need to be provided to keep the energy high. Finally, a CoP needs to settle into a rhythm that works for its members. Box 4.8 lists five factors that a CoP requires to be successful.

Virtual Communities of Practice

Virtual (online) communities play a socialisation role to the same extent as real communities [124, 125]. The theoretical foundation of virtual communities is based on social cognitive theory and situated learning. Henri and Pudelko [124] propose three components of the social context of activity in virtual communities – the goal of the community, the methods of initial group creation, and the temporal evolution of both the goals and the methods of the group – leading to the development of four different types of community. Figure 4.4 illustrates different forms of virtual communities of practice.

BOX 4.8 HOW TO: Requirements for a successful community of practice

Lave and Wenger [16] suggest that the success of a community of practice depends on the following five factors:
- the existence and sharing by the community of a common goal,
- the existence and use of knowledge to achieve that goal,
- the nature and importance of relationships formed among community members,
- the relationships between the community and those outside it, and
- the relationship between the work of the community and the value of the activity.

Wenger [17] later added the idea that achieving the shared goals of the community requires a shared repertoire of common resources – for example, language, stories, and practices.

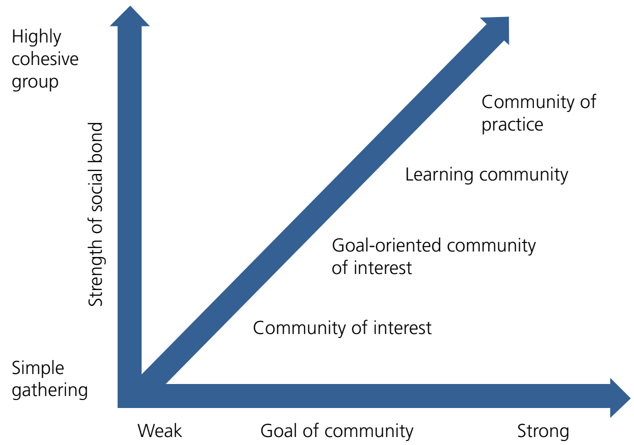


Figure 4.4 Different forms of virtual communities of practice according to their context of emergence. *Source*: Adapted from Henri and Pudelko [124, p. 476].

BOX 4.9 FOCUS ON: Principal descriptors of the four types of virtual communities

	Community of interest	Goal-oriented community of interest	Learners' community	Community of practice
Purpose	Gathering around a common topic of interest	Created to carry out a specific task	Pedagogical activity proposed by the instructor	Stems from an existing, real community
Activity	Information exchange	Sharing of diverse perspectives and production of objects commissioned by the mandate	Participation in discussions of collective topics	Professional practice development through sharing knowledge among members
Learning	Knowledge construction for individual use	Knowledge construction from diverse knowledge systems towards collective use	Knowledge construction by carrying out social situated activities	Appropriation of new practices and development of involvement

Source: Adapted from Henri and Pudelko [124, p. 485].

Box 4.9 dissects the characteristics of four types of community. It demonstrates that although many types of virtual community can exist, they may not be true CoPs. The virtual CoP generally arises from an existing, face-to-face CoP in which professional practice is developed through sharing knowledge among members. Through this interaction, new practices may be developed and identification with the community can occur.

Some writers have distinguished 'soft' from 'hard' knowledge [126]. Soft knowledge can be gathered in a domain through sharing solutions to a particularly difficult problem, describing idiosyncrasies of particular tools, equipment, or processes, and recounting and reflecting on challenging events (i.e. recounting 'war stories'). This refers to the implicit or tacit knowledge in a domain. CoPs are central to the creation and maintenance of soft knowledge. Hard knowledge, in contrast, is stored in databases and documents. It is highly explicit and codified. A key question is whether a virtual CoP can effectively share soft knowledge, which tends to be situated in specific contexts. This is a question that requires further research.

There are currently a number of large virtual CoPs in the medical/health field, such as:
- www.doctors.net.uk
- www.doctorslounge.com
- www.medhelp.org
- http://www.aamc.org/initiatives/diversity/portfolios/cdc/344696/cop.html

However, virtual CoPs are a relatively recent phenomenon, and more studies are needed on factors affecting their effectiveness for enhancing learning. Parboosingh [126] advocates conducting evaluation studies that focus on how the

CoP takes advantage of the technology, rather than how the technology affects the CoP. Resources are available to assist with this [127], and many methods are available to evaluators. Examples include:

- *Case studies*. Investigate changes that happened within a particular project/organisation as a result of a member's participation in the CoP.
- *Contribution analysis*. Consider the extent to which observed results are due to programme (CoP) activities, rather than other factors.
- *Horizontal evaluation*. Combines self-assessment and review by peers.
- *Institutional histories*. Records new ways of creating more effective ways to achieve goals; can be used by the CoP co-ordination team.
- *Institutional linkage diagram*. Illustrates the extent to which individuals, organisations, projects, or services interact with each other; can be used to illustrate how individuals and organisations who never engaged in a CoP before are now in contact because of the virtual CoP platform.
- Web resources such as:
  - https://sites.google.com/a/cgxchange.org/evaluation-cop/home
  - https://www.cdc.gov/phcommunities/resourcekit/evaluate/start_evaluation.html

Implications for Educational Practice

This section on implications integrates the concepts of knowledge transmission, construction, translation, and CoPs. These ideas have many obvious applications in the medical education arena, and a number of CoPs are emerging in various specialties. As an example, this section outlines a CoP for palliative care practitioners and students. This is an excellent CoP application because palliative care is a truly interdisciplinary field that involves subcommunities of various specialties, including oncologists, family physicians, nurses, and social workers. These subcommunities need to interact in CoPs, but the various professional groups also need to interact with each other around specific topics and cases. This furnishes an excellent model for continuing medical education and also provides an environment for training residents, interns, and medical students. Since many participants are acquiring and applying new knowledge in this field, scaffolding learners through an evolving continuum from simulation to participation to co-determined interactions is an effective instructional approach [127]. For example, family physicians, residents, and nurses who have trained with oncology specialists may begin with simulated cases. They then learn to participate in real cases supported by learning materials and/or clinicians, until they are able to operate as full participants. The scaffolding process proposed here uses a staged approach for bridging from a learner (knowledge) identity to a participant (practitioner) identity. This approach is consistent with the constructivist view of learning [124], which espouses the learner as central in the educational process. The advantages of the situated learning approach over the traditional didactic approach are discussed above in the section on situated learning.

A CoP implemented for palliative care in a community-based learning environment could include specialist and non-specialist practitioners in palliative care as well as residents and medical students (clerks). It would aspire to achieve a number of different aims, based on the challenges identified by Richardson and Cooper [128]:

- Engage all trainees in a research culture (i.e. encourage evidence-based practice).
- Provide an opportunity for participants to identify with their peers and supervisors.
- Encourage cross-site discussion to explore shared theoretical, methodological, and practical issues.
- Provide a forum for discussion and a recognised channel for communication and collaboration.
- Facilitate high-quality supervision to ensure adequate access to teaching and learning for all practitioners.
- Foster scholarly interaction and good supervisory practice to stimulate dialogue among learners and supervisors across sites.

CoPs provide a critical resource for professionals who want and need recommendations, pointers, tips and tricks, best practices, insights, and innovations. Part of what makes a CoP strong is the 'aggregation of relevance'; that is, people and information related to a coherent set of topics that certain people will find interesting, useful, and potentially profitable. Linking medical students, their community preceptors, and medical school specialists in an online CoP can greatly enhance the learning and practice experience of all participants. White and Thomas [129] demonstrated that students assigned to community practices for their paediatric clerkship perform as well as, or better than, students assigned to academic medical centres in written examinations; other studies have demonstrated similar outcomes. An online CoP approach can build on this positive finding and perhaps provide an even more effective community experience for medical students. A side benefit could be improvement in teaching and supervisory methods used by their preceptors.

Constructivism

Constructivism has multiple roots in twentieth-century psychology and philosophy. It emerged from Piaget's developmental perspective [87] and Bruner's cognitive psychology [130] and is an amalgamation of both behaviourist and cognitive principles [131]. The constructivist stance maintains that learning is a process of constructing meaning, and this is how people make sense of their experience [131].

There are two major strands of constructivist perspective: (i) cognitive constructivist and (ii) social-cultural (socio-constructivist). Cognitive constructivism is an individualistic perspective based on Piaget's work. He asserts that learning does not occur passively but occurs by active construction of meaning [132]. Piaget explains that when learners encounter an experience that challenges the way they think, a state of disequilibrium or imbalance is created. Learners must then alter their thinking to restore equilibrium or balance. For this purpose, learners make sense of the new information by associating it with what they already know, that is, by attempting to assimilate it into their existing knowledge. When learners are unable to do this, they use accommodations to restructure present knowledge to a higher level of thinking.

Fosnot [133] defines constructivism according to four principles: (i) learning depends on what individuals already know, (ii) new ideas occur as individuals adapt and change their old ideas, (iii) learning involves inventing ideas rather than mechanically accumulating a series of facts, and (iv) meaningful learning occurs through rethinking old ideas and coming to new conclusions about new ideas that conflict with our old ideas.

For constructivism, learning is represented as a constructive process in which the learner is building an internal illustration of knowledge, a personal interpretation of experience. This representation is always open to modification, its structure and linkages forming the ground to which other knowledge structures are attached. Learning is then an active process in which experience has an important role in understanding and grasping meaning. This view of knowledge does not necessarily reject the existence of the real world. However, it contends that reality places constraints on existing concepts and that all individuals' knowledge of the world consists of their interpretations of their experiences. Furthermore, conceptual growth is the result of various perspectives, and the simultaneous changing of individuals' internal representations in response to those perspectives, as well as through their experience [134].

Social constructivism in its modern form has been in existence for approximately 40 years. Strictly speaking, while it is thought of as a learning theory, with roots in cognitive constructivism [135] and sociocultural theory [18], it is more correctly an epistemology or philosophical explanation of the nature of learning [136].

Social constructivism is based on Vygotsky's [18] theories about language, thought, and their mediation by society. Vygotsky states that the process of knowing is affected by other people and is mediated by community and culture. Social constructivism emphasises that all cognitive functions including learning are dependent on interactions with others (e.g. teachers, peers, and parents). Therefore learning is critically dependent on the qualities of a collaborative process within an educational community, which is situation-specific and context-bound [137]. However, learning must be seen as more than the assimilation of new knowledge by the individual; it is also the process by which learners are integrated into a knowledge community.

According to social constructivists, nothing is learned from scratch; instead it is related to existing knowledge, with new information being integrated into and expanding the existing network of understanding. The successful learner is therefore one who embeds new ideas within old ones and for whom understanding expands to encompass the new experience. A social constructivist learner's view of the world will always be subjective, as each individual will interpret experience via a different pre-existing framework of understanding and will develop their own unique view of the world.

Social constructivists assert that knowledge is a human product that is socially and culturally constructed in an active manner and not something that can be discovered [138, 139]. Knowledge is therefore neither tied to the external world nor wholly to the working of the mind, but exists as the outcomes of mental contradictions that result from one's interactions with other people in the environment [137].

Social constructivism maintains that learning is based on real-life adaptive problem solving, which takes place in a social manner through shared experience and discussion with others such that new ideas are matched against existing knowledge and the learner adapts rules to make sense of the world. Social constructivism views learners as part of a social group in which learning occurs through group interaction, not simply within the individual. Learning is an active, socially engaged process [140, 141]. According to social constructivists, the process of sharing individual perspectives or elaborating with others results in learners constructing an understanding together that would not be possible alone [140]. Social constructivism maintains that while it is possible for people to have shared meanings that are negotiated through discussion, it also acknowledges that no two people will have exactly the same discussions with exactly the same people. To this extent social constructivism allows that multiple realities exist.

Social constructivism views learning as an active process where learners should work to discover principles, concepts, and facts for themselves [101]. This approach does not see education as a process in which the teacher/tutor 'pours' knowledge into passive students; instead it emphasises how students should be actively involved in their own learning process.

A main focus of social constructivism is the role that social interaction and social processes play in creating knowledge. Vygotsky [18] believed that learning could not be separated from social context. He argued that all cognitive function begins as a product of social interactions. Social constructivism requires one primary element: two or more participants. These participants must be involved in some form of interaction for knowledge to be constructed, and they must have knowledge of prior social experience [142]. It is a shared understanding among individuals whose interaction is based on common interests that form the ground for their communication. Therefore, during the interaction between participants, this prior knowledge is exchanged in a transaction in order to negotiate meaning. This meaning does not have to be strictly language based but can also be a product of actions.

A key concept in social constructivism is the zone of proximal development, often abbreviated as ZPD. This is defined as the difference between what a learner can do without help and what he or she can do with help [143]. Since Vygotsky's original conception, the definition of the ZPD has been expanded and modified. The ZPD is an area of learning that occurs when a person is assisted by a teacher or peer with a skill set higher than that of the subject. The person learning the skill set cannot complete it without the assistance of the teacher or peer. The teacher helps the learner attain the skill the student is trying to master, in the hope that the teacher will no longer be needed for that task. The ZPD concept is seen as a form of scaffolding referring to the help or guidance received from a teacher or more competent peer to permit the learner to work within the ZPD. Vygotsky never mentioned the term 'scaffolding'; the concept was first developed by Bruner [130] while applying Vygotsky's concept of ZPD to various educational contexts. According to Wass and Golding [144], giving learners the hardest tasks they can do with scaffolding will lead to the greatest learning gains.

Implications for Educational Practice

Hoover [145] argues that constructivism has important implications for teaching. First, teaching cannot be viewed as the transmission of knowledge from enlightened or known to unenlightened or unknown. Constructivist teachers are not monologue teachers who teach completely new lessons. Rather, constructivist teachers have the role of guides for learners, providing them with opportunities to test the adequacy of their current understandings.

Second, constructivist teachers consider the prior knowledge of their learners and provide learning environments that exploit inconsistencies between learners' current knowledge and their new experiences [145, 146]. Differences among learners challenge teachers and do not allow them to use the same method or the same materials while teaching diverse learners.

Third, since constructivism emphasises learners' involvement, teachers must engage learners in learning and bring learners' current understanding to the forefront [145]. Constructivist teachers can ensure that learning experiences include problems that are important to learners and are not just related to the needs and interests of teachers and the educational system.

Fourth, Hoover [145] highlights the need for sufficient time for learners to actively build new knowledge. During this time, learners reflect on their new experiences and try to consider the relationship between these experiences and previous ones in order to have an improved (not 'correct') view of the world.

According to the social constructivist theory, teachers should assume the position of 'facilitators' rather than 'imparters of knowledge'. A facilitator encourages learners to achieve their own appreciation of the content. When the teacher functions as an imparter of knowledge, the learner can quite easily play an unreceptive role; when the teacher acts as a facilitator, learners are encouraged to play a more functional and effective role within their own learning. Therefore, the focus is on learners and what they are able to do. In a social constructivist educational setting, the responsibility for learning falls on the learner, while the teacher is a facilitator who guides direction and promotes new patterns of thinking.

This dramatically different expectation of a teacher acting as a facilitator as opposed to a lecturer suggests that within social constructivist learning, the teacher plays a fundamentally different role from that which is expected in a 'traditional' teacher-directed model [147]. For example, a teacher offers answers, as in a traditional programme, while a facilitator offers strategies that allow the learner to achieve his or her independent conclusions [148]. The learning environment should be designed to support and challenge the learner's thinking. The critical goal is to support the learner in becoming an effective thinker and to develop skills in heuristic problem solving, metacognitive knowledge, creativity, and originality as a by-product of increasing the level of understanding of the topic of interest [149, 150].

Collaborative Learning

Social constructivism stresses the need for collaborative learning. Learning is promoted through collaboration among students and between students and teachers. From a social constructivist perspective, as students share background knowledge and participate in the give and take of collaborative activities they are actually negotiating meaning and building knowledge, not as individuals, but as a group. This collaboration in tasks and discussions allows learners with different skills and backgrounds to arrive at a shared understanding [134]. Social constructivist approaches should require the learners to collaborate and critically analyse issues. Some examples of collaborative learning activities are group problem solving, group inquiry, simulations, and debates. The activities encourage creativity and foster higher-level thinking [151].

Numerous benefits have been described for collaborative learning. Johnson and Johnson [152] and Panitz [153] list over 50 of these, summarised by Laal and Ghodsi [154] into four major categories of social, psychological, academic, and assessment benefits. Examples include:

1. Helping to develop higher level thinking, e.g. critical thinking.
2. Building self-efficacy.
3. Developing social, communication, and teamwork skills.
4. Improving problem-solving skills.

In collaborative learning, learners work towards a common goal and hold responsibility for one another's learning as well as their own. For instance, peer tutoring, where learners in the same group tutor one another, facilitates meaning for both parties, as the peer tutor clarifies his/her own understanding through the teaching process. When a teacher and a few learners form a collaborative group, group members apply four cognitive strategies: questioning, summarising, clarifying, and predicting. This creates a ZPD in which students gradually assume more responsibility for the material, and through collaboration forge group expectations for high-level thinking and acquire vital skills for learning and success in everyday life. Box 4.10 offers suggestions for putting social constructivism into action.

Socio-materiality

Several authors have recently proposed 'socio-materiality' as a theoretical framework that can be used in medical education [155, 156]. Within the social sciences during the last three decades, there has been a significant turn towards the study of how material things – for example, objects, animals, machines, humans, and organisations – might be arranged, manipulated, or enacted to allow particular tasks, activities, or practices to be accomplished. MacLeod et al. [155] have applied this theory to distributed medical education, and Fenwick and Dahlgren [157] have has applied it to simulation in medical education.

BOX 4.10 HOW TO: Putting social constructivism into practice

- Promote discussion, even in lectures, by asking open questions and providing time for responses.
- Encourage the connection of ideas via analysis, prediction, and justification of new ideas.
- Provide tasks just above the learners' competence, and be available to assist if needed.
- Promote peer collaboration around problems and group project work.
- Set up study groups for peer learning.
- Promote the use of technology to provide team-based simulations of real activities, networked writing, and communications.
- Allocate a proportion of grades to peer assessment, training learners in the process.

Science and technology were among the first fields studied using this focus on the material foundation of reality, knowledge, and social life. A major theme running through this literature is that making and distributing knowledge requires the collaboration and interaction of many different material entities – for example, humans and other natural objects, instruments and technologies, texts and images. This literature blurs the usual distinction between the natural world and the social world; it views science as both natural and social.

A related theoretical formulation called Activity Theory, or more correctly, Cultural-Historical Activity Theory or CHAT [156], is older than socio-materiality, is quite developed methodologically, and has been applied to medical education [158, 159].

Activity Theory is not actually a theory in the strict interpretation of the term, but rather consists of a set of basic principles that constitute a general conceptual system. The basic principles are beyond the scope of this chapter but include object-orientedness, the dual concepts of internalisation/externalisation, tool mediation, hierarchical structure of activity, and continuous development [159]. Simply put, activity cannot exist as an isolated entity. The very concept of activity implies that there is an agent who takes action (individually or collectively) directed at something. According to Activity Theory terminology, activity mediates interaction between subjects (agents) and objects (things) [160, 161]. Some writers consider Activity Theory as a conceptual framework that is encompassed by the socio-material perspective, since it emphasises the importance of material artefacts that mediate human activity systems. However, its primary focus is on human activity, i.e. divisions of labour, cultural rules and languages, and social purposes [156].

Ajjawi and Bearman [162] provide an example of how socio-material theory could be useful in a clinical setting. They explain that from the socio-material perspective, a family practice preceptor and trainee can be viewed as acting within two separate and inter-related activity systems – patient care and trainee education. There may even be a third activity system related to research. These activity systems have their own divisions of labour, materials, social rules and routines, policies, and practices (that can even be contradictory in practice). Understanding the supervisory relationship requires an understanding of these activity systems and the tensions and contradictions that occur among them. Unfortunately, some practitioners have found this theory difficult to understand and apply, partly because it is interpreted and explained differently by different authors.

Today the topic of socio-materiality is one of the most popular, most cited, most debated, and most critiqued topics in the fields of information systems and management [163]. The concept of socio-materiality is extremely theoretical. Authors who write about socio-materiality attempt to make a pointedly philosophical statement about the relationship between the social and the material that begins, quite overtly, with the name 'socio-material' – a deliberate fusion of the words 'social' and 'material' [163]. From this perspective, people and things only exist in relation to each other. In other words, entities (whether humans or technologies) have no inherent properties, but acquire form, attributes, and capabilities through their interpenetration [164].

Similar points can be made about medical education. Material elements are foundational to every aspect of social life, including education, but when we think about social issues we tend to focus exclusively on relations among human beings. Inherent technological issues (e.g. Internet infrastructure) are taken for granted, and we often stop thinking about the effects they have in the world and on us unless they stop working. Socio-material theories of science and technology encourage us to unravel the tangle of human and non-human elements. Rigorous study of medical education requires theory that addresses both human and material factors and the many connections and interactions that occur among them.

Using the example of distributed medical education, MacLeod et al. [155] present three ways in which socio-material theory can be useful to medical education. First, it can help to disentangle the materials, technologies, knowledge, physical spaces, nature, and objects of all kinds and to shine a light on previously obscured actors, infrastructure, and other material factors to give a fuller picture. Second, it can help us to understand the dynamics of everyday life, particularly learning. Third, socio-material approaches have the ability to unsettle ideas that we have taken for granted, leading us to think differently and even disrupt traditional approaches.

Fenwick and Dahlgren point out that medical educators are increasingly turning to theories that '… help to elucidate student learning in dynamic contexts of continuous uncertainty and emergent outcomes' [157, p. 360]. They note that more emphasis is being placed on recognising and attending to the ways in which *materials* (objects, texts, technologies, bodies, settings) move in practice and learning, and how they are related to the *social* (texts, symbols, meanings, intentions) in complex systems. Orlikowski, who is often quoted on socio-materiality, says: 'The social and the material are considered to be inextricably related – there is no social that is not also material, and no material that is not also social' [164, p. 1437].

Fenwick and Dahlgren [157] highlight useful questions and approaches offered by socio-material theory to expand and deepen medical students' learning. They base their discussion on the case of simulation-based medical education. In a workplace in which we are doing things, learning shifts from an emphasis on preparing for action by acquiring knowledge, to a process of participating effectively in context.

Box 4.11 illustrates the key concepts of socio-material theory, as well as questions they raise for educators.

Although this theory has only recently emerged in medical education, it has potential as a practice-based theory to increase our understanding and improve our practices. Fenwick eloquently presents the case for a socio-material approach:

> Educators as well as students can look more closely at what material elements most influence their learning and teaching processes, how materials limit or enhance possibilities for learning, why particular educational or learning practices become stabilized and powerful, and when these black boxes create problems. This is not about stuffing more activities into crowded curricula, but about opening out ways of engaging students [156, p. 91].

Implications for Practice

As mentioned above, socio-materiality theory has been applied in the fields of information systems and management [164–166]. This has not been an easy theory to apply, perhaps because the concept of socio-materiality is so theoretical. In the field of medical education, there have been two excellent examples of its application. The first is in the area of distributed medical education (DME). MacLeod et al. [155] explain that DME relies heavily on the adoption and integration of material resources in the form of technologies and that these are associated with social resources. Social considerations include the development of viable organisational strategies, appropriate teaching modalities, suitable assessment standards, and new definitions of meaningful social/professional interaction. They illustrate that considering social presence [167], pedagogical presence [168], and cognitive presence [169] in distance education can lead to insights that are helpful for practice. An example related to social presence is the use of Skype by a small group of learners, who find it difficult to get to know their peers and to interact if the technology doesn't function properly. An example related to pedagogical presence is video-conferencing technology, as a material condition, since it influences the ways in which lecturers are willing and/or able to engage with learners. The technology is not easy to learn for occasional users, and a lecturer could be preoccupied with how to use the technology rather than being focused on how to engage learners. Finally, for cognitive presence, if someone has a question in a video-conferenced lecture setting, there is a sequence of material factors to consider. Questions are placed in a queue and the lecturer responds to them based on the order in which they were asked rather than on relevance to the conversation. This technologically mediated question ordering affects the group's ability to engage in authentic dialogue.

The second example addresses the topic of simulation in medical education. Fenwick and Dahlgren [157] argue that this field lacks a theoretical base. They point out that some writers believe that simulation in this context needs

BOX 4.11 FOCUS ON: Socio-materiality

Common aspects of socio-material approaches to understanding education and some questions these raise for educators

Key socio-material understandings	Questions raised for educators
Focus is on *materials* as dynamic and enmeshed with human activity.	How do particular materials and built environments affect what our students do and think?
Human meanings and decisions are important but are not the only things acting in any situation.	How might we encourage students to notice how materials influence situations in which they practise?
Emphasis is not on individual things and their characteristics, such as an individual doctor's skills or particular technologies, but on their *relationships* and what these produce.	How might students become more actively aware of these relationships and their effects?
Practices themselves are continuously changing *gatherings* of human and nonhuman elements that act on one another in unpredictable ways.	How do different elements act on one another to affect what happens, and how do these different interactions produce particular kinds of knowledge?
The *whole system* affects any particular practice as it continuously adapts and changes pattern.	How is a particular practice interconnected with and affected by other systems?
Uncertainty and *unpredictability* are assumed.	What might be inhibited in professional education dominated by predetermined curricula and planned objectives?

Source: From Fenwick [157, p. 361].

reconceptualising to be more innovative, integrative, and interprofessional. Currently, the literature on simulation focuses more on mastering clinical procedural skills rather than on responding to emergent practice scenarios or interprofessional situations. The authors describe the application of complexity theory, which can be considered a form of socio-materiality theory, to address the limitations of using simulations in medical education.

As discussed briefly earlier, clinical education may provide an excellent opportunity to use socio-material analysis to understand and reconfigure education in clinical contexts. It is clear that tools, materials, and physical space interact with and impact the actions of clinicians, medical staff, students, and even patients. All staff and students in a rotation could benefit greatly from a better understanding of these connections.

Adult Learning Principles

Since undergraduate and postgraduate medical students can be considered adult learners, adult learning theory has been frequently used to study medical education. Several authors [20, 170, 171] have provided excellent summaries of theory-building efforts in adult learning. They conclude that no single theory fares well when judged by the criteria of comprehensiveness (including all types of learning), practicality, and universality of application. They also assert that a phenomenon as complex as adult learning will probably never be adequately explained by a single theory.

Andragogy

Although many theoretical frameworks address adult education, few have actually been applied widely other than Knowles' *andragogy* [20]. The remainder of this section therefore focuses on andragogy, its implications for practice, and an example of its use in undergraduate medical education.

Knowles [20] first introduced the term 'andragogy' to North America, defining it as 'the art and science of helping adults learn'. Knowles did not present andragogy as an empirically based theory but simply as a set of four assumptions, to which a fifth and sixth were later added. The six assumptions underlying andragogy, as theorised by Knowles, are that: (i) adult self-concept is well-developed; (ii) adults bring considerable experience to learning; (iii) adults' readiness to learn depends on need; (iv) adults tend to have a problem-centred focus; (v) adults are generally internally motivated; and (vi) adults need to know why they need to know something [20].

Andragogy has its roots in humanistic psychology through the work of Maslow [172] and Rogers [173]. The foundation of andragogy is that the attainment of adulthood is marked by adults coming to view themselves as self-directed individuals. Knowles' 'model of assumptions' has given adult education a 'badge of identity' that distinguishes the field from other areas of education, for example, childhood schooling [174]. Bard has asserted that andragogy 'probably more than any other force, has changed the role of the learner in adult education and in human resource development' [175, p. xi]. However, it has also caused enormous controversy, debate, and criticism. Early criticism led Knowles to later modify his model by describing andragogy and pedagogy as a continuum, and suggesting that the use of both teaching methods is appropriate at different times in different situations, regardless of the learner's age [10].

Although andragogy is widely accepted and used in adult learning practice, it remains unproven empirically. Pratt [176] has noted that the empirical questions are still unanswered: 'We cannot say with any confidence that andragogy has been tested and found to be, as so many have hoped, either the basis for a theory of adult learning or a unifying concept for adult education' [176, p. 21]. This assertion is supported by Rachal, who concluded that 'the empirical literature examining the efficacy of andragogy remains, after over three decades, both inconclusive and beset by considerable variability in definition, resulting in differing approaches to andragogy's implementation' [177, p. 210]. Despite these criticisms, andragogical assumptions have guided educators in medicine and many other contexts for more than three decades, as their practical application seems to result in a respectful and effective process that enhances learning. Box 4.12 summarises these assumptions.

It is widely accepted that andragogy is not really a theory of *how* adults learn, the assumptions being merely descriptions of the adult learner [174]. Furthermore, even the assumptions have been questioned as prescriptions for

BOX 4.12 FOCUS ON: Andragogical assumptions [20]

1. *Self-concept*. Adults typically want to choose what they want to learn, when they want to learn it, and how they want to learn.
2. *Experience*. Adult learners have a wealth of life experiences that they bring with them into new learning experiences; they can contribute richness to learning from and with each other.
3. *Readiness to learn depends on need*. Adults are ready to learn when they see that what they need to know will help them to deal with life situations.
4. *Problem-centred focus*. Adults need to see the immediate application of learning, so they seek learning opportunities that will enable them to solve problems.
5. *Internal motivation*. Adults seek learning opportunities due to external motivators, but the more potent motivators, such as self-esteem, better quality of life, and self-actualisation, are internal.
6. *Adults need to know why they need to learn something*. Adults need to know how they will benefit from new knowledge, for example, to solve a problem or apply the knowledge immediately.

practice. The general critique is that andragogy lacks the fundamental characteristics of a science because of limited empirical evidence [174, 177]. Some argue that andragogy may, in time, become a theory through empirical studies of the assumptions. Merriam and Cafarella commented that, in his autobiography, Knowles described andragogy as less of a theory of adult learning than a 'model of assumptions about learning or a conceptual framework that serves as a basis for emergent theory' [62, p. 112]. At the least, andragogy captures general characteristics of adult learners and offers guidelines for planning instruction with learners who tend to be at least somewhat independent and self-directed [10, 20].

Implications for Educational Practice

There are several implications for practice derived from adult learning theories that have at their heart the fact that an adult's life situation is quite different from that of a child. Merriam [178] discusses differences between adults' and children's learning in three areas: context, learner, and learning process.

Context

Typically, being a learner is only one of several roles played concurrently by adults. Adults generally learn and function in multiple settings where situation-specific skills are required to resolve relevant problems.

Learner

Adults are self-directing due to their large reservoir of experience, the relationship of their readiness to learn to their social roles, their desire for knowledge that can be immediately applied to current relevant problems, and their internal motivation to learn.

Learning Process

Three non-cognitive factors have been shown to affect adult learning [29]:

- pacing
- meaningfulness
- motivation.

Pacing of learning, through deadlines or other external pressures, may adversely affect learning. Certainly, busy adults rely on deadlines to organise their deliverables; however, the learning required will often be superficial as the focus will be on meeting the deadline rather than on the learning. Also, adults tend to perform poorly on learning tasks that are not meaningful or that do not fall within their domain of interest.

A teacher using andragogical principles focuses on being a facilitator of learning rather than a transmitter of knowledge and an evaluator. Vella [179] lists 12 principles that should be addressed with adult learners. These are: a needs assessment of what is to be learned; a feeling of safety for the learner in the environment; sound relationships between the facilitator and the learners; the sequence of the content presented and its reinforcement; the use of praxis; establishment of respect for learners as decision makers; understanding of learners' ideas, feelings, and actions; immediacy of the learning; clearly established roles on the part of the facilitator and learners; the use of teamwork; the engagement of learners; and accountability.

These ideas can be formulated as a set of principles to guide adult learning activities. Knowles [180] drew seven principles from the assumptions of andragogy, which are presented here along with suggestions for their application (see Box 4.13).

BOX 4.13 HOW TO: Applying the principles of adult learning [180]

1. Establish an effective learning climate. Learners should be comfortable, both physically and emotionally. They should feel safe and free to express themselves without judgement or ridicule.
2. Involve learners in mutual planning of methods and curricular directions. Their involvement will help assure that collaboration occurs in the content and learning process. It will also increase the relevance of content and process to learners' needs.
3. Involve learners in diagnosing their own learning needs. Once again, this will help to ensure meaning and will trigger learners' internal (intrinsic) motivation. It will also promote learners' self-assessment and reflection, and effective integration of their learning.
4. Encourage learners to formulate their own learning objectives. The rationale for this is the same as for points 1–3, above. Learners are encouraged to take control of their learning.
5. Encourage learners to identify resources and devise strategies for using them to accomplish their objectives. This principle connects adult learning needs to practical resources for meeting their objectives and also provides motivation for using the resources for a specific and focused purpose.
6. Help learners carry out their learning plans. Expectation of success is a key element of motivation. Learners will become discouraged and lose their motivation if a learning task is too difficult. Also, too much pressure without support can inhibit learning.
7. Involve learners in evaluating their own learning. This is an essential step in a self-directed learning process that requires critical reflection on experience.

Connections

This chapter has presented 10 theoretical approaches to learning, each of which has the potential to inform our practice as medical educators. For each theoretical approach, the chapter has set out the underlying framework and principles and provided examples of the theory's application. Box 4.14 lists each theory, its main proponent, and one key reference.

The application of educational theory to practice has always been somewhat eclectic. This is not unusual in

BOX 4.14 FOCUS ON: Summary of learning theories

Theory	Proponent	Key references
Social cognitive theory	Albert Bandura	Bandura [11]
Reflective practice	Donald Schön	Schön [12]
Transformative learning	Jack Mezirow	Mezirow [13]
Self-directed learning	Philip Candy	Candy [14]
Experiential learning	David Kolb	Kolb [15]
Situated learning	Jean Lave	Lave and Wenger [16]
Communities of practice	Etienne Wenger	Wenger, Snyder, McDermott [122]
Constructivism	Lev Vygotsky	Gergen [142]
Socio-materiality	Wanda Orlikowski	Fenwick [156]
Adult learning principles	Malcolm Knowles	Knowles [180]

applied sciences such as education. To make this exposition of theory as useful as possible to our educational practice, it is helpful to consider the relationships among the theoretical frameworks, and the consistency of messages and themes that can be drawn from all of them to inform our medical education practice. Some of these common themes are provided here.

All theoretical frameworks view the learner as an active contributor to the learning process. In each of the theoretical approaches discussed here, the learner actively interacts with a changing, complex environment. The curriculum can no longer be viewed as something that is transmitted to, or acts upon, the students, be they undergraduate, postgraduate, or practising physicians. There is an important element of human agency. Moreover, in practice, the physician-learner is stimulated to learn through interactions in the practice environment.

The entire context of learning is more important than any one variable alone. The learning environment is complex. It includes learners, faculty, patients, colleagues, resources, and other workers. It is both the interacting and the independent effects of all these variables that result in a learning environment that is experienced by learners at all levels. Learning is accomplished both through direct experience and vicariously, and from many interactions in this complex system. Consequently, we must analyse as many factors in the environment as possible when planning, implementing, and evaluating our educational programmes.

Learning is integrally related to the solution and understanding of authentic problems. For adult learners, learning is most effective and motivating when it is relevant to the solution of authentic (i.e. real-life) needs or problems. This is obvious in the learning that occurs in reflective practice, where new learning is triggered by surprise encountered in a problem in practice. Experiential learning through authentic problems leads to ongoing mastery and competence. Learning around clinical problems, both in the clinic and in the classroom, represents learning to solve the authentic tasks of the profession and of future professional practice.

Individuals' past experience and knowledge are critical in learning, in actions, and in acquiring new knowledge. At all levels, learning must be connected to relevant experience or compatible with the learner's existing knowledge. Past experience and knowledge will affect perceptions of self-efficacy, which will, in turn, affect the choice of new experiences and goals. In practice, the new learning opportunities identified will depend substantially on the individual's existing experience and knowledge.

Learners' values, attitudes, and beliefs influence their learning and actions, and building learners' self-awareness in this area is important for their development. These values, attitudes, and beliefs are central to learners' willingness to attempt new actions. They affect virtually everything that learners think, as well as their interactions with mentors, peers, and patients. Self-awareness is critically important for the development of professional identity. Various processes exist to modify values, attitudes, and beliefs, such as reflective observation, perspective transformation, role modelling, and feedback on action.

Individuals as learners are capable of self-regulation, that is, of setting goals, planning strategies, and evaluating their progress. Adult learners are viewed as self-motivated and directed, pursuing those learning objectives that are relevant to personal goals. They are inherently self-regulating, and the process of reflection implies a learning that arises directly out of experience. In planning learning experiences we must regard these not as skills we have to teach students, but as skills and abilities that need to be developed and enhanced.

The ability to reflect on one's practice (performance) is critical to lifelong, self-directed learning. At the heart of all these theoretical approaches is the belief that we can learn from our experience, incorporating it into our existing knowledge and skills. This opportunity for reflection requires an early introduction to a systematic approach to facilitate reflection. Reflection is not merely description of experience, but analysis of it. It is not a natural and intuitive ability, and it must be developed through practice. It is critical to becoming an effective lifelong learner, as it also enables learners to develop and apply standards to their performance, decide what further learning needs to occur, and continue their learning over a professional lifetime.

Learning occurs both individually, and in collaboration with others. These theories support the learning that occurs dynamically, through interaction with others; these theories also recognise that learning can occur collectively as individuals share experiences, knowledge, and perspectives. The result of this is that knowledge and understanding are constructed collaboratively or mutually and that all members of the group can contribute to that growth.

Learning involves collaboration and interaction among humans, tools, materials, and physical space. Tools, materials and physical space interact with and impact the actions of clinicians, staff, learners, and patients. All individuals in a clinical rotation could benefit greatly from a better understanding of these connections so that learning can be more strategic.

Applying these theories to medical education requires reflection and practice. As medical educators, we can benefit from reading and considering relevant literature to better understand these theories, and from participation in a community of peers who have a common interest in this area. By practising the application of each theory, receiving feedback from learners and peer observers, and reflecting on practice, medical educators will continue to improve in their educational roles. Through their participation in the community, they will contribute to the construction of shared knowledge and understanding, and to new theories informed by practice. These activities will result in the enhancement of medical education.

Acknowledgements

I would like to acknowledge Dr Karen Mann for her friendship, support, and co-authorship of this chapter in the first two editions of this book. Her intelligence, sensitivity, and kindness will be missed by everyone who knew her.

I would also like to acknowledge Dr Penny Jennett, formerly from the Faculty of Medicine at the University of Calgary, for her co-authorship of the first version of the ASME monograph on which the section on Reflective Practice in this chapter is based.

References

1 Tripp, D. (1993). *Critical Incidents in Teaching. Developing Professional Judgement*. London: Routledge.
2 Rees, C. and Monrouxe, L. (2010). Theory in medical education: how do we get there? *Medical Education* 44: 334–339.
3 Albert, M., Hodges, B., and Regehr, G. (2007). Research in medical education: balancing service and science. *Advances in Health Sciences Education* 12 (1): 103–115.
4 Bordage, G. (2009). Conceptual frameworks to illuminate and magnify. *Medical Education* 43: 312–319.
5 Mann, K.M. (2011). Theoretical perspectives in medical education: past experience and future possibilities. *Medical Education* 45: 60–68.
6 Feldman, M.S. and Orlikowski, W.J. (2011). Theorizing practice and practicing theory. *Organization Science* 22 (5): 1240–1253.
7 Cuff, P.A. and Vanselow, N. ed. (2004). *Improving Medical Education: Enhancing the Behavioral and Social Science Content of Medical School Curricula*. Washington, DC: National Academies Press.
8 Tekian, A. and Harris, I. (2012). Preparing health professions education leaders worldwide: a description of masters-level programs. *Medical Teacher* 34: 52–58.
9 Foundation for Advancement of International Medical Education and Research. Master's Programs in Health Professions Education. http://www.faimer.org/resources/mastersmeded.html (accessed 27 June 2018).
10 Knowles, M., Holton, E.F. III, and Swanson, R.A. (2005). *The Adult Learner*, 6e. Oxford: Butterworth-Heinemann.
11 Bandura, A. (1986). *Social Foundations of Thought and Action. A Social Cognitive Theory*. Englewood Cliffs, NJ: Prentice-Hall.
12 Schön, D.A. (1983). *The Reflective Practitioner: How Professionals Think in Action*. New York, NY: Basic Books.
13 Mezirow, J. (1994). Understanding transformation theory. *Adult Education Quarterly* 44: 222–244.
14 Candy, P.C. (1991). *Self-Direction in Lifelong Learning*. San Francisco, CA: Jossey-Bass.
15 Kolb, D.A. (1984). *Experiential Learning: Experience as the Source of Learning and Development*. Englewood Cliffs, NJ: Prentice Hall.
16 Lave, J. and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. New York, NY: Cambridge University Press.
17 Wenger, E. (1998). *Communities of Practice: Learning, Meaning, and Identity*. New York, NY: Cambridge University Press.
18 Vygotsky, L.S. (1978). *Mind in Society*. Cambridge, MA: Harvard University Press.
19 Orlikowski, W.J. and Scott, S.V. (2008). Sociomateriality: challenging the separation of technology, work and organization. *The Academy of Management Annals* 2: 433–474.
20 Knowles, M.S. (1980). *The Modern Practice of Adult Education: From Pedagogy to Andragogy*, 2e. New York, NY: Cambridge Books.
21 Bandura, A. (1977). *Social Learning Theory*. Englewood Cliffs, NJ: Prentice-Hall.
22 Bandura, A. (1997). *Self-Efficacy: The Exercise of Control*. New York, NY: WH Freeman.
23 Bandura, A. (2006). Guide for constructing self-efficacy scales. In: *Self-Efficacy Beliefs of Adolescents*, vol. 5 (ed. F. Pajares and T. Urdan), 307–337. Greenwich, CT: Information Age Publishing.
24 Schön, D.A. (1987). *Educating the Reflective Practitioner: Toward a New Design for Teaching and Learning in the Professions*. San Francisco, CA: Jossey-Bass.
25 Schön, D.A. (1995). The new scholarship requires a new epistemology. *Change* 27 (6): 26–34.
26 Boud, D., Keogh, R., and Walker, D. ed. (1985). *Reflection: Turning Experience into Learning*. London: Kogan Page.
27 Branch, W. and Paranjape, A. (2002). Feedback and reflection: teaching methods for clinical settings. *Academic Medicine* 77: 1185–1188.
28 Moon, J. (1999). *Reflection in Learning and Professional Development*. London: Kogan Page.
29 Resnick, L.B. (1987). Learning in school and out. *Educational Researcher* 16: 13–20.
30 Atkins, S. and Murphy, K. (1993). Reflection: a review of the literature. *Journal of Advanced Nursing* 18 (8): 1188–1192.
31 Mann, K., Gordon, J., and MacLeod, A. (2009). Reflection and reflective practice in health professions education: a systematic review. *Advances in Health Sciences Education* 14: 595–621.
32 Jennett, P.A., Lockyer, J.M., Maes, W. et al. (1990). Providing relevant information on rural practitioners: a study of a medical information system. *Teaching and Learning in Medicine* 2 (4): 200–204.
33 Lockyer, J., Gondocz, S.T., and Thivierge, R.L. (2004). Knowledge translation: the role and place of practice reflection. *Journal of Continuing Education in the Health Professions* 24: 50–56.
34 Slotnick, H.B. (1996). How doctors learn: the role of clinical problems across the medical school-to-practice continuum. *Academic Medicine* 71: 28–34.

35 Crandall, S. (1993). How expert clinical educators teach what they know. *Journal of Continuing Education in the Health Professions* 13: 85–98.
36 Palmer, A., Burns, S., and Bulman, C. ed. (1994). *Reflective Practice in Nursing: The Growth of the Professional Practitioner*. Oxford: Blackwell Scientific Publications.
37 Epstein, R.M. and Hundert, E.M. (2002). Defining and assessing professional competence. *Journal of the American Medical Association* 287: 226–235.
38 Regehr, G. and Mylopolous, M. (2008). Maintaining competence in the field: learning about practice, from practice, in practice. *Journal of Continuing Education in the Health Professions* 28 (Suppl 1): S19–S23.
39 Branch, W.T. Jr., Frankel, R., Gracey, C.F. et al. (2009). A good clinician and a caring person: longitudinal faculty development and the enhancement of the human dimensions of care. *Academic Medicine* 84: 117–126.
40 Weissmann, P.F., Branch, W.T., Gracey, C.F. et al. (2006). Role-modeling humanistic behavior: learning bedside manner from the experts. *Academic Medicine* 81: 661–667.
41 Shapiro, J. and Talbot, Y. (1991). Applying the concept of the reflective practitioner to understanding and teaching family medicine. *Family Medicine* 23 (6): 450–456.
42 Clift, R., Houst, W.R., and Pugach, M. ed. (1990). *Encouraging Reflective Practice in Education: An Analysis of Issues and Programs*. New York, NY: Teachers College Press.
43 Westberg, J. and Jason, H. (2001). *Fostering Reflection and Providing Feedback. Helping Others Learn from Experience*. New York, NY: Springer Series on Medical Education. Springer.
44 Moon, J. (2004). Using reflective learning to improve the impact of short courses and workshops. *Journal of Continuing Education in the Health Professions* 24: 4–11.
45 Mezirow, J. (1991). *Transformative Dimensions of Adult Learning*. San Francisco, CA: Jossey-Bass.
46 Lockyer, J., Woloschuk, W., Hayden, A. et al. (1994). Faculty perceptions of their role as consultants to practising physicians. *Academic Medicine* 69: S13–S15.
47 Beecher, A., Lindemann, J., Morzinski, J., and Simpson, D. (1997). Use of the educator's portfolio to stimulate reflective practice among medical educators. *Teaching and Learning in Medicine* 9: 56–59.
48 Mezirow, J. (1990). *Fostering Critical Reflection in Adulthood*. San Francisco, CA: Jossey-Bass.
49 Roberts, C. and Stark, P. (2008). Readiness for self-directed change in professional behaviours: factorial validation of the Self-reflection and Insight Scale. *Medical Education* 42: 1054–1063.
50 Sobral, D. (2005). Mindset for reflective learning: a revalidation of the reflection-in-learning scale. *Advances in Health Sciences Education* 10: 303–314.
51 Wald, H., Borkan, J.M., Taylor, J.S. et al. (2012). Fostering and evaluating reflective capacity in medical education: developing the REFLECT rubric for assessing reflective writing. *Academic Medicine* 87: 41–50.
52 Vivekananda-Schmidt, V., Marshall, M., Stark, P. et al. (2011). Lessons from medical students' perceptions of learning reflective skills: a multi-institutional study. *Medical Teacher* 33: 846–850.
53 Beecher, A., Lindemann, J., Morzinski, J., and Simpson, D. (1997). Use of the educator's portfolio to stimulate reflective practice among medical educators. *Teaching and Learning in Medicine* 9: 56–59.
54 Cranton, P. (1994). *Understanding and Promoting Transformative Learning: A Guide for Educators of Adults*. San Francisco, CA: Jossey-Bass.
55 Taylor, E.W. and Cranton, P. (2012). *The Handbook of Transformative Learning: Theory, Research, and Practice*. San Francisco, CA: Jossey-Bass.
56 Janik, D.S. (2005). *Unlock the Genius Within*. Lanham, MD: Rowman and Littlefield Education.
57 Taylor, E.W. (2008). Transformative learning theory. *New Directions for Adult and Continuing Education* 119: 5–15.
58 Brookfield, S.D. (1990). *The Skillful Teacher: On Technique, Trust and Responsiveness in the Classroom*. San Francisco, CA: Jossey-Bass.
59 Maslow, A.H. (1968). *Toward a Psychology of Being*, 2e. New York, NY: Van Nostrand Reinhold.
60 Brockett, R.G. and Hiemstra, R. (1991). *Self-Direction in Adult Learning: Perspectives on Theory, Research and Practice*. New York, NY: Routledge.
61 Knowles, M.S. (1975). *Self-Directed Learning: A Guide for Learners and Teachers*. New York, NY: Association Press.
62 Merriam, S.B. and Caffarella, R.S. (1999). *Learning in Adulthood: A Comprehensive Guide*, 2e. San Francisco, CA: Jossey-Bass.
63 Grow, G. (1991). Teaching learners to be self-directed: a stage approach. *Adult Education Quarterly* 41: 125–149.
64 Hammond, M. and Collins, R. (1991). *Self-Directed Learning: Critical Practice*. New York, NY: Nichols/GP Publishing.
65 Gugliemino, L.M. (1977). Development of the Self-Directed Learning Readiness Scale. Unpublished doctoral dissertation. University of Georgia, Athens, GA, USA.
66 Oddi, L.F. (1986). Development and validation of an instrument to identify self-directed continuing learners. *Adult Education Quarterly* 36: 97–107.
67 Roberts, C. and Stark, P. (2008). Readiness for self-directed change in professional behaviours: factorial validation of the Self-reflection and Insight Scale. *Medical Education* 42: 1054–1063.
68 Eva, K.W. and Regehr, G. (2005). Self-assessment in the health professions: a reformulation and research agenda. *Academic Medicine* 80: 547–554.
69 Sargeant, J., Armson, H., Chesluk, B. et al. (2010). The processes and dimensions of informed self-assessment: a conceptual model. *Academic Medicine* 85: 1212–1220.
70 Brydges, R. and Butler, D. (2012). A reflective analysis of medical education research on self-regulation in learning and practice. *Medical Education* 46: 71–79.
71 Sandars, J. and Cleary, T.J. (2011). Self-regulation theory: applications to medical education: AMEE guide. *Medical Teacher* 33: 875–886.
72 Davis, D.A., Mazmanian, P.E., Fordis, M., Van Harrison, R., Thorpe, K.E. and Perrier, L. (2006). Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *Journal of the American Medical Association* 296: 1094–102.
73 Hacker, D.J., Bol, L., and Keener, M.C. (2008). Metacognition in education: a focus on calibration. In: *Handbook of Memory and Metacognition* (ed. J. Dunlosky and R. Bjork), 429–456. Mahwah, NJ: Erlbaum Publishers.
74 Butler, D.L. and Winne, P.H. (1995). Feedback and self-regulated learning: a theoretical synthesis. *Review of Educational Research* 65: 245–281.
75 Cook, D.A. and Artino, A.R. Jr. (2016). Motivation to learn: an overview of contemporary theories. *Medical Education* 50: 997–1014.
76 Keller, J.M. (2009). *Motivational Design for Learning and Performance: The ARCS Model Approach*. Berlin, Germany: Springer Science & Business Media.
77 Keller, J.M. (1987). Development and use of the ARCS model of instructional design. *Journal of Instructional Development* 10 (3): 2–10.
78 Entwistle, N. and Ramsden, R. (1983). *Understanding Student Learning*. London: Croom Helm.
79 Miflin, B., Campbell, C.B., and Price, D.A. (1999). A lesson from the introduction of a problem-based graduate entry course: the effects and different views of self-direction. *Medical Education* 33: 801–807.
80 Biggs, J. (2003). *Teaching for Quality Learning at University*, 2e. Bury St. Edmonds: Open University Press.
81 McClusky, H.Y. (1970). An approach to a differential psychology of the adult potential. In: *Adult Learning and Instruction* (ed. S.M.

Grabowski), 80–95. Syracuse, NY: ERIC Clearinghouse on Adult Education.
82 Parboosingh, J. (1996). Learning portfolios: potential to assist health professionals with self-directed learning. *Journal of Continuing Education in the Health Professions* 16: 75–81.
83 Smith, S., Kakarala, R.R., Talluri, S.K. et al. (2011). Internal medicine residents' acceptance of self-directed learning plans at the point of care. *Journal of Graduate Medical Education* 3: 425–428.
84 Bravata, D.M.T., Huot, S., Abernathy, H.S. et al. (2003). The development and implementation of a curriculum to improve clinicians' self-directed learning skills: a pilot project. *BMC Medical Education* 3: 7.
85 Lewin, K. (1951). *Field Theory in Social Sciences: Selected Theoretical Papers*. New York, NY: Harper and Row.
86 Dewey, J. (1938). *Experience and Education*. New York, NY: Kappa Delta Phi, Touchstone.
87 Piaget, J. (1971). *Psychology and Epistemology*. Harmondsworth, UK: Penguin Books.
88 Fenwick, T. (2000). Conceptions of experiential learning: a review of the five contemporary perspectives on cognition. *Adult Education Quarterly* 50 (4): 243–272.
89 Holzer, S.M. and Andruet, R.H. (1998). A multimedia workshop learning environment for statistics. Workshop presented at ASEE Conference and Exposition, Seattle, WA, and Virginia Polytechnic Institute and State University, Blacksburg, VA.
90 Lee, P. and Caffarella, R.S. (1994). Methods and techniques for engaging learners in experiential learning activities. *New Directions for Adult and Continuing Education* 62: 43–54.
91 Fry, R. and Kolb, B. (1979). Experiential learning theory and learning experiences in liberal arts education. *New Directions for Experiential Learning* 6: 79–92.
92 Caffarella, R.S. (1992). *Psychological Development of Women: Linkages to the Practice of Teaching and Learning in Adult Education*. Columbus, OH: ERIC Clearinghouse on Adult, Career and Vocational Education.
93 Rogoff, B., Matusov, E., and White, C. (1996). Models of teaching and learning: participation in a community of learners. In: *The Handbook of Education and Human Development: New Models of Learning, Teaching, and Schooling* (ed. D.R. Olsen and N. Torrance), 388–415. Oxford: Blackwell.
94 Cantillon, P., D'Eath, M., De Grave, W. and Dornan, T. (2016). How do clinicians become teachers? A communities of practice perspective. *Advances in Health Sciences Education* 21: 991–1008.
95 Rogoff, B. (2003). *The Cultural Nature of Human Development*. New York, NY: Oxford University Press.
96 Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational Researcher* 27: 4–13.
97 Lave, J. and Wenger, E. (2002). Legitimate peripheral participation in communities of practice. In: *Supporting Lifelong Learning. Perspectives on Learning* (ed. R. Harrison, F. Reeve, A. Hanson and J. Clarke), 1111–1126. London: Routledge Farmer.
98 Borg, E. (2003). Discourse community. *The ELT Journal* 57: 398–400.
99 Steven, K., Wenger, E., Boshuizen, H. et al. (2014). How clerkship students learn from real patients in practice settings. *Academic Medicine* 89: 469–476.
100 Palinscar, A.S. (1998). Social constructivist perspectives on teaching and learning. *Annual Review of Psychology* 49: 345–375.
101 Brown, J.S., Collins, A., and Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher* 18: 32–42.
102 Eraut, M. (2000). Non-formal learning and tacit knowledge in professional work. *British Journal of Educational Psychology* 70: 113–136.
103 Reber, A. (1995). *Implicit Learning and Tacit Knowledge: An Essay on the Cognitive Unconscious*. New York, NY: Oxford University Press.
104 Polyani, M. (1958). *Personal Knowledge: Towards a Post-Critical Philosophy*. London: Routledge and Kegan Paul.
105 Cooke, M., Irby, D., and O'Brien, B. (2010). *Educating Physicians: A Call for Reform of Medical School and Residency*. San Francisco, CA: Jossey-Bass.
106 Hafferty, F. and Franks, R. (1994). The hidden curriculum, ethics teaching and the structure of medical education. *Academic Medicine* 69: 861–871.
107 Bates, J., Konkin, J., Suddards, C. et al. (2013). Student perceptions of assessment and feedback in integrated longitudinal clerkships. *Medical Education* 47: 362–374.
108 Coulehan, J. and Williams, P.C. (2001). Vanquishing virtue: the impact of medical education. *Academic Medicine* 76: 598–605.
109 Dornan, T., Boshuizen, H., King, N., and Scherpbier, A. (2007). Experience-based learning: a model linking the processes and outcomes of medical students' workplace learning. *Medical Education* 41: 84–91.
110 Collins, A., Brown, J.S., and Newman, S. (1989). Cognitive apprenticeship: teaching the crafts of reading, writing and mathematics. In: *Knowing, Learning and Instruction: Essays in Honor of Robert Glaser* (ed. L.B. Resnick), 453–494. Mahwah, NJ: Lawrence Erlbaum Associates.
111 Barab, S.A., Barnett, M., and Squire, K. (2002). Developing an empirical account of a community of practice: characterizing the essential tensions. *The Journal of the Learning Sciences* 11: 489–542.
112 Wenger, E. and Wenger-Trayner, B. (2015). Communities of Practice: A Brief Introduction. http://wenger-trayner.com/introduction-to-communities-of-practice (accessed 25 March 2017).
113 Phipps, D. (2012). Knowledge mobilisation and why does it matter to universities? *Guardian Professional*, 9th March 2012. www.guardian.co.uk/higher-education-network/blog/2012/mar/09/introduction-to-knowledge-mobilisation (accessed 2 December 2012).
114 Caplan, N. (1978). The two communities theory and knowledge utilization. *American Behavioral Scientist* 22: 459–470.
115 Oftek, E. and Sarvary, M. (2002). *Knowledge Exchange and Knowledge Creation: Should the Emphasis Shift in a Competitive Environment* (working paper). Fontainebleau, France: INSEAD.
116 World Health Organization (WHO) (2006). Bridging the 'Know–Do' Gap in Global Health. http://www.who.int/workforcealliance/knowledge/resources/knowdo_gap/en (accessed 27 March 2017).
117 Canadian Health Services Research Foundation (CHSRF) (2003). *The Theory and Practice of Knowledge Brokering in Canada's Health System. A Report Based on a CHSRF Consultation and a Literature Review*. Ottawa, ON: CHSRF.
118 Canadian Institutes of Health Research (CIHR) (2004). *Knowledge Translation Strategy: 2004–2009: Innovation in Action*. Ottawa, ON: CIHR.
119 Davis, D.A., Thomson, M.A., Oxman, A.D., and Haynes, R.B. (1992). Evidence for the effectiveness of CME. A review of 50 randomized controlled trials. *Journal of the American Medical Association* 268: 1111–1117.
120 Triggs, P. and John, P. (2004). From transaction to transformation: information and communication technology, professional development and the formation of communities of practice. *Journal of Computer Assisted Learning* 20: 426–439.
121 Millen, D.R., Fontaine, M.A., and Muller, M.J. (2002). Understanding the benefits and costs of communities of practice. *Communications of the ACM* 45 (4): 69–73.
122 Wenger, E., Snyder, W., and McDermott, R. (2002). *Cultivating Communities of Practice: A Guide to Managing Knowledge*. Cambridge, MA: Harvard Business School Press.
123 Ensley, M.S.C., Kirkegaard, M., and Linares, A. (2005). Working together: communities of practice in family medicine. *Family Practice Management* 12 (1): 28–32.
124 Henri, F. and Pudelko, B. (2003). Understanding and analyzing activity and learning in virtual communities. *Journal of Computer-Assisted Learning* 19: 474–487.

125 Kimble, C., Hildreth, P., and Wright, P. (2001). Communities of practice: going virtual. In: *Knowledge Management and Business Model Innovation* (ed. Y. Malhotra), 220. Hershey, PA: Idea Group Publishing.
126 Parboosingh, J.T. (2002). Physician communities of practice: where learning and practice are inseparable. *Journal of Continuing Education in the Health Professions* 22: 230–236.
127 Hung, D. and Tan, S.C. (2004). Bridging between practice fields and real communities through instructional technologies. *International Journal of Instructional Media* 31 (2): 1–8.
128 Richardson, B. and Cooper, N. (2003). Developing a virtual interdisciplinary research community in higher education. *Journal of Interprofessional Care* 17 (2): 173–182.
129 White, C.B. and Thomas, A.M. (2004). Students assigned to community practices for their pediatric clerkship perform as well or better on written examinations as students assigned to academic medical centers. *Teaching and Learning in Medicine* 16: 250–254.
130 Bruner, J. (1966). *Toward a Theory of Instruction*. Cambridge, MA: Harvard University Press.
131 Amineh, R.J. and Asi, H.D. (2015). Review of constructivism and social constructivism. *Journal of Social Sciences, Literature and Languages* 1 (1): 9–16.
132 Piaget, J. (1977). *Intellectual Evolution from Adolescence to Adulthood*. Cambridge: Cambridge University Press.
133 Fosnot, C.T. (1989). *Enquiring Teachers, Enquiring Learners: A Constructivist Approach for Teaching*. New York, NY: Teachers College Press.
134 Duffy, T.M. and Jonassen, D.H. (1991). Constructivism: new implications for instructional technology. *Educational Technology* 31 (5): 7–12.
135 Piaget, J. (1970). *Structuralism*. New York, NY: Harper and Row.
136 Hyslop-Margison, E.J. and Strobel, J. (2007). Constructivism and education: misunderstandings and pedagogical implications. *The Teacher Educator* 43 (1): 72–86.
137 Schunk, D.H. (2014). *Learning Theories: An Educational Perspective*, 6e. Essex: Pearson.
138 Gredler, M.E. (1997). *Learning and Instruction: Theory into Practice*, 3e. Upper Saddle River, NJ: Prentice-Hall.
139 Ernest, P. (1998). *Social Constructivism as a Philosophy of Mathematics*. Albany, NY: State University of New York Press.
140 McMahon, M. (1997). Social Constructivism and the World Wide Web – A Paradigm for Learning. Paper presented at the ASCILITE conference, Perth, Australia. http://www.ascilite.org/conferences/perth97/papers/Mcmahon/Mcmahon.html (accessed 15 July 2017).
141 Derry, S.J. (1999). A fish called peer learning: searching for common themes. In: *Cognitive Perspectives on Peer Learning* (ed. A.M. O'Donnell and A. King), 197–211. Mahwah, NJ: Lawrence Erlbaum Associates.
142 Gergen, K.J. (1995). Social construction and the educational process. In: *Constructivism in Education* (ed. L.P. Steffe and J. Gale), 17–39. Mahwah, NJ: Lawrence Erlbaum Associates.
143 Kuusisaari, H. (2014). Teachers at the zone of proximal development – collaboration promoting or hindering the development process. *Teaching and Teacher Education* 43: 46–57.
144 Wass, R. and Golding, C. (2014). Sharpening a tool for teaching: the zone of proximal development. *Teaching in Higher Education* 19 (6): 671–684.
145 Hoover, W.A. (1996). The practice implications of constructivism. *SEDL Letter* 9 (3): 1–2.
146 Clements, D.H. (1997). (Mis?)constructing constructivism. *Teaching Children Mathematics* 4 (4): 198–200.
147 Brownstein, B. (2001). Collaboration: the foundation of learning in the future. *Education* 122 (2): 240.
148 Rhodes, L.K. and Bellamy, G.T. (1999). Choices and consequences in the renewal of teacher education. *Journal of Teacher Education* 50 (1): 17–26.
149 Di Vesta, F. (1987). The cognitive movement and education. In: *Historical Foundations of Education Psychology* (ed. J. Glover and R. Royce), 203–233. New York, NY: Plenum Press.
150 Bauersfeld, H. (1995). "Language games" in the mathematics classroom: their function and their effects. In: *The Emergence of Mathematical Meaning: Interaction in Classroom Cultures* (ed. P. Cobb and H. Bauersfeld), 271–289. Mahwah, NJ: Lawrence Erlbaum Associates.
151 Brown, R.G. (1999). Middle school social studies in the cognitive revolution. *The Clearing House* 72 (6): 327–331.
152 Johnson, D.W. and Johnson, R.T. (1989). *Cooperation and Competition Theory and Research*. Edina, MN: Interaction Book Co. Publishing.
153 Panitz, T. (1999). Benefits of cooperative learning in relation to student motivation. In: *Motivation From Within: Approaches for Encouraging Faculty and Students to Excel, New Directions for Teaching and Learning* (ed. M. Theall), 59–68. San Francisco, CA: Josey-Bass.
154 Laal, M. and Ghodsi, S.M. (2012). Benefits of collaborative learning. *Procedia – Social and Behavioral Sciences* 31: 486–490.
155 MacLeod, A. et al. (2015). Sociomateriality: a theoretical framework for studying distributed medical education. *Academic Medicine* 90: 1451–1456.
156 Fenwick, T. (2015). Sociomaterialty and learning: a critical approach. In: *The SAGE Handbook of Learning* (ed. D. Scott and A. Hargreaves), 83–93. Thousand Oaks, CA: SAGE.
157 Fenwick, T. and Dahlgren, M.A. (2015). Towards socio-material approaches in simulation-based education: lessons from complexity theory. *Medical Education* 49: 359–367.
158 Frambach, J.M., Driessen, E.W., and van der Vleuten, C.P.M. (2014). Using activity theory to study cultural complexity in medical education. *Perspectives on Medical Education* 3 (3): 190–203.
159 Johnson, J. and Dornan, T. (2015). Activity theory: mediating research in medical education. In: *Researching Medical Education*, Chap. 7 (ed. J. Cleland and S.J. Durning), 93–103. John Wiley and Sons. http://onlinelibrary.wiley.com/doi/10.1002/9781118838983.ch9/summary (accessed 25 March 2017).
160 Reid, A.M., Ledger, A., Kilminster, S., and Fuller, R. (2015). Can the tools of activity theory help us in advancing understanding and organisational change in undergraduate medical education? *Advances in Health Sciences Education* 20: 655–668.
161 Kaptelinin, V., Kuutti, K., and Bannon, L. (1995). Activity Theory: Basic Concepts and Applications. A summary of a tutorial given at the East West HCI 95 Conference. http://www.ulfblanke.com/downloads/activity_theory/kaptelinin-basics.pdf (accessed 26 March 2017).
162 Ajjawi, R. and Bearman, M. (2012). Sociomateriality matters to family practitioners as supervisors. *Medical Education* 46: 1145–1147.
163 Leonardi, P.M. (2013). Theoretical foundations for the study of sociomateriality. *Information and Organization* 23 (2): 59–76.
164 Orlikowski, W.J. (2007). Sociomaterial practices: exploring technology at work. *Organization Studies* 28 (9): 1435–1448.
165 Carlile, P.R., Nicolini, D., Langley, A., and Tsoukas, H. ed. (2013). *How Matter Matters: Objects, Artifacts, and Materiality in Organization Studies*, vol. 3. Oxford: Oxford University Press.
166 Orlikowski, W.J. (2009). The sociomateriality of organizational life: considering technology in management research. *Cambridge Journal of Economics* 34 (1): 125–141.
167 Anderson, T., Rourke, L., Garrison, R., and Archer, W. (2001). Assessing teaching presence in a computer conferencing context. Journal of Asynchronous Learning Networks http://hdl.handle.net/2149/725 (accessed 14 March 2017).
168 Garrison, R. (2009). Implications of online learning for the conceptual development and practice of distance education. *Journal of Distance Education* 23 (2): 93–104.
169 Garrison, R., Anderson, T., and Archer, W. (2001). Critical thinking, cognitive presence and computer conferencing in distance education. *American Journal of Distance Education* 15 (1): 7–23.

170 Cross, K.P. (1981). *Adults as Learners*. San Francisco, CA: Jossey-Bass.
171 Darkenwald, G. and Merriam, S.B. (1982). *Adult Education: Foundations of Practice*. New York, NY: Harper and Row.
172 Maslow, A.H. (1970). *Motivation and Personality*, 2e. New York, NY: Harper and Row.
173 Rogers, C. (1968). *Freedom to Learn*. Columbus, OH: Charles E Merrill.
174 Hartree, A. (1981). Malcolm Knowles' theory of andragogy: a critique. *International Journal of Lifelong Education* 3 (3): 203–210.
175 Bard, R. (1984). Foreword. In: *Andragogy in Action* (ed. M.S. Knowles), xi. San Francisco, CA: Jossey-Bass.
176 Pratt, D.D. (1993). Andragogy after 25 years. In: *An Update on Adult Learning Theory* (ed. S.B. Merriam), 15–23. San Francisco CA: Jossey-Bass.
177 Rachal, J.R. (2002). Andragogy's detectives: a critique of the present and a proposal for the future. *Adult Education Quarterly* 52 (3): 210–227.
178 Merriam, S.B. (1996). Updating our knowledge of adult learning. *Journal of Continuing Education in the Health Professions* 16: 136–143.
179 Vella, J. (2002). Quantum learning: teaching as dialogue. *New Directions for Adult and Continuing Education* 93: 78–84.
180 Knowles, M.S. (1984). *Andragogy in Action: Applying Modern Principles of Adult Learning*. San Francisco, CA: Jossey-Bass.

5 Principles of Curriculum Design

Janet Grant
CenMEDIC (Centre for Medical Education in Context), London, UK
The Open University, Milton Keynes, UK
University College London Medical School, London, UK
World Federation for Medical Education, Ferney-Voltaire, France

KEY MESSAGES

- A curriculum is an ideological, social, and aspirational statement that should reflect local circumstances and needs.
- A curriculum in medical education is made up of all the experiences learners will have that enable them to reach the intended outcomes in basic and clinical sciences, and in clinical skills at undergraduate and postgraduate levels.
- A curriculum statement should enable learners, teachers, and managers to know and fulfil their obligations in relation to the programme or course. It should describe intended learner achievements, content to be covered (the syllabus), teaching, learning, supervision, feedback and assessment processes, entry requirements, and how the programme will be organised.
- The curriculum design process should first ask 'what is the purpose?' of the programme or course.
- The way in which a curriculum for medical education is constructed depends on the designers' views about how students learn, how medicine is practised, issues of social responsibility and accountability, the role of the knowledge base, professional values, resources, and health service development.
- Curriculum structure and process should take into account what is known from cognitive and educational psychology. A contextual curriculum may also reflect issues raised by empiricists as well as by philosophical, sociocultural, and critical theories.
- There is no body of evidence that supports one best choice for framing a curriculum as a whole or any of its parts. A curriculum should simply be fit for the purpose and context of its place and day.

Introduction

My bookshelves carry an ever-expanding history of medical education, so I chose some books at random to determine whether the years had produced different ideas about curriculum design. Partially they had, and partially they had not. For instance, in 1961 [1], curriculum debates centred on instructional skills and ideas about how students learn. The curriculum was to be made up of objectives and experiences with relatively traditional divisions of content, but all based on the health needs of society, the philosophy of scientific thinking, and the professional characteristics of physicians. In 1972 [2], the advice was to define aims and objectives in behavioural terms (not so different from today's preoccupation with competencies, perhaps), and also that curricula should offer what the student and community require – not what is convenient for medical school staff to offer. Teachers were advised to try to integrate their teaching more effectively and give students some choice over what they learn. By 1982 [3] and 1983 [4], a systems approach to educational design was advocated, with an emphasis on teaching methods aimed at delivering the learning objectives in the belief that active student involvement in learning was a likely effective strategy. By 1989 [5], it seemed reasonable to devote entire books to the question of how the curriculum might be structured to facilitate learning appropriate to clinical practice. In today's publications, we find the theme of social accountability [6] somewhat incongruously placed alongside the equally powerful, but contradictory, rhetoric of the 'post-colonial dilemma' of globalisation [7].

Education in any field reflects wider social trends and values, and to understand what is happening in curriculum philosophy and design in medical education, we need to look beyond that sphere to the role of health care in society and the debates that occupy education in general. Such a study of curriculum shows us that underpinning ideas develop according to economic and social imperatives but continue to have roots in previous thinking. Many 'current' ideas are not new; they also often lack supporting evidence, for example:

- a drive for integration
- a focus on students' learning rather than teachers' teaching

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

- a need for teachers to learn how to do their job well
- an instrumental focus on outcomes (expressed as objectives or competencies)
- a responsibility to respond to societal needs
- an obligation to prepare students for professional practice.

The same ideas can give rise to different curriculum designs and to different processes of reaching that design. The design principles guiding our educational efforts (see Chapter 6 for more on instructional design) are based on the professional choices that curriculum designers make. Those choices are informed by the theories, the beliefs, the dominant rhetoric and social conditions of the day, and by the values and experiences of the medical profession doing its best to produce the next generation of doctors fit for its changing purpose.

Curriculum design in medical education is an arena in which many ideological battles are fought. There are many differing views about, for example, what medical students should learn, how they should learn it, what qualities they should develop, where the science base stands, where skills of communication and examination should be acquired, how long it should all take, and whether to frame their task in terms of outcomes or competencies. The call for more 'professional' content – management, leadership, professionalism, teaching skills – to be included in the curriculum persists [8], alongside an on-going (but largely unevidenced) argument about curriculum overload.

Equally, there are many diverse views about how a curriculum should be developed and structured. Incontrovertible research findings on which to base decisions in education are hard to find; education is a social science, and in social science objective truth is elusive [9]. This means that vogues in curriculum design ebb and flow in response to the dominant concerns of society and those of the professions, just as they ebb and flow in relation to developments in teaching and learning methods, curriculum evaluation, and even the assessment of learning. Eisner [10] talks of 'curriculum ideologies' and 'the value premises from which decisions about practical educational matters are made'. These can be very strong, so that, as Toohey says, 'alternative views are literally "unthinkable" ' [11]. And so zealousness for a particular curriculum model develops, as she says, on beliefs that are 'so commonly held in the discipline, that they are accepted without question'. Integration, learner-centredness, and ideas about adult learning, none of which have agreed definitions or a basis in evidence, come into this category of belief.

And in all this, there is an enduring truth, concisely expressed by Michael Apple [12], that a curriculum can never be a neutral statement, it is the expression a particular set of values, beliefs, and aspirations derived from the immediate local political, cultural, professional, and social surroundings. This is a particularly significant consideration as globalisation threatens to homogenise curricula to meet criteria and standards unrelated to their local contexts.

To complicate things further, predominant concerns in curriculum design at the basic (medical school), postgraduate, and continuing education levels are very different. In medical school, we have students who have everything to learn and a school that has the responsibility and opportunity to ensure that they do, and the right to call on the student's time and fill it with activities that reflect the school's view of curriculum. At the postgraduate level, learning occurs in the context of clinical practice; the student now is a young doctor who still has much to learn and examinations to pass, but also has clinical duties to fulfil. Much of the learning is dependent on the clinical work that is experienced, and teachers and curriculum planners, quite rightly, only have limited power to organise the learning of a postgraduate trainee.

At the stage of continuing professional development (discussed in detail in Chapter 19), every doctor has become an autonomous professional, each with a unique history of experience and many with unique learning intentions arising out of their professional practice. For most, there is little protected time and minimal finance for learning. At this point, the idea of a set curriculum might seem to be an unworkable irrelevance. This, in turn, renders the standardised assessment of practising physicians highly problematic [13]. Instead, we might simply guide senior doctors to identify their own learning needs, design their own learning, and reinforce that in their own practice [14].

In this chapter, curriculum design is discussed only as it applies to undergraduate medical education and postgraduate training. Enduring principles are presented that will stand the test of time, changes of fashion, and the many different contexts across the world in which medical curricula are applied. The principles outlined should be flexible enough to yield different types of curricula in different hands. The curriculum must be appropriate to its context and not a slave to abstract, if well meaning, intent. Effective education must be contextual, rooted in its own culture and conditions.

What is a Curriculum?

A curriculum can be thought of as a managerial, ideological, and planning document that has three major components: *structure*, *content*, and *process*. A curriculum presents a reasoned picture of the subject to be studied and defines the teaching processes, learning processes, and intended outcomes of that study. All curriculum decisions must be made on the basis of a prior statement of vision or mission or values. And that statement must be made for the local context. General statements are of limited value. Contextual statements expressed in concrete terms will drive useful change at all levels.

A curriculum, then, is much more than a 'syllabus', which is just a simple listing of the content or topics of the course or programme. The curriculum is a powerful tool and therefore, both within institutions and across the wider society, is often the focus of battles for power and control over structure, content, and process.

Although much is written about curriculum, definitions are few and far between. As with all social science, each practitioner or researcher must be clear about their own perspective. UNESCO [15] for instance refers to the *intended* curriculum, the *written* curriculum, the *official* curriculum,

BOX 5.1 Definition of curriculum

A curriculum is a managerial, ideological, and planning document that should:

- Tell the learner exactly what to expect – including entry requirements, length and organisation of the course or programme and its flexibilities, the assessment system, and methods of student support.
- Advise the teacher what to do to deliver the content and support the learners in their task of personal and professional development.
- Help the institution to set appropriate assessments of student learning and implement relevant evaluations of the educational provision.
- Tell society how the school is executing its responsibility to produce the next generation of doctors appropriately.

the *implemented* curriculum, the *achieved* curriculum, the *learned* curriculum, the *hidden* curriculum, the curriculum *framework*, the curriculum *system*, and curriculum *processes*. All these terms reflect a different way of looking at a curriculum, its purpose, and the values that underpin that perception. Given all this, Box 5.1 suggests a broad working definition of curriculum.

The definition in Box 5.1 leaves open all the content and process decisions that must be made on the basis of the local health profile, health care service needs, educational culture, and resource availability. Curricula that comply with this definition should offer all stakeholders a clear description of requirements and expectations that are suited to their own needs and circumstances. The definition allows resolution of the tension between increasing prominence of ideas of globalisation and the imperative for a curriculum that reflects local needs. Competing views about the best way of structuring a curriculum are largely theoretical, until decided on the basis of local needs and resources; a curriculum must be contextual [16].

Curriculum Design

Educators and philosophers have addressed the question of what to teach and how to teach at least since Plato wrote *The Republic* in about 360 BCE. It might seem surprising, then, that it is only relatively recently that curriculum design has become a topic of debate in its own right, although the initial concerns about the nature of curricula arose with the advent of mass schooling in the late nineteenth century [17]. Until that point, curricula were defined by elite and specialist groups, and a curriculum statement (whether explicit or implicit) might contain only the content to be studied, and perhaps the time to be taken and the teaching method to be used.

Prior to the 1960s, curriculum change was best described as unplanned 'drift' [18], although, even before that time, curriculum ideology was informed by dominant social ideologies and imperatives. For example, the need to reconstruct the world after the Second World War certainly gave rise to the 'management by objectives' movement and so to objectives-based curricula, in the race to normalise as quickly and efficiently as possible. But from that point, Kelly [19] records that educationalists recognised the need for planned innovation to keep pace with societal changes, while maintaining standards and values and taking advantage of new theoretical underpinnings. At the same time, the idea of the curriculum as a total description of the intentions, mechanisms, context, and outcomes of education took hold.

Nowadays, however, as ideas about the process and management of education have developed, a curriculum statement would be regarded as satisfactory only if it addressed the wider experience of the learner and the context of learning as well as the content and quality control of the enterprise. The curriculum should guide the learner, the teacher, and educational managers. At the same time, it should leave room in its implementation for the creative and individual professionalism of the teacher and for the individual preferences of the learner, given that both are clear about what is to be achieved.

Curriculum design, as discussed later in this chapter, is subject to a wide variety of influences in relation to the profession of medicine, the health care service, and society as a whole. Each curriculum design team must decide for themselves which approach they will use (see Box 5.2), although these choices will often be influenced by explicit external guidance or standards.

BOX 5.2 WHERE'S THE EVIDENCE: For comparative curriculum design

Although there is much research published about different curriculum models and teaching and learning strategies, there is no evidence to suggest that there is 'one best way'. This is partly because a curriculum is made up of many components and there is little evidence to suggest that even for any one of these components there is a preferred choice for all circumstances.

Curricula have many different specific purposes, operate in different contexts, and therefore have many different designs. Their effectiveness can only be judged against their intended purposes. And few share exactly the same purpose, beyond intending to produce safe and responsible doctors. These differences in teaching, learning, culture, resources, and opportunities make comparative or controlled research almost impossible.

Each curriculum designer must decide on the purpose of the curriculum and then search the literature for relevant evidence about the likely effect of each curriculum component in serving that purpose. Convincing evidence in education is often difficult to find. So curriculum designers might best rely on their professional judgement and values and should always seek to gather their own evidence about the effects of their own preferred curriculum design and conduct their own consultations and risk analyses locally.

Curriculum Standards

In some countries, curricula or curriculum standards are set by the state; in others they are set by regulators and professional bodies. Some regulators offer guidance in relation to curriculum outcomes and a few set actual standards for a curriculum, how it should be stated, what its component parts should be, and how it should be developed, implemented, and used. Usually, curriculum standards address more than the syllabus content of the course or programme.

For example, in the UK, the General Medical Council (GMC) offers an outcomes statement at the undergraduate level [20], general standards for curriculum management and delivery at undergraduate and postgraduate levels [21], and specific standards for curriculum design for postgraduate training [22]. In the USA, the Liaison Committee on Medical Education sets accreditation standards that contain guidance on many key aspects of curriculum in medical schools, but not on how to frame the curriculum statement itself [23].

Not surprisingly, most curriculum standards address a similar range of fundamental issues such as:

- educational objectives
- curriculum structure and design
- content
- teaching and assessment
- curriculum management
- roles and responsibilities
- evaluation of curriculum effectiveness.

Additionally, medical educators at all levels can compare their own curricula and medical education and training processes with advisory or indicative sets of standards, such as those published by the World Federation for Medical Education (WFME) [24]. (See Box 5.3). The WFME has set, piloted, and evaluated quality improvement standards for all aspects of medical education at all stages to 'provide a mechanism for quality improvement in medical education, in a global context, to be applied by institutions, organisations and national authorities responsible for medical education' [24]. These are all aspects of curriculum and they support the view that curriculum design must encompass much more than a statement of the content to be covered in the course. The WFME standards are widely used within medical schools and for accreditation purposes. Box 5.3 lists the aspects of the content and development that the WFME considers important to state in postgraduate training curricula [25], although similar headings are used at all levels of medical education and training.

Ultimately, standards for curricula try to decrease the distance between the three coexisting types of curriculum identified by Coles [26]:

- the curriculum on paper
- the curriculum in action
- the curriculum the learner experiences.

The standards cited all require the curriculum designer to think about the intended product and character of the course or programme, its rationale, values, or mission. Without these elements, the development of curricula becomes a dangerous and instrumental undertaking, apt to serve only political or economic purposes. 'Aims-talk', as Noddings [27] calls it, is the first and most important element of curriculum design and its most important standard whereby local relevance can be assured.

BOX 5.3 Components of a postgraduate medical education curriculum statement [25]

1. Mission and outcomes
- 1.1 Mission
- 1.2 Professionalism and professional autonomy
- 1.3 Educational outcomes
- 1.4 Participation in formulation of mission and outcomes

2. Educational programme
- 2.1 Framework of the PGME programme
- 2.2 Scientific method
- 2.3 Programme content
- 2.4 Programme structure, composition and duration
- 2.5 Organisation of education
- 2.6 The relation between PGME and service

3. Assessment of trainees
- 3.1 Assessment methods
- 3.2 Relation between assessment and learning

4. Trainees
- 4.1 Admission policy and selection
- 4.2 Number of trainees
- 4.3 Trainee counselling and support
- 4.4 Trainee representation
- 4.5 Working conditions

5. Trainers
- 5.1 Recruitment and selection policy
- 5.2 Trainer obligations and trainer development

6. Educational resources
- 6.1 Physical facilities
- 6.2 Learning settings
- 6.3 Information technology
- 6.4 Clinical teams
- 6.5 Medical research and scholarship
- 6.6 Educational expertise
- 6.7 Learning in alternative settings

7. Programme evaluation
- 7.1 Mechanisms for programme monitoring and evaluation
- 7.2 Trainer and trainee feedback
- 7.3 Performance of qualified doctors
- 7.4 Involvement of stakeholders
- 7.5 Approval of educational programmes

8. Governance and administration
- 8.1 Governance
- 8.2 Academic leadership
- 8.3 Educational budget and resource allocation
- 8.4 Administration and management
- 8.5 Requirements and regulations

9. Continuous renewal

Ethical Considerations

It has long been recognised that a curriculum is a social and psychological experiment in action, the subjects of which are teachers and learners [28]. Given that curriculum change tends to be driven by academic fashion, social values, and educational ideology, rather than by evidence, and that there is no institutional ethics committee that sanctions such change or tempers unwarranted enthusiasm, it is incumbent upon curriculum designers and developers to identify what problems are being addressed, and to provide a convincing rationale for the proposed solution. Curriculum designers must also exert caution when imposing on learners a new and untested experience that will affect a crucial stage of their personal and professional development and which will have an effect on the health care they provide and the teams they work with, at least in the early stages.

Some may think that designing a curriculum on paper, no matter how cautiously, is not as ethically defensible as working with teachers to help them develop their own practice. In this way, through reflection and action research, the 'teacher as researcher' becomes an extended professional central to both curriculum delivery and curriculum development [29]. The underpinning curriculum theory or model therefore arises from the grounded development of practice.

The Importance of Context

The most powerful emerging influence on thinking about curriculum concerns the role of the local context, and the dangers of importation of curriculum models from different and incompatible cultures and systems [7], even as the international trade in curriculum as a transferable commodity flourishes. And yet there is no evidence that Western models (for the flow of ideas has invariably been from west to east) are any better in their outcomes than other models [7, 30, 31]. A worrying phenomenon has been noted, namely the 'apologetic stance taken by authors in the east about their slowness in adopting Western methods, even though … those methods will demand an "intense re-socialisation of learners into metropolitan Western mindsets"' [7, p. 177].

But this is not simply an east–west issue. Differences in educational and assessment culture have been shown in medical education, even within and between Western countries [32–34]. So a contextual curriculum, one that derives from and is appropriate to its site of implementation, will not place its emphasis narrowly on educational method and the search for the most effective methods of teaching and learning, for which there is no robust differentiating evidence. Instead, the emphasis must be on context, on health benefits, and benefits to the scientific and cultural basis of medicine. In a contextual curriculum, the 'medical education' decisions become secondary.

Before moving on to think about curriculum design in more detail, it is important to be clear about the necessary considerations in a curriculum sensitive to local context. Some of these will be true for any curriculum, some will not. They include:

- Consideration of the body of knowledge, skill, and experience necessary for the practice of medicine in the local context. This may be derived from the scientific base as commonly used and understood but must be done consciously and on the basis of analysis.

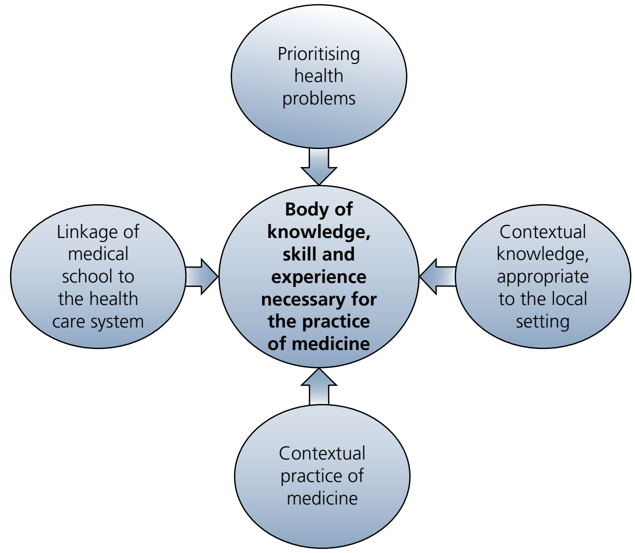


Figure 5.1 The curriculum in context.

- Prioritisation of health problems, which will yield very different results from location to location.
- Contextualisation of knowledge, appropriate to the local setting, which will allow not only appropriate understanding of the context of health and illness, but also of the approach to communication and clinical decision-making.
- Awareness of the diversity of medical practice, according to which even the classification of disease, its manifestation, and treatment are all linked to the local context.
- Linkage of medical school inputs, processes, and outputs to the health care system – without which contextualisation of learning is severely compromised.

These are also represented in Figure 5.1, derived from [16]. In summary, a curriculum must be contextual to be meaningful. Bearing this in mind, we can now consider the more traditional views of curriculum.

Factors that Influence Curriculum Design

While writing a curriculum is a process that demands consideration of values, beliefs, and context, it also deserves a review of current evidence and an explicit development process that sends out messages about quality assurance and recognition of stakeholders. Jolly and Rees recognise this need for a rational, open, and accountable curriculum design process. They eloquently describe the accompanying lack of evidential basis for how best to do this, but conclude that: 'Although curriculum design is an imprecise and arbitrary rubric, such a code is needed: systematic and arbitrary is somewhat better than capricious' [35, p. 22].

The days when subject experts or workforce managers alone wrote down what was to be learnt are now past. Curriculum design now encompasses many other factors that derive from the democratisation of social processes, the development of educational theory, political imperatives, and economic concerns. Box 5.4 highlights some influences

BOX 5.4 Influences and effects on medical curricula

Influence	Effect on medical curricula (example)
Theories of professional practice	• Integrated curricula • Teamwork • Ethics
Theories of learning	• Learner-centred design, e.g. problem-based
Social values	• Socially responsible medical schools • Widening participation curricula
Knowledge base expansion	• Core and options curriculum
Professional	• Communication skills training • Professionalism
Health service development	• Community-oriented curricula • Multiprofessional elements
Political	• Shorter curricula for faster production of medical workforce
Accountability and transparency	• Outcomes-based curricula • Objectives-based curricula

on modern medical curricula and their areas of effect. Each of the influences cited here has left its mark, and the residue of each remains to become incorporated into the new generation of curricula, making each new reformulation richer than the previous models.

Professional Practice

The evolution of curriculum models and learning theories is addressed below. However, other factors, which are not part of the academic discourse, are equally as important in shaping ideas about curriculum. Some of these factors affect the content of the curriculum and some affect its design. For example, concepts of professional practice in the UK have arisen around the ideas embodied in the UK General Medical Council's statement *Good Medical Practice*, which defines a set of core professional behaviours and values [36]. This document covers such issues as standards for clinical care, maintaining good medical practice, educational activity, and relationships with patients, colleagues, and within teams. Although originally intended as a professional guidance document, *Good Medical Practice* has had a significant influence on curriculum design [20].

The Royal College of Physicians and Surgeons of Canada [37] has also issued a similarly influential statement on the essential roles – and therefore key competencies – of specialist physicians (CanMEDS). This statement addresses the qualities of a doctor that every educational programme should facilitate in relation to the professional roles of:

- medical expert
- communicator
- collaborator
- leader
- health advocate
- scholar
- professional.

Such statements not only contribute to the vision that an organisation has of its intended product, but will also affect directly the content and style of the curriculum. On the other hand, social drivers for accountability and transparency have determined the use of clear outcomes, amenable to peer or lay input and review. Political imperatives have often pushed curricula to be more aware of issues of the cost and speed of workforce production. From this, we should be aware that choice of curriculum design or model is not an objective entity but is socially, professionally, academically, and politically constructed. At any one point, curriculum design is a child of its time.

Curriculum Models

Curriculum models have been the subject of academic and management theory since the mid-twentieth century, when Tyler first put forward the idea that: '… it is very necessary to have some conception of the goals that are being aimed at. These educational objectives become the criteria by which materials are selected, content is outlined, instructional procedures are developed and tests and examinations are prepared' [38, p. 52].

Although Tyler adopted a relaxed view of how objectives should be framed, this approach still allowed a 'transmission model' [39] of learning, which focuses on the teacher's, rather than the learner's, activity. Despite Mager coining the subsequent term 'instructional objectives' and taking a harder line on expressing objectives in measurable terms, his simultaneous intention was to change that focus and emphasise the importance of student achievement rather than teacher activity [40]. At the same time, he was much more prescriptive about exactly how those achievements should be specified: in behavioural, observable terms that were amenable to assessment. And so the use of the curriculum as the foundation of assessment became a central tenet.

There followed a raft of curriculum theorists who found that the Mager and Tyler models did not encompass all types of valued learning. So, for example, Eisner [10], thinking about art criticism and connoisseurship, introduced the idea of problem solving and expressive objectives or expressive outcomes that are not predetermined, but are generated out of an activity and then reflected upon in a responsive, evaluative manner. Some might call this true learner-centredness, or perhaps constructivist learning. In medicine, where there is a body of knowledge that must be acquired in its given form, educators might consider whether such concepts are the most appropriate. A new theory of learning might be required.

Some theorists tried to break free from curriculum models that specified outcomes in whatever form. Stenhouse [29], for example, proposed a process model that focused on the processes of acquiring, using, and evaluating the knowledge of the discipline. Outcomes, then, would be truly learner-centred, rather than having the contradictory position of a learner-centred rhetoric aimed at their achievement of outcomes specified by others.

This contradiction has been compounded in more recent times, during which the hegemony of a competence-based curriculum model, which was originally introduced in practical vocational subjects, has held sway. Its suitability as a basis for assessments, its common-sense appeal, its apparent analytical basis, and its implicit message that if we could define competencies we could ensure that learners acquire them, and be assured by relevant testing that this is so, all make a competence model attractive. I have myself argued that competencies alone cannot describe even the skills, much less the performance, of a profession [41]. Some prominent writers, such as Hyland, have suggested that the competence movement in curriculum design is little more than an economically driven derivative of the behavioural school: 'This attempt to specify exactly what is to be achieved and measured is, of course, nothing more than reconstituted behaviourism … Constructed out of a "fusion of behavioural objectives and accountability" … the movement provided irresistible appeal to those seeking accountability and input–output efficiency in the new economic realism of the 1980s' [42, p. 49]. Perhaps this will 'ring some bells' for medical educators today.

The twin factors of accountability and efficiency of education or training appeal to medicine, as it has become increasingly concerned in our bureaucratic, compliance-dominated era about demonstrating transparency and public accountability in times of increasing litigation. The contextual climate of a hard-pressed health service, limited resources, and managerial and political imperatives has made the competence model very alluring. On the other hand, the rise of competency-based models has possibly increased the tendency to 'teach-to-the-test' along with a more instrumental, less creative, approach to learning on the part of the students that might encourage minimalism rather than professional excellence [43].

These examples of curriculum models reveal that their use can be a function of instrumental pragmatism, values and vision, political, social, and managerial imperatives, and of the ideas that are current about how people learn. This means that selection of a curriculum model is a process that requires careful thought and open justification.

Learning Theories

Ideas about learning influence curriculum design in two ways: first, by affecting the structure of the education and training, and second, by affecting the choice of teaching and learning methods. These two are related. In adopting ideas about learning that might influence curriculum design, curriculum designers should differentiate very clearly between what is a theory of learning based on available evidence, and what is simply a framework or an idea or perception of some phenomenon. 'In psychology, theories are used to provide a model for understanding human thoughts, emotions, and behaviours' [44]. The work of Piaget, Vygotsky, Dewey, Bruner, and others come under the heading of cognitive theory. Ideas such as active and passive learning, learning styles, and adult learning are not theories, and have no evidence base. A phenomenon is simply an observed behaviour which might or might not be content- or context- or person-specific. To base a curriculum on the observation of phenomena would not be defensible. To base it on a theory might be more rational. Having said that, some authors fail to differentiate between evidence-based theory and simple observation of phenomena [45] and many of the mantras that drive medical education are not based on evidence [46].

Learning theories and the observation of phenomena have influenced curriculum design. When objectives-based curriculum models were predominant so was behavioural theory. While behavioural theory has declined, however, assessment practice and managerial imperatives have taken over to ensure that the behavioural aspects of curriculum definition still remain, albeit in new guises (defined as competencies, perhaps). The application of adult learning principles [47] to curriculum design provide an example of the effect of observation of phenomena. These principles, which promote 'active' 'self-directed' learning towards personally relevant goals, have guided curriculum design despite their lacking an evidence base and true theoretical status. Some medical schools, such as that of the University of New Mexico, have espoused an adult learning-style curriculum that is 'student-centered, small-group education with early clinical skills through community-based, self-directed instruction' [48]. It is an evolving curriculum, as many are, often based on changing belief and tested only in practice. A history of the fate of such curriculum developments is given in a personal and quite moving account of educational belief meeting the pragmatics of medical school reality [45].

As with every other aspect of education, applying learning theory and the observation of phenomena to pedagogical practice is a never-ending work in progress because social and cultural ideas change. So, when objectives-based curriculum models were predominant, behavioural theory was also in its prime, and the role of the teacher in shaping behaviour was a main focus. However, the focus subsequently moved away from teaching and towards learning. This may well have been in response not only to changing social ideas about how to bring up children in a more liberal and consultative manner (which relates ultimately to the position of women in society), but also in response to the cognitive and constructivist theories of learning and development which educational and cognitive psychologists have advanced, through research [49–51]. Nicol explains that nowadays, the teacher: '… encourages participation, dialogue and interaction by students with course materials and with each other. The teacher should, it is said, function as a facilitator of learning, intellectually critical, stimulating and challenging, but within a learning context that emphasises support and mutual respect' [52]. This is a social value more than a pedagogical view. There is no rationale in educational or cognitive psychology that would inhibit a teacher from teaching. Even teaching facilitates learning!

Considering robust learning theory rather than the observation of phenomena [53], some key premises from educational psychology include:

- Learners first need a strong basis of structured foundational knowledge in their long-term memory – that foundational knowledge is almost certainly the basic sciences since these comprise the most generalisable and structured knowledge.

- Knowledge is best learned by being taught in an organised manner; this might even involve learning by rote.
- 'Learner-centred' and 'active' ideas of learning may overload short-term memory and endanger transfer to long-term memory because of lack of robust structure – a new learner cannot reconstruct the knowledge that experts have compiled over a very long period.
- Knowledge needs to be used repeatedly to tune it to the needs of practice – as begins during the clinical phase of medical school and builds up thereafter – which also constructs an ever-expanding structured store of experience linked to robust knowledge.
- Contextualising learning from the outset by using clinical examples, by sending students into the clinical arena or into the community for short periods, and by teaching them clinical and communication skills, might be helpful on a personal and motivational level and perhaps in terms of both helping them to understand the importance of the foundational knowledge they are learning and developing their identity as future doctors.
- Integration of learning is a function of that structure and use, and integration can only happen inside the learner's head, not in the curriculum.

Cognitive Development

Cognitive theories of development offer additional insights for curriculum designers to consider. For example, the philosopher Immanuel Kant and the psychologists Lev Vygotsky, Edward Bartlett [54], and Jean Piaget [55], among many others, defined cognitive development in terms of schemata. A schema is a cognitive framework or memory structure that helps organise and interpret information. As experiences happen, this new information is used to modify, add to, or change previously existing schemata. The result is learning. 'Assimilation' is the process of incorporating new information into previously existing schemata. 'Accommodation' involves altering existing schemata, or ideas, as a result of new information or new experiences. New schemata may also be developed during this process. The important issue here is the quality of initial schemata as they are laid down.

A key concept for curriculum designers then is the student's trajectory of learning. It might seem surprising that a traditional curriculum, with differentiated preclinical and clinical phases of learning, is more effective than problem-based learning in encouraging clinical problem-solving skills [56]. A well-structured knowledge base is a good springboard towards freedom of creative thought [57]. In an environment which demands constant new problem-solving for each new patient or presentation, a strong and structured base of knowledge, tuned through experience, and supported by skills, is the essential component for effective problem solving. The most effective trajectory of learning, therefore, will initially ensure well-structured knowledge that is almost independent of problems or situations and relates to the learner's stage of mastery of concepts. Such knowledge is therefore transferable, and can be followed or accompanied by its contextual application. But the knowledge must come first and must have its own coherent organisation. It is that which ensures transferability. This might suggest that learning the basic sciences while having the contextual background that, for example, early clinical exposure offers, would indeed yield more effective clinical problem solvers [58–62].

The use of learning trajectories to structure the curriculum has been successfully used at all levels of learning. For example, the approach has been explained convincingly in relation to early childhood mathematics [63]. The notion of using instructional activities to link goals and developmental paths in ways that yield increasingly higher levels of thinking probably resonates with many medical teachers.

More on the emerging field of cognitive neuroscience and the insights it provides for medical education and training can be found in Chapter 3 of this book.

Dominant Discourses

Education reflects social values. It is important, therefore, for any curriculum planner to be aware of the social ideas that influence educational thinking in their own context. Such ideas might be *dominant* (held by the majority), *emerging* (held by a growing number of people), *oppositional* (directly challenging the majority view), or *alternative* (simply offering a different perspective) [64]. Conscious analysis along these lines of the principles that drive curriculum design would be helpful, along with an analysis of whether any of them is contextually appropriate.

Where there is no proper theoretical or evidential basis for curriculum change, trends have risen and then gradually retreated as they fail to deliver hypothesised benefits. The dominant rhetoric is powerful in its effects on educational practice whether or not that dominant rhetoric is based on proper theory or evidence. For instance, the current dominant view of effective learning, requiring activity on the part of the learner, has spawned a panoply of ideas about the components and manifestation of this approach.

Other ideas that medical education has chosen to embrace include the unproven dichotomy between 'deep' and 'surface-level' approaches to learning [65, 66]. The former is said to be characterised by an active concern in the student to seek the underlying meaning, the wider picture, the relationship between different information and experiences, the logic of the argument, and the need to question and understand. Surface-level processors, on the other hand, are said to seek only to learn the content, acquire the knowledge, and get the right answers [67]. But a surface characterisation of learning styles can fail to illuminate the deep strategic thinking that is actually occurring, and can be culturally determined [68].

The inconsistencies of the idea of deep or surface-level learning led to the proposition that there might be a third approach: strategic learning [67]. Strategic learning was described as using different approaches to learning and managing time in order to achieve the highest possible grades, based on an understanding of the assessment system requirements. The recognition that students will study in order to pass the test throws some doubt on conceptual frameworks such as deep and surface-level learning which may well reflect outcomes rather than processes.

These and other ideas about, for instance, learning styles and approaches has had implications for curriculum design in terms of teaching skills and methods, learning opportunities, and assessment [69]. Curricula that dissuade students from simple rote learning (although this might actually be productive repetitive learning) and encourage apparently deep processing (although this can only occur in the presence of acquired knowledge) have now become the dominant form, recognising that curricula might affect a learner's approach to learning. McManus et al. put forward their opinion that:

> Formal education, particularly effective formal education, can also alter study habits and learning styles … Intercalated degrees increase deep and strategic learning and decrease surface learning at medical school … Deep and strategic learning also relate to the clinical experience gained by medical students, making it possible that greater patient involvement during undergraduate clinical training, rather than mere reliance on textbook learning to pass exams, a characteristic of surface learners, will also reduce surface-disorganised approaches to work [70].

The ideas that inform today's curriculum design, despite their lack of evidence, seem to be very far from the behavioural theories of learning, and far from the implication of cognitive psychology that the knowledge base of the discipline must first be learnt before its application can be attempted. Today's trajectory of learning is portrayed as flatter, with integration being the hallmark throughout the course, and deep learning in the context of practice its aim.

At the same time as these liberal developments, we also have seen the rise of more conservative competence-based curriculum frameworks, which seem strangely to hark back to the days when curricula were based on the attainment of set objectives and the underlying theory was distinctly behavioural. This contradiction remains unresolved in the competence-based curricula of today. On the one hand, such a curriculum model specifies predetermined outcomes that the learner must attain. On the other hand, we see simultaneous advocacy for 'student-centred' learning in which knowledge is said to be constructed from the experience of the learner and resides in the mind rather than externally [71]. So, learning reflects the learner's understanding and personal interpretation of the world. This is at odds with a curriculum that states what must be learned, as a medical curriculum must. The acquisition, not construction, of a large body of knowledge and specific skills still lies at the heart of medicine, as it does in any profession. Medical education has developed a populist and different version of student-centred learning which bears little relation to the more profound original set of theoretical ideas deriving from social and philosophical perspectives.

Decisions in Curriculum Design

Despite the differences in perspective that have existed over the years between different practitioners and theorists, all generally agree that the process of curriculum design must answer the following central questions, originally set out by Tyler in 1949 [38].

- What is the purpose of the educational programme?
- How will the programme be organised?
- What experiences will further these purposes?
- How can we determine whether the purposes are being attained?

Systems have been suggested for curriculum development since Tyler's time. In medicine, Kern's [72] six-step approach has appealed to many for its simplicity of approach. The derivative six steps, which build on the work of Tyler and a number of classical authors in the field, are:

- problem identification and general needs assessment
- targeted needs assessment
- goals and objectives
- educational strategies
- implementation
- evaluation (which includes assessment of learning as well as programme evaluation) and feedback.

Although Kern's approach was written for medicine, curriculum theory offers numerous other curriculum models, political frameworks, learning paradigms, professional and social theories, and approaches to curriculum development that could equally as well be applied [16].

Figure 5.2 summarises the decisions that most curriculum theorists agree should be addressed in the process of curriculum design. Although these decisions are presented serially, such decisions often occur in parallel, in a different order, or are iterative, because they are so tightly interdependent and are a function of local conditions. The curriculum designer must make choices about how to answer each of these questions. We have seen that those choices are influenced by a number of contextual factors, but what options are available at each stage? The next sections of this chapter set out some of those choices.

What is the Fundamental Character of the Course?

Underpinning the overall purposes of the curriculum is a set of values that pervade the thinking or the aspirations of the school and describe its character in practice. Many years ago, these value choices were set out in the SPICES [73] model as a series of dimensions between two extremes. Despite its popularity, and its implied preference for 'innovative approaches' (student-centred, problem-based, integrated, community-based, electives, systematic) over 'traditional' ones (teacher-centred, information gathering, discipline-based, hospital-based, standard programme, apprenticeship-based) this model uses vague terminology, mixes decisions at different levels and types, and lacks evidence to underpin either its dimensions or its hierarchy. Curriculum designers should be cautious not to attach automatic value judgements to either dimension; for example, apprenticeship learning is still regarded as fundamental to medical training, and the potential narrow instrumentality of a planned systematic approach is recognised as having its dangers in professional training. Rather than deciding between a narrow set of predefined qualities, the curriculum designer might better read more broadly, consult more widely, and determine their own value set in their own context, based on their own understanding of evidence and purpose.

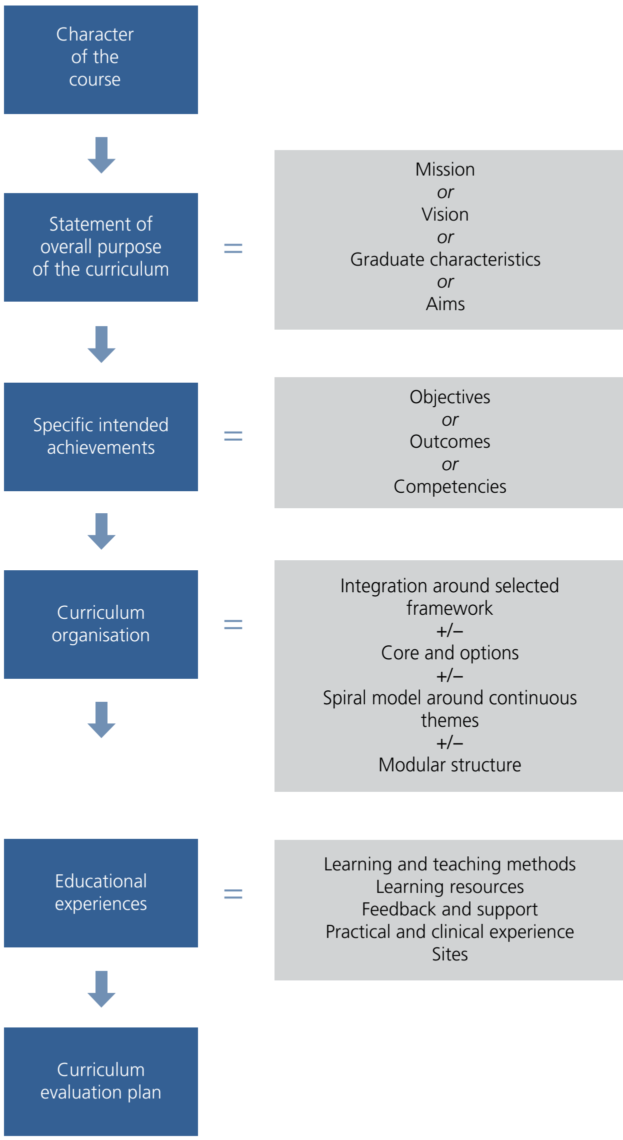


Figure 5.2 Steps and options in curriculum design.

How Should the Overall Purpose of the Curriculum be Expressed?

The purpose of a curriculum is often based on a set of aims, or a mission statement, such as the WFME standards require, or a statement of professional values, such as that embedded within the CanMEDS Roles, or a 'vision statement', such as that developed by the University of Sheffield Medical School. The Sheffield vision statement encompasses the general intentions, values, and characteristics of the curriculum, specifically:

- the instructional approach (a spine of problem, case-based, and patient-based integrated learning activities complemented by a range of other teaching and learning activities, with an increase in systematic teaching of some components to ensure competence in key areas)
- the learning approach (progressively more self-directed, supported by information technology resources, distance learning, and activities)
- the assessment system (formative and summative based on the defined outcomes)
- the curriculum management system and team
- the curriculum monitoring and improvement system.

An effective vision statement addresses all the central curriculum design issues and must be the result of extensive discussion and consultation with the relevant stakeholders and experts, as such consultation is fundamental to a properly managed change process [74]; in Sheffield it took nearly a year to complete [75].

In contrast, Brown University School of Medicine [76] chose to think about the intended achievements of its curriculum in terms of the abilities of successful doctors. It derived, through consultation, nine such abilities, as follows:

1 effective communication
2 basic clinical skills
3 using basic sciences in the practice of medicine
4 diagnosis, management, and prevention
5 lifelong learning
6 self-awareness, self-care, and personal growth
7 the social and community contexts of health care
8 moral reasoning and clinical ethics
9 problem solving.

Curriculum statements can be written in different ways, but ultimately they tend to express very similar ideas. What is important is that the statement of purpose suits the local context. Figure 5.1 highlights that this statement should include reflection on social, academic, and professional issues, as well as a local prioritisation of health problems.

How will the Specific Intended Achievements be Described?

The specification of intended curriculum outcomes (expressed in whatever terms) is, in almost all cases, non-negotiable, not least because the syllabus part of the curriculum is the basis for planning and developing the assessment system. The description of learning opportunities might also be relevant in planning assessments. If there is no agreed curriculum, how can we develop an objective, representative, valid, and reliable assessment system? Simply, we cannot.

There are many ways to express what it is that a curriculum is intended to achieve. We have seen that the choice of expression is often as much a function of social context and educational fashion or belief as it is of any objective evidence of effect. The decisions around intended achievements are important because they both define the content of the course and set the basis of the assessment blueprint.

Not surprisingly, this is another contentious area of curriculum design: every department and teacher will want to have their own subject properly represented in the curriculum, and a team-based approach that matches the organisation of the curriculum is advisable, with iterative consultations following a properly managed change process [74].

Essentially, what the curriculum intends to achieve is most commonly expressed in one of the following ways:

- As *objectives*, expressed as the specific knowledge, skills, and attitudes that the student will display at the end of the course. As we have seen, the objectives model

became predominant after the Second World War, when reconstruction was most efficiently tackled in a managerial way, leading to observable and measurable changes, after the chaos of the preceding period.
- As *intended outcomes*, stated in clear and precise terms, which will allow the designer to specify the learning experiences that will facilitate achievement of the stated outcomes. For many, this is a return to Tyler's original idea of objectives [77].
- As *competencies* to be achieved and assessed, again expressed in terms that bear similarity to objectives but are often thought of in relation to the ultimate intended performance that the competencies underpin. A recent variant on competencies is entrustable professional activities (EPAs), which translate competencies into practice so that supervisors can decide on the required level of supervision [78, 79]. This formulation recognises the reality of clinical medicine by integrating multiple competencies, as they are integrated in practice, rather than atomising integrated practice into component competencies.

There has been and still is a furious debate around the use of these terms, and what they mean, how they differ, what they imply, and how they are used (see Box 5.5). It has been argued that a simple statement of competencies alone cannot reflect the complex nature of a profession or the central skill of professional judgement [41]. It was Stenhouse's belief that a statement of behavioural objectives cannot address socialisation and problem solving [29], which are processes fundamental to a profession. It has also been argued that such 'product-oriented curricula' are disempowering for the learners and take control of learning away from the learner [80], and possibly disempower teachers similarly. In this, an outcomes-based curriculum would be incompatible with a learner-centred approach to learning, yet the two, in many curricula, attempt to coexist. Instability in the definition of these terms might enable that unlikely companionship.

Guidance on the specification of outcomes sounds similar to objectives-based approaches to curriculum design. In outcome-based education:

> Decisions about the curriculum are driven by the outcomes the students should display at the end of the course. In outcome-based education, product defines process. [It] … can be summed up as 'results-orientated thinking' and is the opposite of 'input-based education' where the emphasis is on the educational process and where we are happy to accept whatever is the result. In outcome-based education, the outcomes agreed for the curriculum guide what is taught and what is assessed [81, p. 8].

The instrumental nature of this approach has given rise to some controversy. Key writers have sometimes used the terms 'outcomes-based' and 'competency-based' interchangeably [76], equating the two as the same thing in practical terms [82]. Objectives also have many similarities. An outcome might be: 'Obtains history in relation to possible underlying causes including cardiovascular and non-organic causes' [75]. It would be difficult to say how this differs from a competency or an objective or even an EPA. And it really does not matter, because this debate has no conclusion. What is important is fitness for purpose, and the main purposes of stating the intended achievements of the curriculum are:
- to inform learners of what they should achieve
- to inform teachers of what they should help the learners to achieve
- to form the basis of the assessment system, so that everyone knows what will be assessed
- to reflect accurately the nature of the profession into which the learner is being inducted and the professional characteristics that must be acquired.

Regardless of the rhetoric surrounding these different ways of describing what a curriculum should achieve, the important point is that this is done in terms specific enough to guide planning, assessment, and review, and to give

BOX 5.5 FOCUS ON: Competence and competency

The terms 'competence' and 'competency' seem to be a perpetual focus of educational concern and debate. But a preoccupation with this definition of terms is, perhaps, to miss the point. In most dictionaries, these are alternative words with the same meaning. Both simply mean 'the ability to do something; the ability to perform a given task'. So there is no contest between competence and competency – it is simply a matter of which word you care to use. But this definitional fact does not stop a semantic debate raging.

In the usual curriculum parlance, competencies are specific, measurable entities (units of knowledge, skill, behaviour, etc.) that the learner should display by the end of the programme. But this does not mean that the possessor of the competencies will translate these into *performance*. The underlying pedagogical theory seems to be that if we can define the competencies that make up professional performance, then we can aim the teaching programme at them and make it more efficient and effective. This theory is flawed.

If the acquisition of competencies in turn leads to ability to perform, this will be because the separate competencies have been used repeatedly in concert in the context of complex professional practice to gather information, to process it, to make judgements and decisions, to solve problems, to make interventions, to deal with and interact with peers, colleagues, and patients, and to think in multidimensional terms about personal, interpersonal, ethical, financial, managerial, multiprofessional, and evidence-based factors. To muddy the semantic water further, the 'ability to integrate and apply multiple competencies, not just in familiar and focused settings, but in novel, complex and changing circumstances' [83] is also referred to as *capability* [84].

A curriculum that bases itself on the specification of competencies is only recognising the first step on a path that leads to the precursor of the ultimate complex professional performance. And if we spend too long on debating definitions, perhaps we are no more than sublimating our energies and closing our eyes to more difficult questions.

students and teachers appropriate expectations. Perhaps it is time that medicine found a new and more appropriate way of describing its qualities.

How will the Curriculum be Organised?

Once the overall purpose of the curriculum and its more specific intended achievements are defined and agreed upon, the curriculum must be organised around one or more frameworks, models, or approaches. Some of the widely used organisational models and frameworks are based on vertical or horizontal integration of content or experiences, core and options, themes, spiralling of topics, and systems or life stages modules. These options are not mutually exclusive and many curricula display elements of them all. So an integrated curriculum with a modular core of mandatory content and student-selected options, which contain topics that are revisited in increasing depth at successive stages of the curriculum, is perfectly possible and probably the most common approach among new undergraduate medical curricula across the globe. Some of these organisational approaches are described in greater detail below.

Integration

In a discipline-based curriculum, knowledge and skills are presented as subject areas in their own right and integration has to occur entirely in the student's head through putting it to use in practice. An integrated curriculum organises the material to be learnt around an entity that is more related to practice, or at least tries to relate learning from different disciplines.

Curriculum integration in undergraduate medical education can be managed as either *horizontal* integration between different subject areas or *vertical* integration between the clinical and basic sciences. This can be a threatening development for some departments, especially in basic sciences, which often feel that they are likely to lose their identity. But if integration is properly managed, and the curriculum content properly defined, every department should be able to track its own contribution to the curriculum as a whole.

Integration typically implies a significant reorganisation of the curriculum and so decisions must be made about the basis for the integration. In other words, what will be the framework around which the content of the curriculum will be arranged? There are many choices.

- In Sheffield, the curriculum was designed around an agreed list of presenting clinical problems derived from published sources and other curricula, added to locally and then rated by clinical teachers for their importance. A blueprint for each problem was then constructed, which defined the curriculum content and outcomes [75].
- In Manchester [85], the core problem-based curriculum was organised around index clinical situations (ICSs) for which new graduates must have a required level of competence. These ICSs were derived in consultation with primary and secondary care clinicians, who then defined the knowledge and skills base for each one in a variety of specific domains, including technical, contextual, intellectual, and interpersonal.

Equally, the basis for integration could be bodily systems, age, patient cases, or any other grouping. Each approach has its advantages and disadvantages. Within the chosen framework, however, the specific content to be covered can be specified in terms of repeated and consistent curriculum *themes* that run vertically through the whole course. This is described further below in relation to modular design.

The choice of how to integrate can be informed by some of the learning theories described above (and in Chapters 3 and 4). For example, some schools of thought in cognitive psychology suggest that usable knowledge is the result of structured learning applied repeatedly in practice. To imagine that pre-packaged, integrated knowledge can be presented in a curriculum, rather than being achieved through an organised learning process, may be misguided. Integration happens inside the head of the individual learner, through the combination of prior knowledge with new information and/or experiences. Contextualisation at the time of delivery may well be useful, but the breakdown of structured knowledge for initial learning, for example by rearranging it around cases or conditions, may not provide a sufficiently robust foundational structure for future application. We know that in assessment terms, we must sample performance over many cases for the results to be generalisable [86]. Just as case-specificity applies to assessment so does it apply to learning, and foundational knowledge may be better learned within its own structures and frameworks.

The above argument suggests, perhaps, that a vertically integrated curriculum might produce greater learning efficacy than a horizontally integrated one or a curriculum organised around cases such as problem-based learning. Implications for curriculum designers therefore include:

- ensure that each component (basic and clinical skills) is learned in a structured way
- offer contextualisation
- provide opportunities to apply knowledge.

It is a widely held view [87] that the early clinical experience that vertical integration offers students is beneficial to their motivation and satisfaction, their acclimatisation and professional induction, and their valuing and contextualisation of the scientific base. It might strengthen and broaden learning and intensify the relevance of the course to ultimate clinical practice. However, these assertions still only attain the status of claims.

It is not only theories of how students learn that affect the design of curricula. Immediate relevance to practice [88] and theories about the discipline of medicine itself have also been paramount in changing the face of curricula. As the Case Western Reserve University School of Medicine describes its own history: 'Already a leading educational institution for more than a century, the School of Medicine in 1952 initiated the most advanced medical curriculum in the country, integrating the basic and clinical sciences, focusing on organ systems and featuring an introduction to patients and clinical work in the first year. Many other medical schools followed suit' [89].

Despite the widening adoption of integration as the basis of curriculum organisation, there is still no robust evidence base that shows its actual effects. As with most changes in education, the innovation occurs as a result of belief, rather than evidence, and gains credibility only through custom and practice.

Core and Options

The specification of core (or mandatory) and optional sections of the curriculum was a response to the perceived problem of content overload in medical education. However, 'core' can mean different things in different contexts and if a 'core and options' model is chosen, then the basis on which the core is selected must be known and agreed upon. To date, there is no adequate evidence base to suggest that one way of identifying the core is better than any other [90]. Harden and Davis [91] set out the possibilities:

- essential aspects of each subject or discipline
- essential competencies for practice
- areas of study relevant to many disciplines.

A fourth possibility is to study only those disciplines deemed essential, but this approach 'has caused great alarm among some teachers, and justifiably so' [91]. It is generally thought that students must gain knowledge and experience of all major disciplines at undergraduate level, since they are being prepared for any one of these as they move to the next phase.

There are many ways of determining the content of the core curriculum, ranging from modified Delphi processes [92, 93] and other formal consultations, to statistical and epidemiological methods, critical incident techniques, and informal consultative and team-based work. Whatever method is chosen, it should be well understood and publicised, and properly managed according to a timescale. It should involve all interested parties and stakeholders and bear in mind the vision of the school.

Options can then be built around the core and given timetabled slots or blocks to offer students choices in their learning and career development, and the opportunity for more self-directed study. Some guidance can be given: for example, options can be provided in different categories such as basic sciences, core extension studies, laboratory specialties, social and community sciences, education, and management. Students may then be required to undertake options in a variety of these areas.

Some medical schools have an 'options bank', which departments and teachers add to and students then select from. These would normally be well-defined elements with a specific assessment plan, each of which would be able to accommodate a limited number of students. It is also possible to allow students to design their own options, either within certain headings or freely but according to set criteria about planning, process, and outcomes against which the option can be marked and assessed.

Spiral

The principle of the spiral curriculum, first elaborated by the titan educationalist Jerome Bruner [94], is that students should revisit material at increasing levels of complexity as they progress through the course. This is almost unavoidable in practice. Thus, for example, the themes of clinical methods, ethics, and health promotion, and their accompanying attitudes, knowledge, and skills, were designed into the Dundee curriculum [95] to be revisited in more complex ways during the four main stages of the course, which dealt with normal structure, function, and behaviour, then abnormal structure, function, and behaviour, then clinical practice and, finally, on-the-job-learning.

Thus the features of the spiral curriculum, which might seem not unlike many other types of curricula, and might even seem unavoidable in practical terms, are that [96]:

- topics, themes, or subjects are revisited on a number of occasions throughout the curriculum
- levels of difficulty increase
- new learning relates to previous learning
- the learner's competence increases with progression through the curriculum.

Modular

A module is a self-contained unit of study. It should have its own outcomes (however expressed), activities, and assessments. Learners might take more than one module of study at a time. Modules are planned according to the curriculum framework selected. In an integrated course, modules will tend to have similar structures, with the vertical themes of the course that spiral through the curriculum being addressed in each module. So, for example, a module on cardiovascular disease might have its content decided in relation to curriculum themes of:

- clinical sciences
- basic sciences
- behavioural sciences
- population sciences
- clinical skills
- interpersonal skills and professional behaviours.

The module might then be taught around a number of index cases that illustrate these themes and the necessary content. Modules typically allow some flexibility in the order in which they are taught.

What Experiences will Further the Purposes of the Curriculum?

The experiences that learners have will be selected on the basis of the planning and design work carried out in the previous steps. The choices broadly relate to:

- learning and teaching methods, including learning resources, feedback, and support
- practical and clinical experience, including sites.

Learning and Teaching Methods

Decisions about learning and teaching methods will flow from the planning of previous stages. But there is no one-to-one relationship between course intentions and teaching and learning methods. Every curriculum designer has a range of choices that could lead to exactly the same outcomes. And every strength of any one teaching or learning method is accompanied by weaknesses. There is no pedagogical silver bullet or panacea.

The teaching and learning methods and resources a curriculum designer can choose from include, but are not limited to, the following:

- simulations; clinical skills laboratories, including communication skills training

- lectures
- seminars and tutorials
- independent or guided group work
- practicals
- study guides describing what is to be learnt and relating this to available learning opportunities [97]
- resource-based learning, including e-learning and library work
- formative assessment, appraisal, and feedback on learning
- clinical experiences.

Many of these methods and others are discussed in detail in other chapters.

The curriculum designer should state what balance of these methods might be desirable and expected. But the method alone will not determine effect on learning unless it is used in an appropriate manner. For example, problem-based learning has variable effects on the acquisition of knowledge [98], and any teaching or learning method which has a heavy workload, high contact hours, excessive material or an emphasis on coverage, is likely to push students towards a more rote approach to learning [99]. This might not always be inappropriate – but the curriculum designer needs to be aware of the cognitive effect of the curriculum design. Likewise, any educational method that displays an appropriate motivational context, a high degree of learner activity, interaction with peers and teachers, and a well-structured knowledge base may encourage a more thoughtful or reflective approach [100]. But this is not to set any value on either approach. Both have their value. Even rote learning suggests some inner cognitive activity ('passive learning' is simply an oxymoron) and is passionately defended in some disciplines and cultures. We have no evidence-based reason to demur [101].

The role of assessment as an instrument of learning, especially if used strictly for formative purposes, should not be overlooked and might be considered with other interventions such as appraisal and regular structured and supportive feedback sessions.

Clinical Experience

As with most published work in medical education, the focus of debate about curriculum has tended to be in the undergraduate arena. But the nature and relevance of curriculum is different for undergraduate and postgraduate medical education and also for continuing professional development.

Central to this difference is the role of clinical experience. For the new student, it is possible to define the component clinical and communication skills that must be acquired. These could even be expressed as discrete competencies which form the basis of the future learning trajectory. During medical school, when there is everything to learn, the school has considerable control of the sites, resources, and opportunities for learning. Correspondingly, it is both reasonable and managerially necessary to have a curriculum that guides the programme in practice and the associated assessment system. There might still be some opportunism in the use of clinical experience, but, in general, that clinical experience is designed to deliver curriculum outcomes associated with clinical and communication skills and clinical problem solving in a predictable and organised manner. At this stage, the curriculum is king.

At postgraduate level, things change. The medical student becomes a junior doctor. The skills and knowledge that have been acquired in the previous stage are now used, integrated, and built on in practice. The clinical problem-solving and patient management work of the developing doctor becomes less easy to describe as curriculum outcomes, even as EPAs. The educational programme now is fitted around, and supports, the practice that is central to the learner's professional development. The curriculum is no longer king; its role changes into a statement of things that must be experienced and learned in the unpredictable context of practice. See Chapter 12 for a more fulsome exploration of work-based learning in medical education.

Different countries address this differently. Some maintain some control of the educational programme and put in more specified periods of formal training, while others leave the doctor in training to be self-directed and to learn from their role as a junior doctor, providing care for an unpredictable stream of patients. At this postgraduate level, then, the nature of a curriculum statement might reflect the inherent unpredictability of the clinical context. An effect of this is the partial use of a curriculum as a type of checklist, with logbooks or portfolios, for example, recording which of the curriculum requirements have been met. We might see different assessments, based more on the workplace, in addition to the assessments which are focused entirely on the specialty-specific knowledge and skills that may be set by, for example, the UK medical royal colleges or the American specialty boards.

The final stage of medical education is that of continuing professional development (sometimes called continuing medical education) occurring once postgraduate training is complete and the doctor is an independent practitioner. At this stage, practitioners become increasingly individual and different in their knowledge, their experience, and their practice [102]. They are therefore individual in what they want and need to learn. It has been argued that a set curriculum is therefore inappropriate at this stage and that we should think more in terms of the central role of practice in giving rise to learning needs and wants, the individual doctor's preference for how to acquire that new knowledge or skill, and how then to reinforce that by bringing it back to the practice context which gave rise to it in the first place [14]. This is truly self-directed learning where the whole process is in the control of the learner. It is, nonetheless, a process which can be made transparent and accountable [103]. For more on continuing professional development, see Chapter 19.

In basic, or undergraduate, medical education, and perhaps even beyond, a wide range of knowledge, skills, and attitudes can be acquired as effectively in the community as in hospital settings [104]. So if the curriculum has the intention of producing graduates with an interest in practice in the community [105, 106], then primary care might be developed as a major provider of teaching, learning, and experience, offering effective integrated teaching [107]. Four types of community-based teaching have been identified [108]:

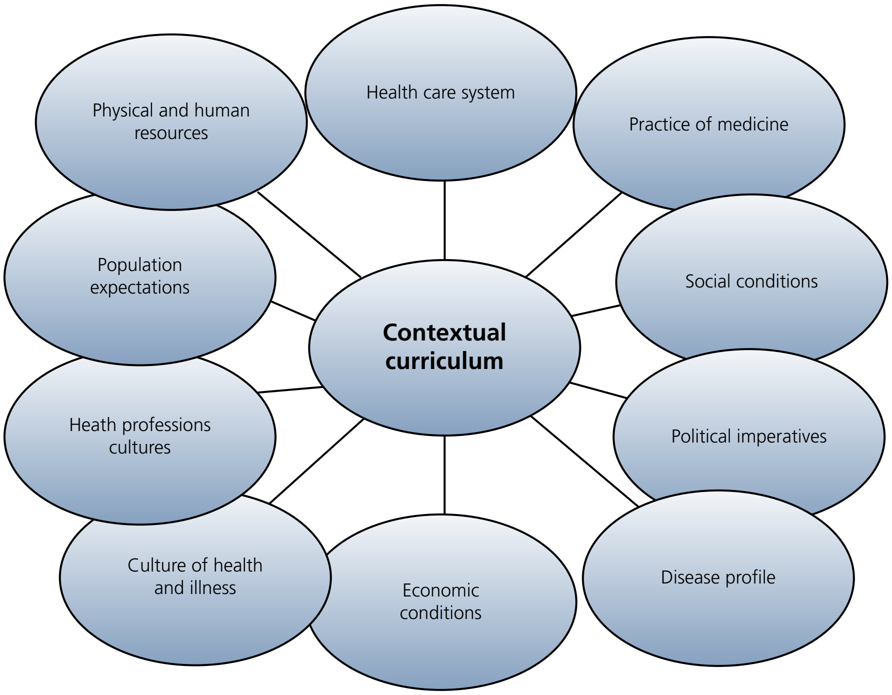


Figure 5.3 The curriculum in context.

- community-orientated teaching: teaching in and about the community
- agency-based teaching: teaching involving community health care providers other than primary care physicians
- general practice-based teaching: either specific clinical teaching or an attachment in primary care
- specialist teaching in the community: specialty subjects taught by hospital practitioners in a community setting.

Equally, such knowledge and skills might also be achieved in hospital settings.

Finally, the role of simulation and skills laboratories in helping students to acquire basic and more advanced clinical and communication skills in a safe, structured environment before using these with patients should also be considered as part of the curriculum-design process.

How will it be Determined Whether the Purposes of the Curriculum are Being Met?

Later chapters discuss assessment and programme evaluation in detail (see Chapters 20–26 and 30). Whether the purposes of the curriculum are met or attained might be measured in two ways. First, a robust assessment system that is properly blueprinted on to the curriculum will measure students' attainment of the intended learning outcomes of the programme. Second, a curriculum evaluation strategy that addresses the views and experiences of all stakeholders will offer information about how the curriculum in practice fulfils or does not fulfil its purposes. On the basis of assessment and evaluation findings, the curriculum can be reviewed and renewed to ensure that it remains fit for purpose.

Throughout all the steps outlined above, and in relation to all the considerations and judgements that are brought to bear in designing a curriculum, there is one principle that must hold sway – the principle of purpose. And purpose must derive from context. That context does not preclude the design of a curriculum that will produce researchers and academics, as they also have a key role in determining the scientific and practice basis of medicine; it does not preclude the production of doctors for secondary, or even tertiary, care, as they too are needed. A contextual curriculum can produce all these. But it does so by recognising local need and circumstances, and not by bench-marking to external contexts which derive from other cultures and practices. Figure 5.3, derived from [16], illustrates the contextual factors that might be brought to bear on contextual curriculum decisions.

Conclusion

Because all decisions about curriculum design are ultimately based on the judgement of the designer(s), working alongside stakeholders, within the constraints of physical resource and human capital, they must be locally driven. And whether the purpose is to produce Nobel laureates who will uncover the basic and clinical science that will improve practice and the health of nations, or to produce primary, secondary, or tertiary care physicians, the only rational and practical choice that will embed the medical school within its own locale might be to write and implement a curriculum that is contextual.

While this tells us what the curriculum should address, it tells us nothing about the curriculum structure, or the teaching and learning processes to be employed. In the face of lack of evidence, these are important but secondary decisions to be made cautiously and systematically on the basis of context, culture, resources, values, and beliefs. A curriculum, after all, is an ideological statement.

References

1 Miller, G. ed. (1961). *Teaching and Learning in Medical School*. Cambridge, MA: Harvard University Press.

2 Simpson, M.A. (1972). *Medical Education. A Critical Approach*. London: Butterworths.

3 Cox, K.R. and Ewan, C.E. ed. (1982). *The Medical Teacher*. Edinburgh: Churchill Livingstone.

4 Newble, D. and Cannon, R. (1983). *A Handbook for Clinical Teachers*. Boston, MA: MTP Press.

5 Balla, J.I., Gibson, M., and Chang, A.M. (1989). *Learning in Medical School. A Model for the Clinical Professions*. Hong Kong: Hong Kong University Press.

6 Boelen, C. and Woollard, B. (2009). Social accountability and accreditation: a new frontier for educational institutions. *Medical Education* 43 (9): 887–894.

7 Bleakley, A., Bligh, J., and Browne, J. (2011). *Medical Education for the Future. Identity, Power and Location*. London: Springer.

8 Gillam, S. (2011). Teaching doctors in training about management and leadership. *British Medical Journal* 343: d5672.

9 Nakkeeran, N. (2010). Knowledge, truth and social reality: an introductory note on qualitative research. *Indian Journal of Community Medicine* 35 (3): 379–381.

10 Eisner, E.W. (1994). *The Educational Imagination: on the Design and Evaluation of School Programs*, 3e. New York, NY: Macmillan.

11 Toohey, S. (1999). *Designing Courses for Higher Education*. Buckingham: Society for Research into Higher Education and Open University Press.

12 Apple, M.W. (2004). *Ideology and Curriculum*, 3e. London: Routledge Falmer.

13 Norcross, W.A., Henzel, T.R., Freeman, K. et al. (2009). Toward meeting the challenge of physician competence assessment: the University of California, San Diego Physician Assessment and Clinical Education (PACE) Program. *Academic Medicine* 84 (8): 1008–1014.

14 Grant, J. (2012). *The Good CPD Guide*, 2e. Abingdon: CRC Press, Taylor & Francis Group.

15 UNESCO (2017). Different meanings of 'curriculum'. http://www.unesco.org/new/en/education/themes/strengthening-education-systems/quality-framework/technical-notes/different-meaning-of-curriculum (accessed 15 March 2017).

16 Grant, J., Abdelrahman, M., and Zachariah, A. (2013). Curriculum design in context. In: *Oxford Textbook of Medical Education* (ed. K. Walsh), 13–24. Oxford: Oxford University Press.

17 Flinders, D.J. and Thornton, S.J. (2004). *The Curriculum Studies Reader*, 2e. New York, NY: Routledge Falmer.

18 Hoyle, E. (1971). How does the curriculum change? In: *The Curriculum: Context, Design and Development* (ed. R. Hooper), 375–398. Edinburgh: Oliver and Boyd in association with The Open University Press.

19 Kelley, A.V. (2009). *The Curriculum: Theory and Practice*, 6e. Thousand Oaks, CA: Sage Publications Ltd.

20 General Medical Council (2015). Outcomes for Graduates. General Medical Council, London. http://www.gmc-uk.org/education/undergraduate/undergrad_outcomes_overarching.asp (accessed 15 March 2017).

21 General Medical Council (2015). Promoting Excellence: Standards for Medical Education and Training. http://www.gmc-uk.org/education/standards.asp (accessed 15 March 2017).

22 General Medical Council (2017). Excellence by Design: Standards for Postgraduate Curricula. http://www.gmc-uk.org/education/postgraduate/excellence_by_design.asp (accessed 28 May 2017).

23 Liaison Committee on Medical Education (2016). Functions and Structure of a Medical School. http://lcme.org/publications (accessed 7 August 2017).

24 World Federation for Medical Education (2012). WFME Global Standards for Quality Improvement. WFME Office: University of Copenhagen, Denmark. http://www.wfme.org/standards (accessed 28 December 2012).

25 World Federation for Medical Education (2015). Postgraduate Medical Education. WFME Global Standards for Quality Improvement, 2015 revision. WFME Office, Ferney-Voltaire, France. http://wfme.org/standards/pgme/97-final-2015-revision-of-postgraduate-medical-education-standards/file (accessed 28 May 2017).

26 Coles, C.R. and Gale Grant, J. (1985). ASME Medical Education Research Booklet No 1. Curriculum evaluation in medical and health-care education. *Medical Education* 19 (5): 405–422.

27 Noddings, N. (2003). *Happiness and Education*. Cambridge: Cambridge University Press.

28 Elliott, J. (1990). *The Curriculum Experiment. Meeting the Challenge of Social Change*. Buckingham: Open University Press.

29 Stenhouse, L. (1975). *An Introduction to Curriculum Research and Development*. London: Heinemann.

30 Rao, K.H. and Rao, R.H. (2007). Reflections on the state of medical education in Japan. *The Keio Journal of Medicine* 55: 41–51.

31 Rao, R.H. (2006). Perspectives in medical education. Implementing a more integrated, interactive and interesting curriculum to improve Japanese medical education. *The Keio Journal of Medicine* 56: 75–84.

32 Segouin, C. and Hodges, B.D. (2005). Educating physicians in France and Canada: are the differences based on evidence or history? *Medical Education* 39: 1205–1212.

33 Hodges, B.D., Maniate, J.M., Martimianakis, M.A. et al. (2009). Cracks and crevices: globalisation discourse and medical education. *Medical Teacher* 31 (10): 910–917.

34 Jippes, M. and Majoor, G.D. (2008). Influence of national culture on the adoption of integrated and problem-based curricula in Europe. *Medical Education* 42 (3): 279–285.

35 Jolly, B. and Rees, L. ed. (1998). *Medical Education in the Millennium*. Oxford: Oxford University Press.

36 General Medical Council (2013). *Good Medical Practice*, 2e. London: General Medical Council.

37 Frank, J.R., Snell, L., and Sherbino, J. ed. (2015). CanMEDS 2015 Physician Competency Framework. The Royal College of Physicians and Surgeons of Canada, Ottawa. http://canmeds.royalcollege.ca/uploads/en/framework/CanMEDS%202015%20Framework_EN_Reduced.pdf (accessed 28 May 2017).

38 Tyler, R.W. (1949). *Basic Principles of Curriculum and Instruction*. Chicago: University of Chicago Press.

39 Jacques, D. (2000). *Learning in Groups. A Handbook for Improving Group Work*, 3e. London: Routledge Falmer.

40 Mager, R. (1975). *Preparing Instructional Objectives*, 2e. Belmont, CA: Fearon-Pitman Publishers.

41 Grant, J. (2000). The incapacitating effects of competence. A critique. *Journal of Health Sciences Education* 4 (3): 271–277.

42 Hyland, T. (1995). *Competence, Education and NVQs*. London: Continuum International Publishing Group Ltd.

43 Schwartz, B. and Sharpe, K. (2010). *Practical Wisdom*. New York, NY: Riverhead Books.

44 Cherry, K. (2016). What is a theory? Verywell Psychology. https://www.verywell.com/what-is-a-theory-2795970 (accessed 16 March 2017).

45 Richter, D.A. (2008). *The University of New Mexico School of Medicine Experiment: Transforming Learning to Transform Through Learning*. New Mexico: University of New Mexico School of Medicine. http://www.menninconsulting.com/resources/unm-som.pdf (accessed 17 March 2017).

46 Christodoulou, D. (2014). *Seven Myths About Education*. London: Routledge.

47 Knowles, M. (1979). *The Adult Learner: A Neglected Species?* Houston, TX: Gulf.

48 University of New Mexico School of Medicine (2018). America's Best Graduate Schools. https://som.unm.edu (accessed 18 June 2018).

49 Piaget, J. (1954). *The Construction of Reality in the Child*. New York, NY: Basic Books.
50 Bruner, J.S. (1977). *The Process of Education*. Cambridge, Mass: Harvard University Press.
51 Daniels, H. (1996). *An Introduction to Vygotsky*. London: Routledge.
52 Nicol, D. (1997). *Research on Learning and Higher Education Teaching. UCoSDA briefing paper 45*. Sheffield: Universities and Colleges Staff Development Agency.
53 Bates, B. (2016). *Learning Theories Simplified*. Los Angeles: Sage.
54 Bartlett, F.C. (1932). *Remembering: A Study in Experimental and Social Psychology*. Cambridge, England: Cambridge University Press.
55 Piaget, J. and Cook, M.T. (1952). *The Origins of Intelligence in Children*. New York, NY: International University Press.
56 Goss, B., Reid, K., Dodds, A., and McColl, G. (2011). Comparison of medical students' diagnostic reasoning skills in a traditional and a problem-based learning curriculum. *International Journal of Medical Education* 2: 87–93.
57 Cholowsk, K.M. and Chan, L.K.S. (2001). Prior knowledge in student and experienced nurses' clinical problem solving. *Australian Journal of Educational and Developmental Psychology* 1: 10–21.
58 Patel, V., Groen, S.J., and Scott, H.M. (1988). Biomedical knowledge in the explanations of clinical problems by medical students. *Medical Education* 22: 398–406.
59 Woods, N.N., Brooks, L.R., and Norman, G. (2005). The value of basic science in clinical diagnosis: creating coherence among signs and symptoms. *Medical Education* 39: 107–112.
60 De Bruin, A.B.H., Schmidt, H.G. and Rikers, R.M.J.P. (2005). The role of basic science knowledge and clinical knowledge in diagnostic reasoning: a structural equation modeling approach. *Academic Medicine* 80: 765–773.
61 Schmidt, H.G. and Rikers, R.M.J.P. (2007). How expertise develops in medicine: knowledge encapsulation and illness script formation. *Medical Education* 41: 1133–1139.
62 Grant, J. and Marsden, P. (1988). Primary knowledge, medical education and consultant expertise. *Medical Education* 22: 173–179.
63 Clements, D.H. and Sarama, J.A. (2009). *Learning and Teaching Early Math: The Learning Trajectories Approach (Studies in Mathematical Thinking and Learning Series)*. New York, NY: Routledge.
64 Williams, R. (1980). *Culture and Materialism. Selected Essays*. London: Verso.
65 Howie, P. and Bagnall, R. (2012). A critique of the deep and surface approaches to learning model. *Teaching in Higher Education* 18 (4): 389–400.
66 Marton, F. and Säljo, R. (1976). On qualitative differences in learning I and II. *British Journal of Educational Psychology* 46 (1/2): 4–11. 128–48.
67 Entwistle, N. (2000). Promoting deep learning through teaching and assessment: conceptual frameworks and educational contexts. Paper presented at the ESRC Teaching and Learning Research Programme, First Annual Conference, University of Leicester, November 2000. www.leeds.ac.uk/educol/documents/00003220.htm (accessed 16 March 2017).
68 Kee-Kuok Wong, J. (2004). Are the learning styles of Asian international students culturally or contextually based? *International Education Journal* 4 (4): 154–166.
69 Newble, D.I. and Entwhistle, N.J. (1980). Learning styles and approaches: implications for medical education. *Medical Education* 20 (3): 162–175.
70 McManus, I.C., Keeling, A., and Paice, E. (2004). Stress, burnout and doctors' attitudes to work are determined by personality and learning style: a twelve-year longitudinal study of UK medical graduates. *BMC Medicine* 2: 29. http://www.biomedcentral.com/1741-7015/2/29 (accessed 8 August 2017).
71 Semple, A. (2000). Learning theories and their influence on the development and use of educational technologies. *Australian Science Teachers Journal* 46 (3): 21–28.
72 Kern, D.E., Thomas, P.A., and Hughes, M.T. ed. (2009). *Curriculum Development for Medical Education. A Six-Step Approach*, 2e. Baltimore, MD: Johns Hopkins University Press.
73 Harden, R.M., Sowden, S., and Dunn, W.R. (1984). Educational strategies in curriculum development: the SPICES model. *Medical Education* 18: 284–297.
74 Gale, R. and Grant, J. (1997). Managing change in a medical context: guidelines for action. AMEE guide no. 10. *Medical Teacher* 19: 239–249.
75 Newble, D., Stark, P., Bax, N., and Lawson, M. (2005). Developing an outcome-focused core curriculum. *Medical Education* 39: 680–687.
76 Smith, S.R. and Dollase, R. (1999). AMEE guide no. 14. Outcome-based education. Part 2 – planning, implementing and evaluating a competency-based curriculum. *Medical Teacher* 21 (1): 15–22.
77 Prideaux, D. (2000). Emperor's new clothes: from objectives to outcomes. *Medical Education* 34: 168–169.
78 ten Cate, O. (2005). Entrustability of professional activities and competency-based training. *Medical Education* 39 (12): 1176–1177.
79 ten Cate, O. (2013). Nuts and bolts of entrustable professional activities. *Journal of Graduate Medical Education*. March 5 (1): 157–158.
80 Rees, C. (2004). The problem with outcomes-based curricula in medical education: insights from educational theory. *Medical Education* 38: 593–598.
81 Harden, R.M., Crosby, J.R., and Davis, M.H. (1999). AMEE guide no. 14. Outcome-based education. Part 1 – an introduction to outcome-based education. *Medical Teacher* 21 (1): 7–14.
82 Harden, R.M., Crosby, J.R., Davis, M.H., and Friedman, M. (1999). AMEE guide no. 14. Outcome-based education. Part 5 – from competency to meta-competency: a model for the specification of learning outcomes. *Medical Teacher* 21 (6): 546–552.
83 Neve, H. and Hanks, S. (2016). When I say … capability. *Medical Education* 50: 610–611.
84 Stephenson, J. (1998). The concept of capability and its importance in higher education. In: *Capability and Quality in Higher Education* (ed. J. Stephenson and M. Yorke), 1–13. London: Kogan Page.
85 O'Neill, P.A., Metcalfe, D., and David, T.J. (1999). The core content of the undergraduate curriculum in Manchester. *Medical Education* 33 (2): 114–120.
86 Wimmers, P.F. and Fung, C.C. (2008). The impact of case specificity and generalisable skills on clinical performance: a correlated traits-correlated methods approach. *Medical Education* 42 (6): 580–588.
87 Dornan, T., Litttlewood, S., Margolis, S.A. et al. (2006). BEME guide. How can experience in clinical and community settings contribute to early medical education? A BEME systematic review. *Medical Teacher* 28 (1): 3–18.
88 Fish, D. and Coles, C. (2005). *Medical Education: Developing a Curriculum for Practice*. Maidenhead: Open University Press.
89 Case Western Reserve University School of Medicine (2012). http://bulletin.case.edu/schoolofmedicine (accessed 27 June 2018).
90 D'Eon, M. and Crawford, R. (2005). The elusive content of the medical school curriculum: a method to the madness. *Medical Teacher* 27 (8): 699–703.
91 Harden, R.M. and Davis, M.H. (1995). AMEE guide no. 5. The core curriculum with options or special study modules. *Medical Teacher* 17 (2): 125–148.
92 Alahlafi, A. and Burge, S. (2005). What should undergraduate medical students know about psoriasis? Involving patients in curriculum development: modified Delphi technique. *British Medical Journal* 330: 633–636.
93 Syme-Grant, J., Stewart, C., and Ker, J. (2005). How we developed a core curriculum in clinical skills. *Medical Teacher* 27 (2): 103–106.
94 Bruner, J. (1977). *The Process of Education*, 2e. Cambridge, MA: Harvard University Press.
95 Harden, R.M., Davis, M.H., and Crosby, J. (1997). The new Dundee medical curriculum: a whole that is greater than the sum of the parts. *Medical Education* 31: 264–271.

96 Harden, R.M. and Stamper, N. (1999). What is a spiral curriculum? *Medical Teacher* 21 (2): 141–143.
97 Harden, R.M., Laidlaw, J.M., and Hesketh, E.A. (1999). AMEE guide No. 16: study guides – their use and preparation. *Medical Teacher* 21 (3): 248–265.
98 Albanese, M.A. and Mitchell, S. (1993). Problem-based learning: a review of the literature on its outcomes and implementation issues. *Academic Medicine* 68: 52–81.
99 Gibbs, G. (1992). Improving the quality of student learning through course design. In: *Learning to Effect* (ed. R. Barnett), 149–165. Buckingham: SRHE and Open University Press.
100 Biggs, J.B. (1989). Approaches to the enhancement of tertiary teaching. *Higher Education Research and Development* 8 (1): 7–25.
101 Larsen-Freemen, D. (2012). On the roles of repetition in language teaching and learning. *Applied Linguistics Review* 3 (2): 195–210.
102 Grant, J. and Marsden, P. (1987). The structure of memorised knowledge in students and clinicians: an explanation for diagnostic expertise. *Medical Education* 21: 92–98.
103 General Medical Council (2012). Continuing Professional Development. Guidance for all Doctors. https://www.gmc-uk.org/education/standards-guidance-and-curricula/guidance/continuing-professional-development (accessed 27 June 2018).
104 Murray, E., Jolly, B., and Modell, M.A. (1999). A comparison of the educational opportunities on junior medical attachments in general practice and in teaching hospital: a questionnaire study. *Medical Education* 317: 170–176.
105 Bligh, J. (1999). Is it time for a community-based medical school? *Medical Education* 33 (5): 315.
106 Nazareth, I. and Kaya, M. (1999). Medical education in the community – the UNITRA experience. *Medical Education* 33: 722–724.
107 Worley, P., Silagy, C., Prideaux, D. et al. (2000). The parallel rural community curriculum: an integrated clinical curriculum based in rural general practice. *Medical Education* 34: 558–565.
108 McCrorie, P., Lefford, F., and Perrin, F. (1994). *Medical Undergraduate Community-Based Teaching: A Survey for ASME on Current and Proposed Teaching in the Community and in General Practice in UK Universities*. Edinburgh: ASME. Association for the Study of Medical Education.

Further Reading

Apple, M. (2004). *Ideology and Curriculum*, 3e. Abingdon: Routledge.
Flinders, D.J. and Thornton, S.J. (2004). *The Curriculum Studies Reader*, 2e. New York, NY: Routledge Falmer.
Kelley, A.V. (2009). *The Curriculum: Theory and Practice*, 6e. Thousand Oaks, CA: Sage Publications Ltd.

6 Instructional Design: Applying Theory to Teaching Practice

Linda Snell[1], *Daisuke Son*[2], *and Hirotaka Onishi*[3]

[1]Centre for Medical Education and Department of Medicine, McGill University, Montréal, Quebec, Canada
[2]Department of Medical Education Studies, International Research Centre for Medical Education, University of Tokyo, Tokyo, Japan
[3]Department of International Cooperation for Medical Education, International Research Centre for Medical Education, University of Tokyo, Tokyo, Japan

KEY MESSAGES

- There is a 'know-do' mismatch between the science of Instructional Design (ID) and its implementation in practice.
- This difference can be addressed by using learning, education, and instructional theories to inform ID models and instructional strategies.
- ID starts with an analysis of the learner's needs, level, and motivation.
- ID sets learning goals (for example curricular outcomes or competencies to be acquired) and focuses on developing effective instructional strategies to address the gap between these goals and results of the learner analysis.
- An iterative model, with a spiral of 'design and develop, implement, evaluate' followed by reflect and repeat is efficient.
- Active learning methods should be emphasised to enhance intrinsic motivation and personalising the learning.

Introduction

It has been said that despite the findings of education research, 'a "know-do" gap exists between the science of instructional design and its implementation within medical education settings' [1, p. 536]. Cees van der Vleuten has aptly described 'a remarkable difference in attitude between university staff as teachers and as researchers'. Researchers use a scientific approach, look for underlying theory and supporting evidence, and rigorously train new scholars. However, in education: 'as teachers we seem to have a different attitude. We do the things we do, because that is … the way it has been done for many years, even centuries' [2, p. 246]. We are convinced what we do as teachers is correct, do not use evidence in education, and feel that a professional qualification (e.g. as a physician or scientist) has adequately prepared us to teach. Instructional design (or ID) directly addresses this issue by helping a medical educator use underlying theory and principles to create better learning experiences. In this chapter, we will discuss what ID is and differentiate it from curriculum development, why ID is important to medical educators, the education theories underpinning ID and leading to ID models, and the design considerations including the use of relevant teaching and learning strategies. These design principles will be applied to example cases to demonstrate their use throughout the chapter.

What is Instructional Design?

Instructional design is the practice of 'creating instructional experiences which make the acquisition of knowledge and skill more efficient, effective, and appealing' [3, p. 6], basing decisions on sound education (learning and instructional) theory. In other words, ID involves designing experiences 'to help people learn better. It describes a variety of methods of instruction (different ways of facilitating human learning and development), and when to use – and not use – each of those methods' [4, p. xii]. The ID process is systematic, ensures the quality of instruction, and includes an analysis of learning needs and goals, the development of a delivery system to meet those needs, the development of instructional materials and activities, and the trial and evaluation of all instruction and learner activities [5].

Mager has suggested that the instructional designer's task is to answer three questions:

- Where are we going (goals of instruction)?
- How will we get there (strategies and media)?
- How will we know when we have arrived (assessment strategies, programme evaluation)? [6].

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

In What Contexts in Medical Education can Instructional Design be used?

ID models provide a framework or guidance for teachers in health professions education as they develop their class, online module, course, simulation scenario, continuing education conference, or clinical teaching session. ID principles and theories apply to formal large group classes, small group discussions, workshops, laboratory teaching, or online learning. ID can (some would say should) also be used to improve teaching in the clinical context such as bedside teaching or learning procedural skills. The clinical setting offers several challenges to the clinical teacher: these include the opportunistic nature of the cases, the different levels of learners, a potential for the teacher to not be an 'expert', the presence of the patient, and the potential conflict of teaching and clinical care.

In medicine, where professionals continue to learn across the 'continuum of medical education', ID can be used for the instruction of medical students, postgraduate trainees, and physicians in practice. All of these groups have specific characteristics that must be incorporated into ID decision-making. In some other areas of medical education (e.g. early medical school classes, formal instruction in residency, continuing medical education), the instructor delivers a scripted lesson plan written by a central organization or national/state government. Clinical teaching is often opportunistic and while each teacher may have created scripts through experience, these are not determined in advance by an external entity.

What is the Difference between Instructional Design and Curriculum Design?

ID is different from curriculum design (CD) but the two are complementary. CD is about the content or material (including facts, concepts, models, topics, vocabulary, and standards) that a student learns. CD also addresses the order in which that material is presented and very generally how it is presented. ID in contrast, focuses in more detail on how the material is taught, sometimes known as pedagogical methodology, or 'a set of procedures that a teacher can develop in order to help all students learn' [7 p. 2]. A purist might say that ID addresses the 'how' of teaching, whereas CD addresses the content or 'what'. However, there is a large amount of crossover, and in some definitions, ID is subsumed within CD.

Why do we need Instructional Design?

ID is vitally important when planning a learning intervention: it helps the designer or teacher to create effective, efficient, and engaging learning experiences. Well-designed instruction is student-centred and results in better learning [8]. Smith and Ragan have pointed out three other advantages of using systematic ID [9]:

1 ID facilitates a team effort: the collaboration and coordination of designers, developers, and those who will implement the instruction.
2 ID promotes 'congruence' among goals, teaching and learning strategies, and assessment methods such that what is being taught and assessed is what is needed.
3 ID also provides a systematic framework for dealing with learning problems.

Finally, effective ID in medical education not only benefits the learner, but also benefits the teacher, institution, and eventually the patient [9].

Theoretical Concepts Informing Instructional Design

> 'The art of instructional design is knowing when and how to apply the science of learning principles' [1, p. 536].

If ID refers to 'the systematic and reflective process of translating principles of learning and instruction into concrete plans for materials, activities and resources' [9, p. 4], then it is important to examine the relevant theories, concepts, and principles. Learning theories attempt to describe what is going on when people learn. Many of the education theories described elsewhere in this volume can be applied to ID, and Chapter 4 discusses the implications for education practice of many of these. In Box 6.1 we describe some of the principles of ID and link these with the relevant learning theories. It is likely that most teachers and designers do not adhere strictly to one theory, but choose the approach most effective for the context.

In most of theories described in Box 6.1 the learner is seen as actively involved in their own learning, using their prior knowledge and experience to enhance learning. Any design should encourage learners to reflect and allow them some control over learning goals, content, and strategies. In medical education, instruction using real-life problems and activities in authentic settings will allow more effective learning. ID practice in medical education should also take into consideration the fact that learning may occur in groups as well as individually.

Instructional Design Models Relevant to Health Professions Education

A model is not a theory. Theories are conceptual frameworks; models apply the theory or link the theory to practice. Many models of ID have been proposed (see some examples in Box 6.2) and a number have been linked to the complex contexts of medical education. We have chosen to describe in more detail four that we find useful in medical education, and then we will look at their common elements.

The ADDIE Model

One of the most fundamental ID models is ADDIE: Analyse, Design, Develop, Implement, and Evaluate. This is cyclical, with 'evaluate' being followed by another 'analyse'. This model is similar in structure to Kern's six-step approach to curriculum development (see chapter 5) and draws on educational engineering [20]. An early ADDIE model originally developed for the US Army details a series of steps for each stage (Figure 6.1) [21]. Using these steps novice instructors can design an effective education programme. However, some experienced instructors might view these steps as too linear or detailed to use regularly.

BOX 6.1 Linking instructional design principles with theoretical concepts

Instructional design 'principle'	Relevant theory or concept (and associated authors)
Design should allow activities that will enable learners to discover knowledge for themselves and to build on what they already know and can do. The role of the teacher is not to transmit knowledge but to facilitate learning.	Constructivism (Piaget, Vygotsky)
Design should consider the learner's background and use their experience, foster active involvement in the learning process, and establish a positive learning environment where learners feel safe and can express themselves.	Andragogy (Knowles)
Learning methods should be consistent with what we know about how information is processed, stored, and retrieved. This will reduce cognitive processing that does not serve an instructional objective, manage processing aimed at representing the essential material in working memory, and foster processing aimed at making sense of the material. This is particularly applicable to multimedia, clinical learning [10].	Cognitivism and cognitive load theory (Merrill, Gagné, Bruner, Sweller)
The teacher guides the learner in the process of acquiring and using cognitive or psychomotor skills, through a series of steps: modelling, coaching, scaffolding, fading, reflection, articulation, and exploration.	Cognitive apprenticeship (Collins [11])
Design should allow reflection and self-assessment, for example using debriefing, feedback, portfolios, journals, logbooks, trigger questions, watching self (e.g. on video), and explicit modelling of teacher reflection in or on action.	Experiential learning Reflection, reflective practice (Kolb, Schön)
Design should allow learners opportunity for deliberate practice and feedback, where learners master each part or level prior to progressing.	Mastery learning, deliberate practice (Bloom, Ericsson [12])
Teachers should reinforce motivation to learn by demonstrating benefit or usefulness, exploring expectations, linking theory and practice, and using motivating assessments and feedback.	Motivation theories Cook, Pelaccia [13–15])
Design should include participation in authentic (or close to authentic) settings in the real world, allowing the learner to observe experts and provide social interaction with different levels of participation that make the student willing and able to learn (readiness).	Situated learning Communities of practice (Lave, Wenger)
Design should involve the learner in the planning (diagnosing their own needs, developing objectives, choosing resources, and strategies) and evaluation of their instruction. It should build in opportunity for learners to practise skills that improve self-directed learning by asking questions, appraising new information, identifying gaps in knowledge and skills, and reflecting on learning process and outcomes.	Self-directed/self-regulated learning (Panadero [16])

Note: these theories are discussed in more depth in many chapters of this book. References are provided above though where they are particularly relevant or the topic is not explained elsewhere.

BOX 6.2 Models of instructional design

ADDIE (Analyze, Design, Develop, Implement, Evaluate) – Gagne [22, 23]
AGILE (Align, Get set, Iterate and implement, Leverage Evaluate) – Gottfredson [17]
ARCS (Attention, Relevance, Confidence, Satisfaction) – Keller [25]
Cognitive Training Model – Sink [29]
4C/ID (Four component ID model) – van Merrienboer [26]
Kemp model – Morrison [18]
Nine events of instruction – Gagne [22, 23]
Pebble in a pond – Merrill [30]
Systematic design of Instruction – Dick & Carey [19]

An Example Using the ADDIE Model

Dr T. is the Director of Patient Safety; the hospital quality assurance data has shown an increase in medical errors. Despite dissemination of the data, most hospital staff seem unaware of the implications and actions that are needed. Dr T. is concerned that patient safety issues are not being addressed, and wants you to help her design a programme to address a number of aspects including awareness of the issues, prevention, interprofessional communication, and disclosure of adverse events. How will you proceed?

You decide to use the ADDIE model to advise Dr T. You recognise that the general objectives have already been derived from the needs data, but a learner analysis is needed. Focus groups held with key stakeholders show

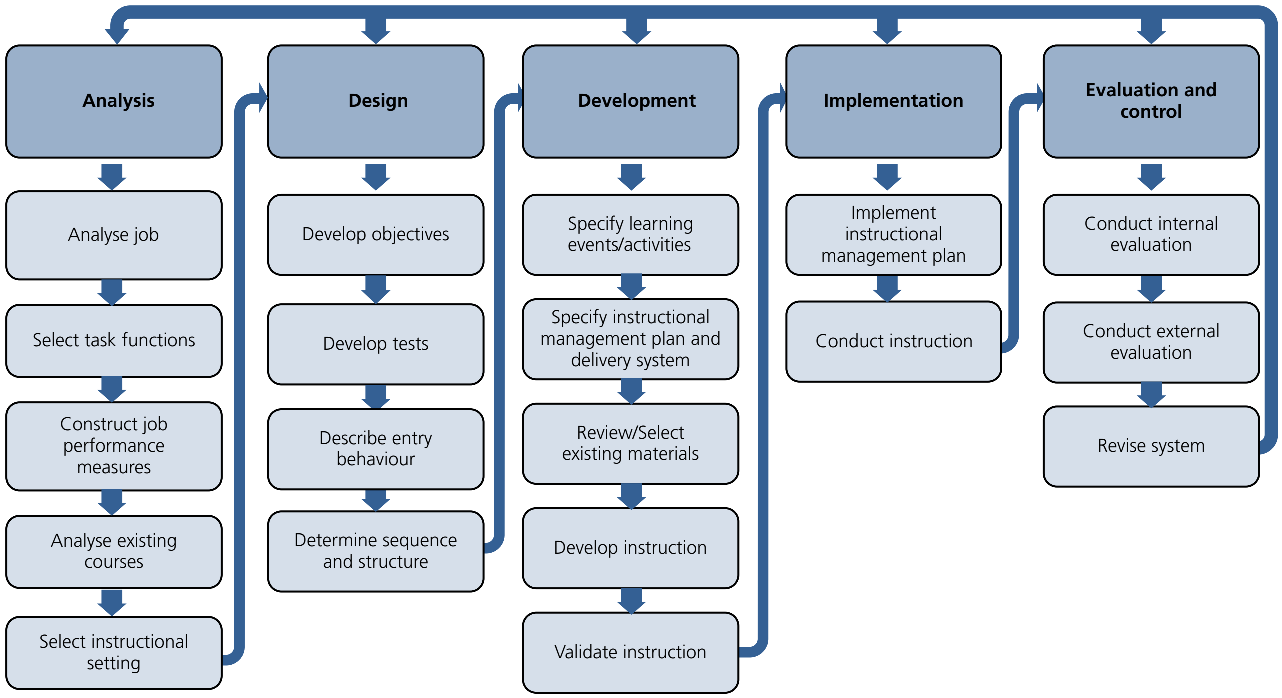


Figure 6.1 Detailed ADDIE model.

that they are not aware of the problem, and that they prefer to learn with their colleagues in the workplace. From this you develop objectives addressing knowledge (of patient safety issues), attitudes (awareness and willingness to change), and skills (communication and disclosure). You decide that three strategies will be used: a brief online module to provide knowledge, which will be mandatory for all staff; a presentation to be done at formal multidisciplinary rounds for each unit, which will provide local quality improvement data and discuss patient safety strategies; and an interdisciplinary simulation activity that can be done in the hospital, where interprofessional communication and disclosure of adverse events can be practised. Evaluation will be done using pre-post quizzes for the rounds presentation, follow-up focus groups, and a longitudinal look at changes in the quality assurance data.

Gagné's Nine Events of Instruction

Gagné and colleagues used several instructional practices and learning theories, drawing particularly on motivation theories, mastery learning, and cognitivism [22, 23]. He regarded instruction as the external conditions or stimuli to promote learning, and listed nine 'events of instruction' to facilitate this. The 'How to' box (Box 6.3) shows these instructional events and their relation to learning processes and strategies that can be used. The list of instructional events will help instructors design experiences for better learning. To use the list, designers should first target one desired learning outcome (intellectual skill, cognitive strategy, verbal information, attitude, motor skills) then choose instructional 'events' which will differ depending on the desired outcome.

An Example using Gagné's Nine Events

Dr Y. has been in clinical practice for 15 years. She knows how to diagnose and manage her patient problems but is not confident that she understands what medical students at different levels need to know. She learned about the 'One-minute preceptor' approach in a recent faculty development workshop and thinks it might help her analyse the need of each medical student [24]. When a resident early in the first year of training presented a case of a 60-year-old man with upper abdominal pain Dr Y. thought it was important to exclude coronary heart disease. However, rather than immediately requesting urgent diagnostic tests she asked the resident for a differential diagnosis and the reasoning behind it – it did not include heart disease. Dr Y. then explained the possibility of heart disease and necessity of tests to exclude it. The resident looked disappointed but the teacher did not understand why. Dr Y. comes to you asking how to improve her teaching around cases in the clinical setting. What will you advise?

After discussion with you Dr Y. realised that as part of work-based learning the resident needs to improve clinical reasoning skills, and she needed to find out what the resident already knows. She used the 'One-minute preceptor' approach, but she was not sure how to ask effective questions to assess the learner's prior knowledge or level, or how to focus the learner on specific goals. The first three steps of Gagné's Nine Events of instruction could be helpful, with an introduction that would gain attention and inform the learner of the objectives of the session, e.g. the importance of ruling out other serious diseases. An example of stimulating recall in the case would be asking the resident to differentiate visceral, somatic, and referred pain, or the mechanisms of referred

BOX 6.3 HOW TO: use Gagné's Nine Events of Instruction with related instructional techniques

	Instructional event	Relation to learning process	Instructional techniques
1.	Gaining attention	Reception of patterns of neural impulses	Stimulate or appeal to the learners' interests.
2.	Informing learner of the objective	Activating a process of executive control	Let them understand what they will be able to do after learning.
3.	Stimulating recall of prerequisite learning	Retrieval of prior learning to working memory	Intellectual skill: Recall prerequisite rules and concept. Cognitive strategy: Recall simple prerequisite rules and concept. Verbal information: Recall well-organised bodies of knowledge. Attitude: Recall the situation and action involved in personal choice; remind learner of the human model and the model's characteristics. Motor skill: Recall the 'executive subroutine,' and part-skills if appropriate.
4.	Presenting the stimulus material	Emphasising features for selective perception	Intellectual skill: The skill is explained and demonstrated, often with a variety of examples from a variety of contexts, to facilitate generalisation of the skill. Cognitive strategies: when and how to employ the strategy should be explained and demonstrated with examples. Verbal information: should be delivered visually and orally, organised in ways that are meaningful to the learners. Attitude: The situation requiring a choice of action is made clear, and the preferred choice is demonstrated by a respected human model. Motor skills: should be demonstrated, with emphasis on the executive subroutine and the stimulus features that cue each action.
5.	Providing learning guidance	Semantic encoding; cues for retrieval	Intellectual skill: Give varied concrete examples of concept or rule. Cognitive strategy: Provide verbal description of strategy, followed by example. Verbal information: Elaborate content by relating to larger bodies of knowledge; use images, mnemonics. Attitude: Human model describes or demonstrates action choice, followed by observation of reinforcement of model's behaviour. Motor skill: Continue practice, with informative feedback.
6.	Eliciting the performance	Activating response organisation	Ask a learner 'Do and show it to me.'
7.	Providing feedback about performance correctness	Establishing reinforcement	Give specific and motivational feedback to the learner to improve performance.
8.	Assessing the performance	Activating retrieval; making reinforcement possible	Assess the learner in reliable (multiple observations) and valid (relevance to objectives and width of information sources) manner.
9.	Enhancing retention and transfer	Providing cues and strategies for retrieval	Give similar and advanced activities for the learner to check if the performance is good.

Source: Adapted from Gagné et al. [22, 23].

abdominal pain. She could then build broader differential diagnoses based on this knowledge.

ARCS Model

How to motivate learners is a key issue in ID. 'ARCS' is an acronym for Attention, Relevance, Confidence, and Satisfaction, categories created from the literature about motivation introduced by Keller and discussed in Gagné's book [22, 25]. Box 6.4 is a list of process questions in the categories of the ARCS model that an instructional designer should ask themselves. In this model, instructors and others who present (performers, writers, movie makers, etc.) use the steps of drawing attention of participants, maintaining the attention for the entire programme, adapting

BOX 6.4 Motivational categories of the ARCS model

Categories and subcategories	Process questions
Attention	
Perceptual arousal	What can I do to capture their interest?
Inquiry arousal	How can I stimulate an attitude of inquiry?
Variability	How can I maintain their attention?
Relevance	
Goal orientation	How can I best meet my learners' needs? (Do I know their needs?)
Motive matching	How and when can I provide my learners with appropriate choices, responsibilities, and influences?
Familiarity	How can I tie the instruction to the learners' experiences?
Confidence	
Learning requirements	How can I assist in building a positive expectation for success?
Success opportunities	How will the learning experience support or enhance the students' beliefs in their competence?
Personal control	How will the learners clearly know their success is based on their efforts and abilities?
Satisfaction	
Natural consequences	How can I provide meaningful opportunities for learners to use their newly acquired knowledge/skill?
Positive consequences	What will provide reinforcement to the learners' successes?
Equity	How can I ensure consistent consequences for meeting standard criteria?

Source: From Gagné et al. [22, p. 118].

the contents relevant to them, facilitating them to provide learners with a sense of control, and satisfying them. This model draws on motivation theories, constructivism, and adult learning concepts.

An Example using ARCS Model

Dr S. is a basic science teacher new to the university, who planned a lecture series on physiology for junior medical students. Yesterday he gave the first lecture and found he was not satisfied. In fact, he was angry with the students, many of whom fell asleep as the lecture went on. He did not understand why, as he was confident that he was teaching essential physiology content. As there was a lot to teach he included more than 100 PowerPoint slides in the 90-minute didactic lecture. He thought that cutting-edge topics on physiological research, including his own work, would entertain and engage the students, and he is excited about what he taught. Dr S. is concerned that the same thing will happen in subsequent lectures and comes to you for advice. What will you tell him?

With your guidance, Dr S. decided to use an ID approach. First, he interviewed a few students who said that his lecture was 'interesting' because of cutting-edge research discussion, but difficult to understand and link with clinical knowledge. He revised and clarified the objectives to include understanding of basic physiology and application to disease mechanisms. He followed the ARCS model to motivate students. At the beginning, he presented a clinical case and asked 'What would you do in such a case?' in order to gain attention. He explained the physiological knowledge needed to understand the case (relevance). He designed active learning by group discussion with problem solving, then provided explanations (confidence and satisfaction). He concluded with direction for future learning on other areas of basic physiology, including some 'cutting-edge' research findings (satisfaction).

4C/ID Model

The four-component instructional design model (4C/ID-model) is an evidence-based ID model suggested by van Merriënboer and Kirschner [26]. It draws from multiple frameworks including cognitivism and complexity theory. Although the model can be used generally, it is especially useful when designing the process of complex learning in medical education. Examples would be a full course or a clinical rotation, rather than one lecture or a brief clinical discussion. Complex learning is 'the integrated acquisition of knowledge, skills, and attitudes and the coordination of a variety of constituent skills, and involves the transfer of what is learned in school and training settings to professional settings' [27, p. 2]. This model also provides evidence-based principles for education practice. The model consists of four major components, all interrelated, and with each part contributing to the development of the skill (see Figure 6.2). The four components are (i) the *learning tasks*, authentic experiences which together comprise a broad skill or ability in medical education, organised from simple to complex or easy to difficult, with progressively decreasing support; (ii) *supportive information* provided at any time, which gives the learner a high level of information on how to approach or organise the task; (iii) task-specific *procedural information*, provided as needed ('just-in-time' instruction) to support and build the routine aspects of learning; and (iv) additional *part-task practice* so that

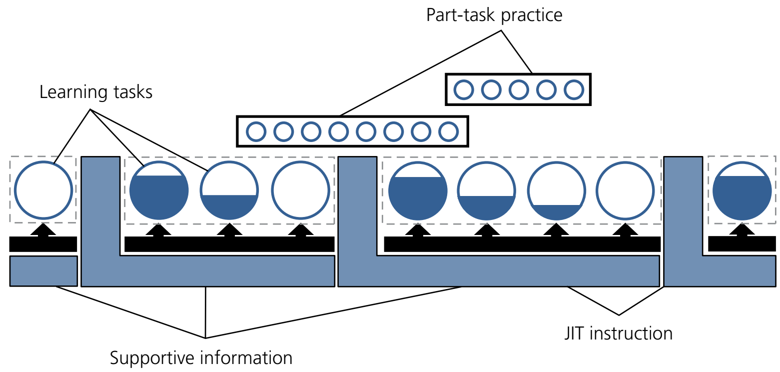


Figure 6.2 Focus on: the four component instructional design (4C/ID) model.

selected, often critical, aspects of the task can be repeated and strengthened until automatic.

Figure 6.2 can be interpreted as follows, adapted from Vandewaetere et al. [28]:

- Circles are learning tasks which provide the backbone of the educational programme.
- Filling of the circles indicates amount of learner support and guidance.
- Support and guidance will typically decrease as learners acquire more expertise.
- An L-shaped figure is supportive information which helps learners to perform non-routine aspects of tasks: aspects that require problem solving, reasoning, and decision-making.
- A rectangle with upward pointing arrows is procedural information or just-in-time information, offered at the task level to inform learners how to perform 'recurrent' aspects to be performed as routines after the programme has been completed.
- The small circles grouped in a rectangle represent part-task practice or repetitive practice provided when a high level of automaticity is required for an aspect of the task.

When putting the 4C/ID model into a design process, 10 steps or activities are proposed to organise and transfer learning materials, as listed below [27]:

A. Learning tasks
  1 Design learning tasks: design a set of typical learning tasks that represent the whole complex skill.
  2 Sequence learning tasks: order tasks in such a way that the learning process is optimised – by increasing complexity and by decreasing support and guidance.
  3 Develop assessment instruments: articulate the to-be-reached standards, in order to inform learners about the criteria or performance goals.

B. Supportive information
  4 Design supportive information: connect units of supportive information to task classes; more complex task classes require more or more elaborated supportive information.
  5 Analyse cognitive strategies: identify the cognitive strategies that proficient task performers use to solve problems as presented in the task classes.
  6 Analyse mental models: analyse the mental models that describe how the domain is organised.

C. Procedural information
  7 Design procedural information: connect units of procedural information to tasks, provide just-in-time how routine aspects of tasks needs to be carried out.
  8 Analyse cognitive rules: identify the condition–action pairs (i.e. cognitive rules) that drive routine behaviours.
  9 Analyse prerequisite knowledge: analyse the knowledge that is prerequisite to a correct use of cognitive rules.

D. Part-task practice
  10 Design part-task practice: design part-task practice items with just-in-time information for repetitive practice with feedback.

An Example using the 4C/ID Model

Dr V. has been asked to teach a course on advanced life support (ALS) skills to senior medical students, and knows that there are standard courses which do this. However, many of these courses are aimed at learners at a higher level of training and Dr V. is not sure that an existing course will be effective for the students.

Dr V. uses the 4C/ID model to assess the level of the students and revise the existing course. Dr V. starts by outlining the *learning tasks* that are needed for performing ALS, and sequences them in a logical order from simple to complex. She then adds to the material from the existing course to provide cognitive support (*supportive information*); the goal of this is to add to what the students have already learned in anatomy and pharmacology and from their early clinical rotations. As it is the first time the students will be performing many resuscitative skills, they break down each into a series of steps and algorithms that lead to different actions (*procedural information*). Finally, Dr V. designs the course so that there is ample time for the students to practise the needed skills (especially the complex parts) and receive feedback (*part-task practice*).

Common Elements and Issues

There are elements common to the ID models described above. Analysis of the learner is the first step in many models. This includes developing an understanding of their current level of knowledge, skill, or performance, and their motivation to learn. Determination of the goal of learning – both general and specific – is the next step. Comparing the goal with the learner analysis provides a gap, or learning need. Instructional experiences are designed and implemented to address this gap. The best instruction is effective (facilitates learning), enduring (encoded in long-term memory), efficient (in terms of time and resources), and appealing (interests and motivates the learner). Assessing whether the learner has acquired the needed capabilities follows, and finally evaluating the ID process with the goal of revising the process closes the loop. The ADDIE and Gagné models list the desired learning outcomes and then select the most appropriate learning strategy to achieve the desired goal. The ARCS and 4C/ID models focus more on the instructional experiences needed to learn the complex abilities needed for health professionals. Those involved in ID must recognise completing these ID activities takes time, which in turn may inhibit implementation. Adopting an iterative model, with a spiral of 'design and develop, implement, evaluate', followed by reflection and repeat, will be more efficient.

Design Considerations

Many of these models were developed to address classroom learning; however, they apply (with some variation) equally well to other contexts such as clinical learning,

BOX 6.5 Questions to ask during the design process

'W5'	Classroom, non-clinical setting	Clinical setting
Who	What is the learners' prior knowledge and experience? Are the learners motivated? Is it a mandatory or elective class? Will there be team teaching or teaching assistants available? What are the strengths and weaknesses of the teacher? What is the relationship with the teacher and students (e.g. one class, ongoing)?	What is the learners' level, past clinical experiences, interests, and characteristics? How will you find out? How many learners are there and are they at different levels? What is the teacher's content knowledge or skill? What is the teacher's experience, strengths, and weaknesses regarding the specialty? What is the teacher's experience with specific strategies?
What	Are learning objectives practical and articulated so they can be assessed? Are the objectives feasible? Are they related to the prior knowledge or learning? How are current learning objectives positioned in the whole curriculum? Resources: Is there support to develop resource materials? What kind of materials will you use as education resources? (text, audiovisual, handouts, simulated patients, e-learning, mobile devices, etc.)	How will you choose the content/objectives, given the opportunistic nature of clinical setting? (e.g. based on case mix, learner level, learner need) Is the patient or case mix appropriate for the learners? Any clinical 'triage' issue (e.g. sick patient, clinical demands)?
When	Are the topics appropriate to be taught in the time provided in relation to other topics in the curriculum? How much time do you need to teach the content? Should the content be broken up into parts, with different activities?	How much time will it take to teach? How much time is actually available? Is teaching at the bedside or with the patient present needed for all the time, and if not, what *must* be taught at the bedside? Will the teacher or learner be involved in concurrent patient care?
Where	What kind of classroom or other space will you need? Is there space for group interaction? Is there access to electronic devices? What are the available classroom resources? (e.g. white board, projector, speakers)	Is patient safety assured? Is patient confidentiality and privacy ensured? Is the space an appropriate size for the number of learners? Is there access to online resources in the clinical setting for just-in-time learning?
How	Does the teaching strategy match the learning objective? Have you provided time for practice and feedback (formative assessment)? How will you use active learning? How might classroom simulation be used? How will you motivate the learners?	Does the teaching strategy match learner level and independence, learning goal, patient considerations? Can simulation be used if the case mix is not conducive?

simulation learning, and e-learning. In Box 6.5 we outline some specific considerations for these settings in the form of questions to ask during the design process. These questions were developed from the authors' experience and from practical tips and advice provided in the descriptions of the models.

To acknowledge the learning theories and concepts and the ID models discussed above, instructional strategies in medical education, whether in the classroom, clinical, or online context, should emphasise active learning methods, self-reflection, just-in-time instructional support, and ongoing formative assessment. Strategies should enhance intrinsic motivation, personalise the learning, and empower the learner. Collaborative methods and peer learning are good options in health professions education as they can lead to outcomes other than solely knowledge acquisition, such as teamwork and communication skills. The role of the teacher should be to assist learners in goal setting and selecting tasks, facilitate, coach, scaffold, mentor, and assess; the role of learners is to be active, self-regulated, and to collaborate with peers [4]. Current and emerging education technology can assist in planning, record keeping, and providing instructional support.

Sink and colleagues proposed that cognitive approaches to training such as Gagné's nine events of instruction discussed earlier in this chapter should be linked to strategies and tactics for helping learners acquire cognitive skills [29]. For instance, the tasks learners do to learn should be associated with the elements designers put into lessons to facilitate learning. These elements provide a link between the theoretical and practical, or between education concepts and designer activities.

These authors describe five learner tasks along with associated methods and strategies, which are summarised here.

1. The learner must select information to attend to, focusing on new knowledge. Designers can focus learners' attention on this knowledge using techniques to tell learners 'WIIFM – What's in it for me?' or 'YCDI – You can do it' regarding learning the new knowledge.
2. Learners need to put new knowledge in an existing framework by recalling prior knowledge and relating the new knowledge to the old. Designers can use strategies to enhance the recall of prior knowledge onto which new knowledge is built. They can also relate, or compare and contrast, new and old knowledge, so that the new knowledge is tied to the old.
3. Learners need to organise the new information in a way that relates to the organisation of existing knowledge; this makes learning easier, stresses relevant information, and decreases confusion. Designers can facilitate this by structuring content, specifying desired behaviours and goals of learning, limit the amount of content to match human information processing capacity, and use visual aids such as text lay-out and diagrams to assist learners' organisation and assimilation.
4. Learners need to integrate new and old knowledge to produce a 'new unified, expanded, and reorganized set of knowledge'. Designers can foster this by presenting knowledge in a way that makes it easiest to understand, and by using real-life examples.
5. Learners need to strengthen new knowledge so that it will be remembered and can be used in the future. Designers can facilitate this by incorporating practice and feedback, summaries, and opportunities to use the knowledge in authentic contexts.

A number of instructional methods have been listed in Chapter 4. These include lectures, symposia and other large group sessions, seminars, tutorials, workshops and other small group sessions, independent work, guided reading, e-learning (individual or group), technology-enhanced learning, simulation methods including skills training, standardised patients, high fidelity activities, practical or clinical experiences in various settings, and formative assessment. Most of these strategies are described in detail in other parts of this book.

Whatever the instructional method, learning is promoted when:

- learners are engaged in solving real-world problems
- existing knowledge is activated as a foundation for new knowledge
- new knowledge is demonstrated to the learner
- new knowledge is applied by the learner
- new knowledge is integrated into the learner's world [30].

In the health professions, there are generic goals of learning, usually related to overarching competencies needed by graduates. These competencies are usually organised into frameworks such as CanMEDS [31] used in Canada and other countries or the ACGME competencies [32] used in the USA. Acquiring these competencies may occur in multiple contexts and each goal may be attained using a number of potential instructional methods, which will differ in clinical, classroom, and online contexts. The challenge is matching the instructional or learning method to the goal, and we have proposed here some instructional strategies to best achieve each goal (see Box 6.6).

Evaluation

Programme evaluation in medical education has been described as the 'systematic approach to the collection, analysis and interpretation of information about any aspect of the conceptualization, design, implementation and utility of education programmes' … 'for subsequent judgement and decision-making' [34–36]. Any ID should be evaluated to see if it has been effective in achieving goals and addressing learner needs and the gap. Measures of success should be determined early in the design process. Typical areas to evaluate might include instructional materials and strategies, quality of teaching, assessment instruments, resource use and return on investment, support from the organisation, as well as whether the steps of the design process have been executed. The aim is to understand and improve the ID, so the evaluation looks at the process and elements of the design as well as the outcomes. In many ways, it is like a programme evaluation (Chapter 30); however, it focuses much more on the actual design process and outcomes. Relevant models of

BOX 6.6 Where curriculum design meets instructional design: linking instructional methods with goals of learning

Goal of learning	Potential instructional methods or learning strategies
Acquire basic biomedical and clinical knowledge	Lectures, including variants such as symposia, panel discussions, debates Directed reading Self-instructional modules, including online Technology-enhanced learning
Apply knowledge to the diagnosis and management of patient problems	Case-based methods Problem-based learning Small group discussions, seminars High fidelity simulation Work-based learning Clinical supervision with feedback
Obtain information from patient (history taking and physical examination skills)	Simulated or standardised patients Role play Work-based learning Clinical supervision with feedback
Develop clinical reasoning skills	Case discussion Online cases Work-based learning Clinical supervision with feedback
Perform procedural and hands-on skills [33]	Task trainers and skills labs Videos, demonstrations Work-based learning Clinical supervision with feedback Coaching Logbooks
Communicate with patients and colleagues	Simulated or standardised patients One-on-one clinical supervision and feedback Mentoring, coaching
Develop collaborative and leadership skills	Basic knowledge acquired through lectures, reading Workshops (e.g. teamwork, leadership) Learn and work in an interprofessional environment Reflective exercises
Work within a system	Work-based learning Mentoring
Develop ethical reasoning	Basic knowledge acquired through lectures, reading Ethics case discussion Debates Portfolios Reflective exercises
Develop critical thinking skills	Journal club Clinical supervision with feedback
Participate in developing and transmitting new knowledge	Research training/mentorship Teacher training/mentorship
Develop professional behaviours and identity	Role modelling Mentoring Reflection

programme evaluation that can be used to frame the evaluation step of ID include Kirkpatrick, CIPP, or the Logic Model [37–39].

Often the evaluation process of design–evaluate–revise is iterative or cyclical, occurring throughout the ID process (sometimes called internal or formative) as well as at the end (sometimes called external or summative). The ADDIE model (see Figure 6.3) reflects this concept well, with the circular arrows representing an evaluation and feedback loop. However, even when it is not emphasised in the model being used, some form of evaluation should accompany any ID process.

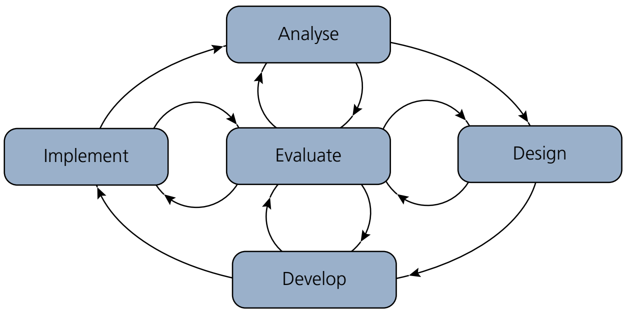


Figure 6.3 The iterative (small circles) and cyclical (large circle) evaluation process in the ADDIE model.

Conclusions

Instructional design is 'the process by which instruction is improved through the analysis of learning needs and systematic development and evaluation of learning experiences' [40]. By using a rigorous design process for classroom, online, or clinical instruction in medical education, the result will be enhanced learning and application of what is learned, better trained health professionals, and ultimately, better patient care.

Acknowledgement

We wish to thank Dr Elaine van Melle for her helpful discussions on the 'Evaluation' section.

References

1 Levinson, A.J. (2010). Where is evidence-based instructional design in medical education curriculum development? *Medical Education* 44 (6): 536–537.
2 van der Vleuten, C., Dolmans, D., and Scherpbier, J.A. (2000). The need for evidence in education. *Medical Teacher* 22 (3): 246–250.
3 Merrill, M.D., Drake, L., Lacy, M.J., and Pratt, J. (1996). Reclaiming instructional design. *Educational Technology* 36 (5): 5–7.
4 Reigeluth, C.M., Beatty, B.J., and Myers, R.D. (2017). *Instructional-Design Theories and Models: The Learner-Centered Paradigm of Education*. New York and London: Routledge.
5 Definitions of Instructional Design, adapted from 'Training and Instructional Design', Applied Research Laboratory, Penn State University. http://www.umich.edu/~ed626/define.html (accessed 27 June 2018).
6 Mager, R.F. (1984). *Preparing Instructional Objectives*. Belmont, CA: Pitman Learning.
7 Teodoro, V.T. and Mesquita, M. (2003). Development of Pedagogical Methodology. https://moodle.fct.unl.pt/pluginfile.php/20500/mod_resource/content/0/diversos/D13_Development_of_pedagogical_methodology.pdf (accessed 20 February 2017).
8 Issa, N., Schuller, M., Santacaterina, S. et al. (2011). Applying multimedia design principles enhances learning in medical education. *Medical Education* 45 (8): 818–826.
9 Smith, P.L. and Ragan, T.J. (2004). *Instructional Design*. Hoboken, NJ: Wiley & Sons.
10 Mayer, R.E. (2010). Applying the science of learning to medical education. *Medical Education* 44 (6): 543–549.
11 Collins, A., Brown, J., and Holum, A. (1991). Cognitive apprenticeship: making thinking visible. *American Educator* 15 (3): 1–18.
12 Ericsson, K.A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine* 79 (10 Suppl): S70–S81.
13 Cook, D.A. and Artino, A.R. (2016). Motivation to learn: an overview of contemporary theories - online supplement Table S1. Summary of practical applications of motivation theory. *Medical Education* 50 (10): 997–1014. medu13074-sup-0001-TableS1-S2.docx.
14 Cook, D.A. and Artino, A.R. Jr. (2016). Motivation to learn: an overview of contemporary theories. *Medical Education* 50 (10): 997–1014.
15 Pelaccia, T. and Viau, R. (2017). Motivation in medical education. *Medical Teacher* 39 (2): 136–140.
16 Panadero, E. (2017). A review of self-regulated learning: six models and four directions for research. *Frontiers in Psychology* 8: 422. 28 April 2017 https://doi.org/10.3389/fpsyg.2017.00422.
17 Clark, T. and Gottfredson, C. (2009, December). Agile learning: Thriving in the new normal. *Chief learning officer* 8 (12): 18–21. http://www.nxtbook.com/nxtbooks/mediatec/clo1209/index.php?startid=18#/20 Accessed Aug 21, 2018.
18 Morrison, G.R., Ross, S.M., and Kemp, J.E. (2004). *Designing Effective Instruction*. Hoboken, NJ: Wiley & Sons.
19 Dick, W., Carey, L., and Carey, J.O. (2009). *The Systematic Design of Instruction*. Upper Saddle River, NJ: Merrill/Pearson.
20 Kern, D.E., Thomas, P.A., and Hughes, M.T. (2009). *Curriculum Development for Medical Education: A Six-Step Approach*, 2e. Baltimore, MD: Johns Hopkins University Press.
21 Branson, R.K., Rayner, G.T., Cox, J.L. et al. (1975). *Interservice Procedures for Instructional Systems Development*. Monroe, VA: Army Training and Doctrine Command.
22 Gagné, R.M., Briggs, L.J., and Wager, W.W. (1992). *Principles of Instructional Design*, 4e. Orlando, FL: Holt, Rinehart and Winston.
23 Gagné, R.M., Wager, W.W., Golas, K.C. et al. (2005). *Principles of Instructional Design*, 5e. Belmont, CA: Wadsworth/Thompson Learning.
24 Neher, J. and Stevens, N. (2003). The one-minute preceptor: shaping the teaching conversation. *Family Medicine* 35 (6): 391–393.
25 Keller, J.M. (1987). Development and use of the ARCS model of instructional design. *Journal of Instructional Development* 10 (3): 2–10.
26 van Merriënboer, J.J.G. and Kirschner, P.A. (2007). *Ten Steps to Complex Learning*. Mahwah, NJ: Erlbaum.
27 van Merriënboer, J.J.G. and Kirschner, P.A. (2012). *Ten Steps To Complex Learning: A Systematic Approach to Four-Component Instructional Design*. New York, NY: Routledge.
28 Vandewaetere, M., Manhaeve, D., Aertgeerts, B. et al. (2015). 4C/ID in medical education: how to design an educational program based on whole-task learning: AMEE Guide No. 93. *Medical Teacher* 37 (1): 4–20.
29 Sink, D.S. (2014). Design Models and Learning Theories for Adults, Section III, Chapter 11: Designing and Developing Effective Learning. Table 11–1. Cognitive Training Model, p. 192.
30 Merrill, M.D. (2002). First principles of instruction. *Educational Technology Research and Development* 50 (3): 43–59.
31 Frank, J.R., Snell, L., and Sherbino, J. (2015). *CanMEDS 2015 Competency Framework*. Ottawa: Royal College of Physicians and Surgeons of Canada.
32 Swing, S. (2007). The ACGME outcome project: retrospective and prospective. *Medical Teacher* 29: 648–654.
33 Cheung, L. (2016). Using an instructional design model to teach medical procedures. *Medical Science Educator* 26: 175–180.
34 Wall, D. (2010). Evaluation: improving practice, influencing policy. In: *Understanding Medical Education: Evidence, Theory and Practice.* (ed. T. Swanwick). Association for the Study of Medical Education.

35 Mohanna, K., Wall, D., and Chambers, R. (2004). *Teaching Made Easy – A Manual for Health Professionals*, 2e. Oxford: Radcliffe Medical Press.

36 Goldie, J. (2006). AMEE guide no. 29: evaluating educational programmes. *Medical Teacher* 28: 210–224.

37 Kirkpatrick, D.I. (1967). Evaluation of training. In: *Training and Development Handbook* (ed. R. Craig and I. Mittel), 87–112. New York: McGraw Hill.

38 Stufflebeam, D.E. (1983). The CIPP model for program evaluation. In: *Evaluation Models* (ed. G. Madaus, M. Scriven and D. Stufflebeam), 117–141. Springer.

39 Van Melle, E. (2016). Using a logic model to assist in the planning, implementation, and evaluation of educational programs. *Academic Medicine* 91, (10) 1464.

40 http://www.instructionaldesign.org (accessed 17 August 2017).

Further Reading

Gagné, R.M., Wager, W.W., Golas, K.C. et al. (2005). *Principles of Instructional Design*, 5e. Belmont, CA: Wadsworth/Thompson Learning.

Keller, J.M. (2010). *Motivational Design for Learning and Performance: The ARCS Model Approach*. New York, NY: Springer.

Mayer, R. (2010). Applying the science of learning to medical education. *Medical Education* 44: 543–549.

Reigeluth, C.M., Beatty, B.J., and Myers, R.D. ed. (2017). *Instructional Design Theories and Models*, vol. IV. New York and London: Routledge.

7 Quality in Medical Education

Diane N. Kenwright[1] and Tim Wilkinson[2]

[1]Department of Pathology and Molecular Medicine, University of Otago, Wellington, New Zealand
[2]Otago Medical School, University of Otago, Christchurch, New Zealand

KEY MESSAGES

- A quality framework or cycle is the basis of all quality improvement.
- Setting or locating appropriate standards or outcomes is required in order to measure quality.
- Choosing appropriate measuring tools is important to make sure the data gathered are reliable and valid.
- Quality assurance identifies and disseminates good practice.
- The ultimate measure of quality is the improvement in patient outcomes.

Introduction

Quality is used as a descriptor at all levels of a course or curriculum; we want the best 'quality teaching', good 'quality assessments', a high 'quality medical curriculum'. All these reflect the use of the word 'quality' to mean a *measure of the degree of excellence* of something. Indeed, 'excellence' can replace quality in many of the sentences in this chapter (even the title) and not significantly change the meaning. Quality assurance is the process that maps our journey to achieving excellence; it finds our less than excellent parts and asks us to improve them. It is a challenging and uncomfortable process.

Quality can be applied to many aspects of an educational programme. Examples include governance, the learning environment, social accountability, selection, the learning outcomes, the teaching and learning methods, assessment, and continuing professional development. In this chapter we aim to provide guidance on processes and tools that could be used, and cover concepts of quality improvement, quality assurance, quality control, benchmarking, and accreditation. How these are applied to which aspects of any programme will depend on the purpose of the activity. Quality is also an emerging and important curriculum content area – what students and trainees should learn about quality and safety improvement – but this is not the focus of this chapter. Similarly, methods for programme evaluation are left to be considered in detail in Chapter 30.

Quality Perspectives

Perspectives on quality – the degree to which excellence is attained – are observer dependent. When teachers see excellence, they see best practice teaching and learning, an environment that encourages lifelong learning, and evaluation processes that result in continuous improvement. When learners see excellence, they see teaching that prepares them for their examinations and practical guidance to become great doctors. When university or course administrators see excellence, they see value for money, external accreditation, and strategic plans being fulfilled. When employers see excellence, they see the production of doctors who are safe and fit for purpose. When hospital administrators see excellence, they see improved patient outcomes, better value, and waste reduction.

In any of these perspectives you measure and document findings – this is called *quality assurance*. After you have measured quality you will find areas that need improvement – acting to improve and then re-documenting them is continuous *quality improvement*. Such documentation should be valid and reliable, often descriptive, sometimes numerical, and should include description of processes and good practice. See Box 7.1.

Quality Standards

Before you begin measuring quality you need a set of standards to measure it against; these may be *internal* or *external*. Internal standards will be found in a variety of places within an institution – in strategic planning documents, graduate attributes frameworks, and policy statements. External standards are developed and published by regulators and by the accreditation bodies of medical schools and postgraduate colleges. Some examples of these are provided in subsequent paragraphs, but other external standards should also be considered – such as those implicit

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

BOX 7.1 FOCUS ON: Quality, cost, and value

Value is defined as outcomes relative to costs.

Cost is measured by the total inputs needed to achieve those outcomes. Outcomes are usually defined by the end user of the service (the patient) and are the actual results of medical care. Safety, or the avoidance of errors, is one type of outcome in the overall set of outcomes for any system.

Quality can be usefully defined in terms of patient outcomes. Thus, quality improvement aims to improve patient outcomes. Cost reduction, without regard to the patient outcomes achieved (the quality), may reduce value. Reducing the total costs involved, not necessarily minimising the cost of individual services, increases value, as expensive but truly effective services increase outcomes more than cost.

To reduce cost and increase value without affecting patient outcomes (i.e. to improve quality), the best approach is often to spend more on some high-value services and less on low-value services.

within government policies, or the wider public expectations of the public, often expressed as social accountability. See Box 7.2.

The World Federation for Medical Education publishes standards on basic medical education, postgraduate medical education, and continuing professional development [1]. It does not accredit programmes, leaving that to the national accrediting bodies, but it does evaluate and recognise the accrediting bodies, examples of which include:

- Committee on Accreditation of Canadian Medical Schools (Canada)
- Liaison Committee on Medical Education (United States of America)
- Korean Institute of Medical Education and Evaluation (Republic of Korea)
- Japan Accreditation Council for Medical Education (Japan).

An example of a WFME postgraduate standard in teamwork is: 'The programme provider(s) must ensure experience of working in a team with colleagues and other health professionals.'

BOX 7.2 FOCUS ON: Social accountability in medical education

Social accountability means that the needs of patients and societies are fundamental in planning and delivering the curriculum. The continuum of adoption of social accountability [5] includes:

- *Social responsibility*: the health education provider is committed to the welfare of society, producing health practitioners to meet society's health needs.
- *Social responsiveness*: the health education provider produces graduates with the skills to care for the society's most vulnerable people and professional values lead them to work in underserved areas.
- *Social accountability*: the health education provider works in partnership with all relevant groups, health care organisations, health professionals, and patient representatives to improve health equity and make the greatest impact on people's well-being.

Social accountability is often measured during the accreditation processes of medical education providers, and can be driven by the need to meet these accreditations.

The ASPIRE programme [6] suggests four domains within which undergraduate health education providers should endeavour to achieve excellence in social accountability:

1. Organisation and function – social accountability is a prime directive in the school's purpose and mandate and is integrated in its day-to-day management.
2. Education of doctors, dentists, and veterinary practitioners – admissions, learning experiences, and faculty development supports social accountability.
3. Research activities – community/regional/national health needs inspire the school's research including knowledge translation.
4. Contribution to health services – the school's graduates and its health service partnerships have a positive impact on the health care and the health of its community/region/nation.

The associated documentation includes:

- Plans, including concepts and goals evident in its organisation and function.
- Actions, evident in its education and research programme activities.
- Impacts, evident in positive effects of its education, research, and service, and its graduates and partnerships on the health care and health of its community/region/nation.

Postgraduate education providers are provided with less explicit standards of social accountability than medical schools. For instance, the UK General Medical Council's 'Excellence by design: standards for postgraduate curricula' includes the standard (CS1.1): 'The curriculum has a stated and clear purpose based on scope of practice, service, and patient and population needs.' [7] This reflects the socially responsible descriptor but not that of social accountability. Jamison et al. [8] were 'unable to find any documents or published literature that describe a systematic process by which health disparities are identified, characterized, and addressed by any of the organizations responsible for overseeing PGME [postgraduate medical education] within Canada'; the ability to undertake postgraduate training in rural and underprivileged settings within family medicine being an exception, reaching the level of social accountability.

The Accreditation Council for Graduate Medical Education (ACGME) is a private organisation that sets standards for US graduate medical education (residency and fellowship) and accredits programmes that comply with these standards. There are 800 ACGME-accredited institutions sponsoring approximately 10 000 residency and fellowship programmes in 150 specialties and subspecialties. The ACGME Common Programme Requirements [2] set standards covering the institution, programme personnel and resources, trainee appointments, educational programme, evaluation of trainees and faculty, and the learning and working environment. An example of an ACGME learning environment standard in teamwork is: 'Residents must care for patients in an environment that maximizes effective communication. This must include the opportunity to work as a member of effective interprofessional teams that are appropriate to the delivery of care in the specialty.'

The Academy of Medical Educators (AoME) sets standards over five domains for medical educators, with levels matching the progress of educators through their careers [3]. Medical educators submit a self-assessment and peer review demonstrating their competence in designing and planning learning, teaching and facilitating learning, assessment of learning, educational research and scholarship, and educational management and leadership. An example of an AoME standard for the learning environment is: 'Monitors and manages the safety and effectiveness of complex learning environments.'

The Committee on Accreditation of Canadian Medical Schools (CACMS) sets standards for the MD degree awarded in the 17 Canadian medical faculties. The CACMS Standards and Elements document and that of the United States MD accrediting body, the Liaison Committee on Medical Education (LCME) [4], comprehensively describe the standards for a medical school, and include standards on medical student health services, personal counselling, and financial aid services that reflect the North American educational environment. An example of a (CACMS) financial aid standard is: 'A medical school provides its medical students with effective financial aid and debt management counselling and has mechanisms in place to minimize the impact of direct educational expenses (i.e. tuition, fees, books, supplies) on medical student indebtedness.'

Quality Processes

The ultimate aim of quality assurance and quality improvement is the improvement of patient outcomes. The contingent outcomes of an individual health education programme are defined in detail, as standards to be met, and the quality assurance of each of these means that the stakeholders, health administrators, teachers, external bodies, and the public know that a programme is delivering the outcomes it advertises. Quality measurement can detect and improve any part of the programme that is underperforming in terms of teaching, learning, or assessment and can also uncover good practice. Detection of underperforming individuals and documenting improvement of their practice are functions of quality assurance programmes in continuing professional development. In undergraduate education and postgraduate training programmes, students and trainees failing to meet graduate outcomes are identified using examinations and workplace assessments, and quality assurance processes around these ensure that the processes are robust.

The term *quality control* describes an internal system used to determine whether processes and outcomes are reaching a predetermined standard. For example, within an education programme a standard might be that 'all units of teaching and learning have defined learning objectives'. During an annual audit a department might check that all units are meeting this standard. If a unit (say) is determined not to have learning objectives then this will be corrected and re-audit performed.

Quality assurance is something that happens outside the usual cycle of quality control. External bodies usually perform quality assurance. These can be external to the individual course or medical school but still within the university, hospital, or postgraduate college programme. For example, a university might have a quality assurance committee that determines whether the medical school programmes meet wider university standards. The medical school might have an audit process, sometimes called an internal review process, to assure that all parts of the programme are meeting the internal standards. In postgraduate training, training posts are quality assured by the colleges – they have to meet standards for supervision, provision of education, and access to learning resources before trainees can be appointed to the post. The highest level of quality assurance comes from the national medical education accrediting bodies – for example the General Medical Council in the UK and the Australian Medical Council in Australia and New Zealand. Quality assurance at every level is concerned at every level with helping doctors to practise safely.

To summarise:

- *Quality assurance* is the process of measuring attributes of a course or programme against internal or external standards. It provides confidence that the quality of the educational qualifications awarded are meeting the standards and that graduates have the attributes claimed by the programme.
- *Quality improvement* is an on-going process of detection of areas for improvement and implementing the positive changes that result.
- *Quality control* is an internal process that matches internally defined standards against observations.

Quality and Continuing Professional Development

Before moving on, we pause to consider quality assurance as it applies to the third phase of medical education, continuing professional development, as this presents a rather different set of issues. Continuing professional development (CPD) programmes aim to promote lifelong learning based on on-going assessments by self or others, and aim to improve or maintain professional skills in domains such as clinical, leadership, administration, and education. As such, a CPD programme can provide quality assurance of practising individuals.

For a CPD programme to function as a quality assurance exercise for individuals, the following criteria should be met:

- It should measure achievement against predefined outcomes or standards.

- The measurement of achievement must include more than self-assessment – this might include audits of practice, recertification examinations, or portfolios of evidence.
- If the programme just records activities, for example, hours spent at a conference, it is not a quality assurance activity.

CPD, when undertaken within a quality framework, can provide public assurance that doctors remain fit to practise. The UK General Medical Council's revalidation process is an example of continuing professional development occurring within a quality framework. Doctors are measured against a clear set of outcomes and processes are in place for appraisal and remediation.

A CPD programme can also be subject to quality assurance itself. In this case the programme can be regarded like any other course or programme of study and the activities outlined elsewhere in this chapter are applicable. Outcomes to be measured might include impact on the participants' learning, the ability of the programme to show improved practice for most participants, the ability of the programme to detect underperforming practitioners, etc.

Chapter 19 provides a more detailed consideration of CPD and the role of regulators and accrediting bodies.

The Quality Cycle

The processes of assessing quality, comparing data to standards, implementing changes, and reassessing, form the basis of the 'quality cycle' (see Figure 7.1.) The quality cycle is a continuous process and several premises must hold true for it to be useful:

- The first premise is that the data collection is accurate, reflecting the true nature of the teaching, learning, and assessment that is occurring.
- The second premise is that the standards against which the data are judged are appropriate.
- The third premise is that there is ability and will to make the improvements.
- The fourth premise is that organisational systems are in place to monitor the quality process itself.

The quality cycle begins by setting appropriate standards. As discussed earlier, standards are set at the level of the faculty or school or based on the requirements of the external quality assurance requirements. Usually (but not always) these will be based on current best practice and/or published evidence in the literature. Standards can be adapted to the local situation of medical education delivery, taking into account cultural and resourcing needs and priorities. Chapter 5 considers the influence of context on curricula in more depth.

Choosing appropriate measuring tools is important to ensure the data gathered are reliable and valid. Instruments used to measure quality should be based on validated instruments, derive evidence from several sources (sometimes called triangulation), and be of quantitative and qualitative types. It is important to emphasise that evidence does not need to be numerical or quantitative. Text-based, or qualitative, outcomes are also important to use, depending on the standards of interest, and free-text comments in response to open-ended questions can be very informative in interpreting answers to more standardised questions. When measuring quality, also consider which voices need to be heard. This will usually include the voices of the student, the teacher, the administrator, and any external stakeholders, such as patients and employers. *Validity* comes from ensuring the measures relate to the standards and are of value and importance to the people involved. *Reliability* comes from choosing good tools, using a variety of tools, and using a variety of observers or viewpoints.

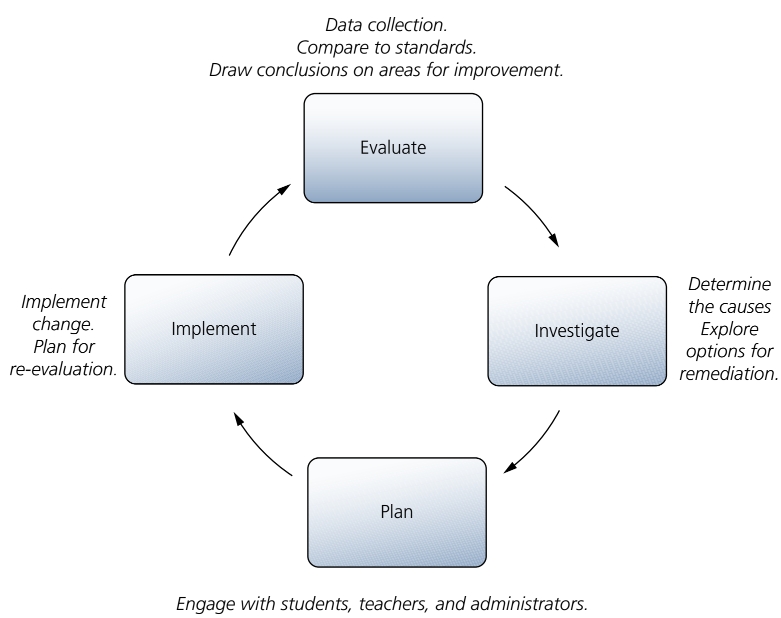


Figure 7.1 A generic quality improvement cycle.

Here is an example: A convenor (programme director) is asked to conduct an annual quality review of her course. She has identified those quality measures her institution requires her to evaluate against, and has included some additional outcomes that are of interest to her and her colleagues. She undertakes a quantitative and qualitative student survey, conducts a discussion at the end of the course with all the tutors answering a predetermined set of questions, and enquires of several convenors in years further along the course to determine whether the students appear to have learnt material that is appropriate for them to build on. The survey and focus group questions include those from the bank containing validated instruments and held by the local education unit. Once the teacher has gathered the data she compares it to the school's standards, noting those that have not been met. There are several groups that have an interest in the data the teacher has gathered. These include the committee or group responsible for quality at the level of the school, the students and teachers who contributed to generating the data, and the external accrediting bodies. During her course evaluation, the teacher will have noted learning activities or practices that are going well. It is valuable to share those with the other teachers on the committee so others can learn from her discoveries. Data collection is considered in more detail further on in this chapter.

Once data are gathered and analysed, areas requiring improvement are noted. Before designing interventions that will improve the course, the reason for the failure to meet any standards must be understood. Reasons may be apparent from the data already gathered but more usually further investigation will be required. Lack of knowledge, lack of time, and/or lack of resources can all generate an observation of the same deficiency, but different strategies to make improvement will be needed. Implementing the changes needed also requires consultation and buy-in from interested and affected parties. These can include teachers, students, and administrators at other learning units running within the programme. Involving the people who have to implement or design the intervention or those who are affected by the intervention will make it more likely that it succeeds. Not only do they have knowledge of on-the-ground conditions but without the explicit energy and conviction of those responsible for implementing a change, any intervention will fail.

Returning to our example, there are several groups that have an interest in the data our convenor has gathered. These include the committee responsible for quality at the level of the school, the students and teachers who contributed to generating the data, and the external accrediting bodies. Before she can report to the interested parties she must determine the underlying reasons why the standards are currently not being met. To do this, she has a discussion with individual teachers and gathers more data about the problem detected. When the teacher feels she has a full understanding she reports back to the teacher group and together they design an intervention that will improve the situation. Once the plan is made, the teacher reports to the school curriculum committee, indicating the deficiency detected and the plans to address it. She takes advice from this wider group of experience and an agreed final plan is made. She also mentions the concerns raised and where action is either not possible or is inappropriate.

Once an improvement plan is formulated and agreed, it should then be reported upwards. The educational organisation with an effective quality cycle will have an ability to centrally monitor that quality cycles are occurring, to review data and plans generated, and to determine whether they are being implemented. This is usually via a nominated individual or committee within the school, university, or postgraduate training provider.

Any quality cycle should be repeated at defined intervals. This is usually done whenever the course is repeated; annually in most medical schools and postgraduate training providers. In the next repetition of the quality cycle any interventions designed to improve the problem must be explicitly investigated.

The cycle above describes the processes needed to improve a programme. There is another aspect to quality assurance – identifying and disseminating good practice. What if the convenor had discovered that her course was going exceptionally well and needed little intervention? She could reflect on why this was so as part of the 'investigate' phase. The 'plan' phase might include a plan on disseminating the success, and 'implement' might be helping others change their courses. An example would be if the teacher had successfully used *backwards design* in planning her course – writing the outcomes, then planning the assessment based on these, with the teaching based on both the assessment and the outcomes. The students commented that the course was well organised, that they felt they could succeed, and they were highly motivated because the assessments examined the objectives, which had been clearly taught. The teacher might report to the supervising committee that backwards design is a key element in the success of her course and suggest that educating other convenors in this process should be a priority. The committee would then consider the current faculty development programme and how this good practice could be implemented in other modules.

An example of the quality cycle operating at both undergraduate and postgraduate level is the UK General Medical Council's quality assurance framework [9]. Evaluation (monitoring) occurs using several tools, including visits, regional review reports, medical school reports, postgraduate dean's reports, national training survey and trainer survey reports, annual specialty reports, and thematic review reports. Reporting to interested parties occurs through online dissemination of reports and publications. Investigation and remediation of problems occurs through enhanced monitoring. Good practice is shared using several methods, including publication of good practice case studies and good practice sharing events bringing together quality leads from medical schools, postgraduate bodies, and medical royal colleges and faculties. Accountability for the GMC quality assurance framework occurs through a Quality Scrutiny Group, which includes members of the public, doctors in postgraduate training, and students.

Data Collection

The quality of a programme has multiple aspects that require evaluation: governance and leadership, admissions (including selection and issues of equality and inclusion), course design (including documentation and alignment with other parts of the programme), course implementation (including resources), course delivery (including staff teaching skills), assessment, staff experience and support, the physical and online learning environment, student experience, and patient safety [10].

The choice of tool depends on the purpose of the measurement. If engaging in quality improvement you will choose tools that provide rich data about *why* various aspects of the course are succeeding or failing. If engaging in quality assurance you will choose tools that look at *whether* various aspects of the course are succeeding or failing. In reality most people use a mixed methods approach with both quantitative survey and qualitative data from a variety of sources. As such, this has parallels with the distinction between formative and summative assessment. Quality improvement approaches are similar to formative assessments – designed to help improvements and to guide future development – what can we do better? Quality assurance approaches are similar to summative assessments – designed to make higher-stakes decisions – is the course good enough?

Selecting the Right Tools

As emphasised above, crucial to any quality exercise is the selection of valid instruments and a range of methods. Just asking students what they think will not result in a sufficiently valid evaluation. The tools you use must be designed to look at specific aspects of the course, and must be aligned to the areas of interest. They must be multifaceted and of several types. One way to do this is by formulating key questions, another is by using pre-developed tools.

Assessment teams visiting educational institutions often use *key questions* or group these as *key lines of enquiry*. Key questions can also be used in internal assessment. For example, a visiting team might ask the following key questions relating to whether the teaching and learning environment includes mechanisms to identify issues in patient safety.

- Could you describe how students and doctors in training raise concerns?
- How are trainees encouraged to raise concerns?
- Could you demonstrate how your current system on raising concerns works?

There are many tools available 'off the shelf' to measure educational quality [11]. Your institution may have tools that it prefers to use, and talking to your local education unit will help you find out about these. Always decide what you're trying to measure before you choose the tool. Also, consider how 'acceptable' the instrument is likely to be; for instance, questionnaires containing more than 30 items may be unacceptable to students and using subsections of some questionnaires may be preferable.

Box 7.3 provides examples of a range of evaluation tools and the aspects of medical education and training they are designed to measure.

BOX 7.3 HOW TO: Evaluate the quality of medical education

To evaluate the learning environment

- The Dundee Ready Educational Environment Measure (DREEM) measures students' perception of learning, teachers, academic self-perceptions, atmosphere, and social self-perceptions [12, 13].

To evaluate clinical teaching

- The Medical Instructional Quality in ambulatory care (MedIQ) tool measures preceptor activities, environmental interactions, learning opportunities, and learner involvement in patient care [14].
- Student Evaluation of Teaching in Outpatient Clinics (SETOC) instrument measures establishing the learning environment, clinical teaching, 'general teaching', clinical competence, and global rating [15].
- The Stanford Faculty Development Programme Clinical Teaching Framework measures establishing a positive learning climate, control of teaching session, communicating goals, promoting understanding and retention, evaluation, feedback, and promoting self-directed learning [16].
- The Maastricht Clinical Teaching Questionnaire (MCTQ) measures modelling, coaching, exploration, articulation, and safe learning environment [17].

To evaluate individual teachers in preclinical teaching

- Student evaluation of teaching remains a cornerstone, but should be supplemented by other methods, especially peer or professional observations of teaching. As student evaluations of teaching are used for academic progression as well as quality assurance, your in-house tools should be used; in fact they are probably mandated by your employing university. They are usually synthesised within a teaching portfolio that incorporates reflection and response to evaluations.
- If an outside evaluation is used, one of the better validated and most widely used is the Student Evaluation of Educational Quality Questionnaire (SEEQ) which measures nine distinct components of teaching effectiveness: learning/value, enthusiasm, organisation, group interaction, individual rapport, breadth of coverage, examinations/grading, assignments, and workload/difficulty [18].

To evaluate courses

- The UNSW Medicine Student Experience Questionnaire (MedSEQ) measures learning, teaching, and assessment; organisation and student understanding of the programme; community interaction and value; student support; and resources [19].
- The Medical Course Experience Questionnaire (MCEQ) uses 18 questions clustered into four factors: clinical practice, becoming a professional, influences on health delivery, professional support [20].

Using Qualitative Data

Qualitative data provides insight into the underlying reasons for the effects observed from quantitative evaluations of courses. Data can be generated in several ways, such as comments from written and online evaluations, focus groups, interviews, and notes from semi-formal encounters such as staff/student committees.

Focus groups should be run by facilitators who can elicit negative comments, not just positive, so are usually better not undertaken by teachers of the course in question but by people at arm's length who are skilled at the method. They are useful for gathering opinions, beliefs, and attitudes about issues identified in a qualitative evaluation, testing possible solutions to a problem, and encouraging discussion about a particular topic. It takes time to plan a focus group. Start several weeks in advance, recruit a facilitator, plan for between six and twelve participants, and develop your questions [21].

Analysis of qualitative data produced by any method can be complex. A typical analysis follows the following pattern. Raw data – the statements as respondents said them – are recorded in summary form or transcribed. The data are then ordered or categorised by natural levels or themes. Descriptive statements are generated that summarise respondents' comments and provide illustrative examples using the raw data. Quotes are used to illustrate points. The final interpretation builds on the descriptive process relating the themes to underlying educational methods and theory. Qualitative methods and the analysis of qualitative data are described in detail in Chapter 29. Overall the goal is to gain deeper understanding of a known problem or to explore possible solutions, rather than just to determine if there is a problem in the first place.

Here is another example, this time taken from postgraduate medical training.

A postgraduate training programme director analyses the end-of-year trainee survey. Tutorials are getting average grading of 2.5/5 and written comments at the end of the evaluation say 'tutorials are boring', 'there is too much detail', and 'why am I even going when it is just irrelevant'. She decides to run a focus group to find out more about what is happening in tutorials and to test possible solutions. She contacts the local education unit and recruits a facilitator. Together they develop questions exploring what trainees think about the tutorials, what they would like to gain from tutorials that are run, and what the objectives of tutorials should be. Participants are recruited, and the focus group is recorded and transcribed. Analysis reveals that some tutorials are liked by the trainees ('Not all of them are bad, just the ones run by X') and learners wanted 'to see how the stuff we are learning is useful'. Analysis reveals that some tutors think students need to know more than is currently in the curriculum and so run a lecture. This clashes with the tutorial objective of using active learning techniques 'to learn how to apply knowledge in medical practice'. The supervisor realises that tutorials need to include more opportunities for active learning, application of knowledge, problem solving, and participation. She discusses with her education unit whether a faculty development course could be made available to all tutors on the course, and an expectation created that tutors would attend.

Pitfalls and Confounding Factors

Low response rates to surveys intended to evaluate quality are a widespread problem and have the possibility of introducing bias, especially if the motivated respondents lie at the extremes of experiences. Rates can be increased by timetabling contact in order to fill-in an evaluation questionnaire either online or on paper. Brevity also encourages completion, as does decreasing frequency of evaluation [22].

Teachers, trainees, and students are more motivated to take time to participate in quality improvement surveys if they are reassured that some of it might be acted upon. Explaining to participants why you are asking them to provide evaluations and what you intend to do with the information can increase response rates and the quality of the responses.

Dissatisfied trainees and students may give a higher response rate. For this reason, student surveys should always be triangulated by using other tools such as peer evaluation or patient surveys. The same effect can apply in focus groups that include one or two disaffected and vocal members.

Some unexpected effects occur – such as when using a scale, more positive results will occur if the positive anchor is placed on the left, and students, when uncertain, choose the second highest option [23]. The same study also showed that students do not always understand educational concepts, for example they seemed particularly confused about what 'feedback' means, interpreting feedback only occurred when given personally.

More understandably, anonymous evaluations typically yield less favourable ratings [24], and students and trainees are swayed by the charisma of the teacher to the detriment of effective content learning [25]. Science and theory-based teaching typically receive less favourable ratings than clinical teaching, and lectures yield worse evaluations than small-group formats [26].

It is important not to over-interpret student ratings, use only statistically significant results, and note evidence that suggests that student-rated quality of teaching does not correlate well with student achievement. Aspects that are disliked by one group of students may be liked by another – overall trends and aggregated data give a better picture [27–29].

Quality Assurance of Assessment

Assessment can be considered from the perspective of the whole course, the unit, and the individual assessment or test. Regardless of the focus, the same principle applies – assessments should align with the learning objectives and the learning methods. A useful framework when thinking about assessing clinical competence is 'Miller's Pyramid' [30]. If the initial objectives are written with the desired level of performance in mind then the assessments that will test them can be crafted appropriately. Quality assurance in this area is concerned with whether assessments are well-documented, fair, objective, and appropriately test the objectives and outcomes. They should have a positive impact on learning, that is, encourage students to learn the desired aspects of the course that are valued by the teachers.

To determine whether the objectives over a teaching unit, year, or the whole course are appropriately assessed, a blueprint is written. A blueprint documents where and how each outcome is tested. To do this, start by tabulating the curriculum content and weighting each item by importance (the importance is subjective and might be based on the frequency of a condition or the severity). Next, assign each item a descriptor based on Miller's framework – whether the item should be tested as knowledge through to performance. Similarly, code formative and summative assessments, mapping each item tested, the weight, and the descriptor, and compare the two lists. The revealed discrepancies can then be analysed.

Standard setting of assessments can be conducted in several ways and methods are discussed elsewhere (see Chapter 24), but the role of the programme evaluator is to ensure that a valid method is being used and documented. It is important to minimise variation in the standard of assessment occurring at different sites and with varying examiners. There may be understandable variation in how examiners rate students on individual assessments, but it is the consistency of decisions arising from the synthesis of several assessments that is of most salience to students and others.

The concept of 'programmatic assessment' recognises that there will be variation between examiners and assessment situations and allows for assessment decisions to be based on multiple information points. An example of its use is the assessment of a trainee or resident at the end of their postgraduate clinical years. Many items from multiple domains will make up the final decision as to whether a trainee is fit to practise as a consultant or general practitioner. A records system should be in place to accurately document progress through the whole training, one that is objective, fair, and transparent. Quality assurance here should focus as much on the assessment processes and decisions as on the assessment tools.

The ultimate aim in quality assurance of assessment is to ensure that it does what it advertises it will do – that is, accurately determine whether graduates are fit for clinical practice or not. There are key organisations and people interested in this who lie outside the university or training provider – in particular future potential employers and patients. Quality assurance of assessment should include determining whether these stakeholders have been consulted on what competencies the graduating students have mastered and whether the competencies are documented in a form that allows employers, students, and the university to transparently see graduating students are fit for practice [31].

Acting on Results and Implementing Change

Training supervisors and educational leaders are in a classic middle management position. They must influence the practice of their colleagues around them and have a supervisory responsibility towards tutors, yet access to resources needed to implement any desired changes lies with their head of department, dean, or health unit manager. Negotiating these relationships requires skill, tact, and persuasion. The key skill required is communication; this must start as early as possible when any need for change is identified by a quality assessment. The first consultation should be with colleagues, tutors, and head of department. Don't present them with solutions; ask them for their thoughts on what their solutions might be. Develop an intervention together. Identify committees that need to be involved and need to approve any changes that will occur. Work out feasibility, resources needed, and barriers that might be encountered. Communicate the problem and the plan to all those that will be affected. Run pilot programmes as needed. Evaluate as you go, making modifications [32]. Box 7.4 provides a worked example.

What if it doesn't work? Change is most commonly derailed when the culture of the department fails to actively support change and innovation or permits unprofessional or obstructive behaviour. The head of department ultimately sets the culture in a department, and as in any management situation if your boss is a problem then you will need to seek help outside the hierarchy. Fortunately, in undergraduate medical education this is usually possible by turning to the local medical education unit for advice. More on the leadership and management of educational change can be found in Chapter 37.

External Accreditation and Benchmarking

External accreditation agencies are keen to ensure appropriate evaluation processes are in place and acted upon but they are more interested in the big picture rather than small changes, trusting that if the processes are in place, and followed, then areas requiring remediation will be identified. For instance, the Australian Medical Council, which accredits primary medical providers in Australasia, defines evaluation as 'the set of policies and structured processes by which a medical education provider regularly assesses and determines the extent to which its training and education functions are achieving their outcomes' [33].

Medical education regulators and accrediting bodies are also increasingly concerned with the quality and safety of patient care. This outward focus on the patient in medical education and training is emphasised, for example, by the UK GMC in relation to the learning environment: 'We will [now] make sure that education and training takes place where patients are safe, the care and experience of patients is good, and education and training are valued.' 'In non-clinical learning environments, there should also be a culture of promoting patient safety' [34].

The external accreditor asks for evaluation to occur from multiple points of view – students, teachers, graduates, and external stakeholders. Quality assurance activities to meet these standards must look wider than the medical school and consult the employers and communities into which medical students will graduate. A range of stakeholders from hospital chief executives and professional postgraduate

BOX 7.4 HOW TO: Implement change: An example

Through quality assurance activity, an undergraduate tutor has discovered changes that need to be made to tutorials. Students would like them to be an active learning experience, and not 'mini-lectures'. What should she do now? Her first thought was to talk to the head of department and present her with a solution – a faculty development course. But is this the right way to go about it? She uses a change management sequence to make a plan.

<u>Identify a shared problem.</u>
Students have identified that some tutorials are didactic and too detailed. They have asked for active learning in context.

<u>Consult with stakeholders.</u>
She meets with colleagues and discusses the student responses. They agree together on the nature of the problem and discuss the potential educational benefits of moving to active learning in context.

<u>Develop an intervention together.</u>
Several ideas are explored – more formally structured tutorials, faculty development, observed practice. Through identifying the solution themselves, her colleagues are more likely to implement it.

<u>Leverage formal committees and processes.</u>
The institutional education committee have heard through the staff/student forum that students are complaining about tutorials. It is important they are updated on changes that are being planned.

<u>Consider feasibility, barriers, and resources.</u>
Colleagues agree to more structured tutorials. These need to be written and the tutor negotiates with her head of department for time to do this. In order to make sure they meet best practice, a senior colleague is asked to review them.

<u>Communicate the plan and the solution to all affected.</u>
The new tutorials will incorporate clinical material. Some of this might overlap with other modules, and may involve real patient data. Clinical heads of department are consulted and permission obtained to use anonymous material. The solution is communicated to students and their feedback obtained.

<u>Run pilot programmes.</u>
The new initiative is piloted in two tutorials.

<u>Re-evaluate and make modifications.</u>
The tutor convenes a focus group to evaluate the new tutorials. She makes further modifications as recommended. Reports are provided back to colleagues, tutors, head of department, curriculum committee, and students.

colleges to patient advocacy groups should all be consulted as part of an external evaluation.

One of the purposes of external accreditation is to share best practice and to help bring everyone up to the level of the good. This doesn't have to wait for accreditation though and a medical school or postgraduate provider should avail itself of opportunities to compare itself to peers. Benchmarking programmes at the undergraduate level are usually based on sharing examination materials then collating data to compare student achievement across medical schools; they include sharing preclinical MCQs (AMSAC), sharing OSCE stations (ACCLAiM), and sharing a large assessment bank for medical education on an international scale (IDEAL) [35]. Benchmarking is a tool, and as such is able to be used usefully – to identify areas for improvement – but can be misused, for example to build league tables. Benchmarking is one tool of many in evaluating a programme but does not readily support the evaluation of aspects more suited to qualitative exploration [36].

Conclusion

Quality assurance and continuous quality improvement are daunting tasks but the pursuit of excellence should be seen as a journey, not a destination. Like any journey, the important thing is to begin, have a map, and take one step at a time. It is rare for a course convenor or programme director to conclude that their course or programme is good and requires no further improvement. More often the course will be good in parts. The aim of a quality assurance process is to identify those areas that need improvement and affirm those areas that are working as intended. Overall, quality processes should be targeted at the things that really matter to learners and, more importantly, to patients.

References

1 World Federation for Medical Education (2015). Standards. http://wfme.org/standards (accessed 15 August 2017).
2 Accreditation Council for Graduate Medical Education (2017). Common Programme Requirements. http://www.acgme.org/What-We-Do/Accreditation/Common-Program-Requirements (accessed 15 August 2017).
3 Academy of Medical Educators (2014). Professional Standards: for medical, dental and veterinary educators. http://www.medicaleducators.org/Professional-Standards (accessed 15 August 2017).
4 Liaison Committee on Medical Education (2016). Functions and Structure of a Medical School: Standards for Accreditation of Medical Education Programs Leading to the MD Degree. http://lcme.org/publications (accessed 15 August 2017).
5 Global Consensus for Social Accountability of Medical Schools (2011). www.healthsocialaccountability.org (accessed 15 August 2017).
6 Association for Medical Education in Europe (2017). ASPIRE Programme. https://www.aspire-to-excellence.org (accessed 15 August 2017).
7 General Medical Council (2017). Excellence by Design: Standards for Postgraduate Curricula. http://www.gmc-uk.org/education/postgraduate/excellence_by_design.asp (accessed 15 August 2017).
8 Jamieson, J., Snadden, D., Dobson, S. et al. (2011). Health Disparities, Social Accountability and Postgraduate Medical Education. A paper commissioned as part of the Environmental Scan for the Future of Medical Education in Canada Postgraduate Project. https://www.afmc.ca/pdf/fmec/01_Jamieson_Health%20Disparity.pdf (accessed 27 June 2018).

9 General Medical Council (2017). Quality Assurance Framework. http://www.gmc-uk.org/education/qaf.asp (accessed 15 August 2017).
10 Gibson, K.A., Boyle, P.M., Black, D.A. et al. (2008). Enhancing evaluation in an undergraduate medical education program. *Academic Medicine* 83 (8): 787–793.
11 Schiekirka, S., Feufel, M.A., Herrmann-Lingen, C., and Raupach, T. (2015). Evaluation in medical education: a topical review of target parameters, data collection tools and confounding factors. *German Medical Science* 13: Doc15.
12 Roff, S. (2005). The Dundee Ready Educational Environment Measure (DREEM) – a generic instrument for measuring students' perceptions of undergraduate health professions curricula. *Medical Teacher* 27 (4): 322–325.
13 Roff, S., McAleer, S., Harden, R.M. et al. (1997). Development and validation of the Dundee Ready Education Environment Measure (DREEM). *Medical Teacher* 19 (4): 295–299.
14 James, P.A. and Osborne, J.W. (1999). A measure of medical instructional quality in ambulatory settings: the MedIQ. *Family Medicine* 31: 263–269.
15 Zuberi, R.W., Bordage, G., and Norman, G.R. (2007). Validation of the SETOC instrument—student evaluation of teaching in outpatient clinics. *Advances in Health Sciences Education* 12 (1): 55–69.
16 Litzelman, D.K., Stratos, G.A., Marriott, D.J., and Skeff, K.M. (1998). Factorial validation of a widely disseminated educational framework for evaluating clinical teachers. *Academic Medicine.* 73 (6): 688–695.
17 Stalmeijer, R.E., Dolmans, D.H., Wolfhagen, I.H. et al. (2010). The Maastricht Clinical Teaching Questionnaire (MCTQ) as a valid and reliable instrument for the evaluation of clinical teachers. *Academic Medicine* 85 (11): 1732–1738.
18 Marsh, H.W. (1982). SEEQ: a reliable, valid, and useful instrument for collecting students' evaluations of university teaching. *British Journal of Educational Psychology* 52 (1): 77–95.
19 Boyle, P., Grimm M., Scicluna, H., and McNeil, H.P. (2009). The UNSW Medicine Student Experience Questionnaire (MedSEQ): A Synopsis of its Development, Features and Utility. http://handle.unsw.edu.au/1959.4/41547 (accessed 15 August 2017).
20 Fontaine, S., Wilkinson, T.J., and Frampton, C. (2006). The medical course experience questionnaire: development and piloting of questions relevant to evaluation of medical programs [online]. *Focus on Health Professional Education: A Multi-disciplinary Journal* 8 (1): 1–11.
21 Krueger, R.A. and Casey, M.A. (2014). *Focus Groups: A Practical Guide for Applied Research*. London: Sage Publications.
22 Alok, K. (2011). Student evaluation of teaching: an instrument and a development process. *International Journal of Teaching and Learning in Higher Education* 23 (2): 226–235.
23 Billings-Gagliardi, S., Barrett, S.V., and Mazor, K.M. (2004). Interpreting course evaluation results: insights from thinkaloud interviews with medical students. *Medical Education* 38 (10): 1061–1070.
24 Afonso, N.M., Cardozo, L.J., Mascarenhas, O. et al. (2005). Are anonymous evaluations a better assessment of faculty teaching performance? A comparative analysis of open and anonymous evaluation processes. *Family Medicine* 37 (1): 43–47.
25 Naftulin, D.H., Ware, J.E. Jr., and Donnelly, F.A. (1973). The Doctor Fox Lecture: a paradigm of educational seduction. *Academic Medicine* 48 (7): 630–635.
26 Remedios, R. and Lieberman, D.A. (2008). I liked your course because you taught me well: the influence of grades, workload, expectations and goals on students' evaluations of teaching. *British Educational Research Journal* 34 (1): 91–115.
27 Boysen, G.A. (2015). Significant interpretation of small mean differences in student evaluations of teaching despite explicit warning to avoid overinterpretation. *Scholarship of Teaching and Learning in Psychology* 2: 150–162.
28 Surgenor, P. (2013). Obstacles and opportunities: addressing the growing pains of summative student evaluation of teaching. *Assessment and Evaluation in Higher Education* 38 (3): 363–376.
29 Carrell, S.E. and West, J.E. (2010). Does professor quality matter? Evidence from random assignment of students to professors. *Journal of Political Economy* 118 (3): 409–432.
30 Miller, G.E. (1995). The assessment of clinical skills/competence/performance. *Academic Medicine* 65 (9): S63–S67.
31 Schuwirth, L.W.T. and Van der Vleuten, C.P.M. (2011). Programmatic assessment: from assessment of learning to assessment for learning. *Medical Teacher* 33 (6): 478–85.
32 Gale, R. and Grant, J. (1997). AMEE Medical Education Guide No. 10: managing change in a medical context: guidelines for action. *Medical Teacher* 19 (4): 239–249.
33 Australian Medical Council (2012). Standards for Assessment and Accreditation of Primary Medical Programs. www.amc.org.au/files/d0ffcecda9608cf49c66c93a79a4ad549638bea0_original.pdf (accessed 15 August 2017).
34 General Medical Council (2015). Promoting Excellence: Standards for Medical Education and Training. http://www.gmc-uk.org/education/standards.asp (accessed 15 August 2017).
35 O'Mara, D.A., Canny, B.J., Rothnie, I.P. et al. (2015). The Australian medical schools assessment collaboration: benchmarking the preclinical performance of medical students. *The Medical Journal of Australia* 202 (2): 95–98.
36 Wilkinson, T.J., Hudson, J.N., McColl, G.J. et al. (2015). Medical school benchmarking. From tools to programmes. *Medical Teacher* 37 (2): 146–152.

Further Reading

Barzansky, B., Hunt, D., Moineau, G. et al. (2015). Continuous quality improvement in an accreditation system for undergraduate medical education: benefits and challenges. *Medical Teacher* 37 (11): 1032–1038.

Blouin, D. and Tekian, A. (2017). Accreditation of medical education programs: moving from student outcomes to continuous quality improvement measures. *Academic Medicine* 93 (3): 377–383. Jul 25. (Epub ahead of print).

Davis, N.L., Davis, D.A., Johnson, N.M. et al. (2013). Aligning academic continuing medical education with quality improvement: a model for the 21st century. *Academic Medicine* 88 (10): 1437–1441.

Vroeijenstijn, A.I. (1995). Quality assurance in medical education. *Academic Medicine* 70 (7): S59–S67.

Part 2

Teaching and Learning

8 Lectures and Large Groups

Dujeepa D. Samarasekera[1], Matthew C. E. Gwee[1], Andrew Long[2,3,4], and Bridget Lock[5]

[1]Centre for Medical Education, Yong Lo Lin School of Medicine, National University of Singapore, Singapore, Singapore
[2]Royal College of Paediatrics and Child Health, London, UK
[3]University College London, London, UK
[4]Great Ormond Street Hospital, London, UK
[5]South London Health care NHS Trust, London, UK

KEY MESSAGES

- Lectures are learning strategies which are still popular and widely used for the delivery of instruction to students in large groups, in both undergraduate and postgraduate education, despite evidence to suggest that they have little impact on 'deep' learning.
- A well-structured lecture with explicit learning objectives, defined contextual relationships, and linkage of theory to experience is more likely to maximise the learning opportunity for those attending the lecture.
- Interactivity during lectures is an important adjunct to allow questioning and engagement of the learners, which is likely to improve the learning experience.
- A range of different techniques may be employed to engage larger groups in constructive learning processes that lead to better concentration and sustained enjoyment for all participants.
- Teachers themselves must undergo a major role change based on three key educational paradigm shifts, namely: from just informing to involving students in the instructional process; from teacher-centred instruction to student-centred learning; from a master–apprentice type relationship to that of partnership-and-bonding in learning.
- The design of large-group lectures must ensure that transformative learning principles are applied and be expected to contribute to the ultimate institutional goal of transforming today's students to become tomorrow's practitioners.

Introduction

Lectures are learning strategies that are still popular and widely used for the delivery of instruction to medical students, residents, or qualified doctors in large groups, usually in a formal setting such as a large hall or lecture 'theatre'. In undergraduate education, lectures are generally considered as cost efficient, mainly because one teacher can deliver instruction to many students. Consequently, lectures have great appeal in situations of dwindling budgets and/or increasing student numbers in some institutions of higher learning. This chapter aims to provide some educational insights on the use of lectures for the delivery of instruction to large groups. We will begin with a brief discussion of the place of the lecture within medical education, followed by a dissection of the didactic lecture and its impact on learning. Some key concepts in medical education will be highlighted and the implications for those in the re-casting of the didactic lecture as a vehicle for large-group teaching, with a focus on undergraduate medical education. The expected outcomes that can be derived from such teaching will also be reviewed. We conclude with some specific strategies that may be deployed in the large-group or lecture setting.

The Lecture in Medical Education

Traditional lectures, during which information is imparted to a passively receptive audience, have been the mainstay of undergraduate and postgraduate education for centuries, but there is an increasing body of evidence questioning the place of this style of learning in medical education. Only a small percentage of the information delivered during a lecture is retained, and often what is told does not equate to what is learned [1]. Despite this widely recognised fact, most undergraduate courses continue to have a significant lecture component within their curricula and practising clinicians continue to choose to attend lecture-based meetings at the local, national, and international level as part of their professional development. This suggests that even in the information-rich twenty-first century lectures still have a place [2]. Like medicine itself, professional learning is a social activity. The challenge is to

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

reconcile the inherent educational problems of the lecture format with the expressed need of the professional community for this activity and to use valuable professional time productively.

Over many years, the predominant mode used for the delivery of instruction to students was the didactic lecture. The primary role of the teacher then, as the discipline expert, was that of an information provider, delivering, mainly factual, course content to students. Why did medical teachers assume such a role? In the past, textbooks were relatively expensive and, therefore, not readily available to students and the internet did not exist. Teachers of that time, as discipline experts, had the unenviable and challenging task of 'informing' students about advances made in their respective disciplines. Consequently, teachers regularly delivered expert, factual content knowledge – 'pearls of wisdom' – for student learning. Soon, teachers gained the reputation as the source or fountain of knowledge, the 'sage on the stage' [3–7].

From an educational viewpoint, the traditional lecture can be viewed as a monologue involving the unilateral transmission of mainly factual discipline knowledge by teachers. The student role was simply to sit, listen, and take notes and then memorise, recall facts, and regurgitate these in exams. Unfortunately, this is usually followed by knowledge fade, often a fairly rapid decline. Such shortcomings suggest that the didactic lecture is inappropriate and inadequate for health professional education in the twenty-first century [8].

Although the didactic lecture is an efficient instructional strategy – one teacher can deliver instruction to many students – its key design features (highly teacher-centred, and discipline-specific) reveal several weaknesses. Such a learning environment leads to students becoming passive learners who are likely to engage in fact memorisation, i.e. students are likely to undertake rote (superficial) learning. As rote-learners, students will lack the ability to organise or connect related facts and ideas into meaningful conceptual frameworks (schemas), although it is schemas which are expected to facilitate knowledge and subsequent recall or retrieval [9]. Students also become highly dependent on teachers for their learning needs. Finally, the knowledge fade which occurs soon after exams are over often provides the basis of complaints from clinical teachers that students do not seem to remember or understand what they learned previously. The highly discipline-specific design is also likely to encourage discipline-specific, 'compartmentalised', learning (learning in silos), at the expense of learning across disciplines (integrated learning).

Although less obvious to many, the design of the didactic lecture also imposes a kind of 'master–apprentice' relationship that reveals 'the instructor's [teacher's] power over the student': the teacher, as the discipline expert, is the 'master' whereas students, as the 'novice learners', are the 'apprentices'. The expert teaches the novice learners who are often assumed to be devoid of any knowledge! This somewhat implies that the master–apprentice type of relationship presumably boosts a teacher's sense of having power over students [10]. More positively, the lecture can also be argued to be part of the process of professional enculturation (see Chapter 17) in which entrants to the community have to learn the tradition and knowledge of the profession for its continuity. Working within a professional community of practice and engaging in conversation about the work of the community allows for the participation of newcomers [11]. Through these 'professional' conversations, the student learns to 'walk-the-walk' and 'talk-the-talk' of professional practice [12]. Later on in a career, social specialist subgroups are bound together through their professional duty to engage in continuing professional development. This includes both learning and teaching of specialist professional knowledge in its widest sense, and assessment of the value of new information for patients through questioning of, and deliberation on, emerging research knowledge within a 'safe' professional grouping. Continuing professional development is discussed in more detail in Chapter 19 but for the remainder of this chapter we shall focus on the lecture in the context of undergraduate medical education.

Lectures and Learning

'… the burden of information that is imposed taxes the memory, but not the intellect' (GMC [13]).

Key to understanding the shortcomings of the traditional lecture, and how presenters have sought to mitigate these, are the emerging insights of cognitive neuroscience (see Chapter 3). Theories of how memory is laid down and knowledge retrieved for later use suggest how a presenter might support his or her audience in their learning.

As words are heard and understood they are assimilated into working or short-term memory through the recognition of patterns and context. Without context, the information is forgotten if not learnt through immediate repetition. To secure information in long-term memory, filing or 'coding' must occur [14]. Two forms of long-term memory have been identified; *procedural* or non-declarative (implicit) memory is related to learning 'unconscious' skills (e.g. riding a bicycle), whereas *declarative* (or explicit) memory refers to that which can be consciously recalled (e.g. facts or knowledge). Declarative memory is thought to have two inter-dependent components; an episodic memory, which stores personal experiences and events, and a semantic memory, which stores factual knowledge that does not rely on personal experience [15]. Episodic memory makes multiple associations or links with events, sequences, sounds, smells, or sights over time. Recall is dependent on the importance of the event and on developed connections to other pieces of retained knowledge in its many forms (see Box 8.1). Something heard or seen may 'strike a chord' in relation to a patient seen previously or an event experienced. Coding episodes into long-term memory is highly personal, dependent on the attention paid, and dependent on the value ascribed to the episode by the individual, as well as how it connects with other previously retained knowledge [16]. Semantic memory refers to the development of concept-based knowledge built from

BOX 8.1 FOCUS ON: Attention and recall

During the course of an hour-long lecture:

- attention is at its height during the first 10–20 minutes and the final 5 minutes [17]
- the most note-taking occurs during the first 10 minutes [18]
- only 42% of the key points of a lecture can be recalled immediately afterwards
- this drops to 20% within one week.

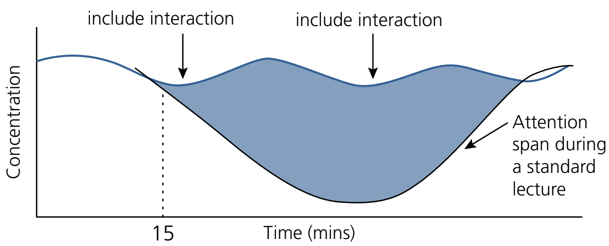


Figure 8.1 Audience concentration during standard lecture with and without interaction. *Source*: Higher Education Academy Engineering Subject Centre [21].

BOX 8.2 WHERE'S THE EVIDENCE: For lectures

To examine the effect of lecturing on course performance, Freeman et al. performed the largest ever, and most comprehensive, meta-analysis of undergraduate science, technology, engineering, and mathematics (STEM) programmes to date. 225 studies were included in the analysis, which looked at data on examination scores or failure rates when comparing student performance in courses that used traditional lecturing versus active learning [24].

On average, student performance on examinations and concept inventories increased by 0.47 SDs with active learning (n = 158 studies), and the odds ratio for failing was 1.95 with traditional lecturing (n = 67 studies). The results indicate that examination scores improved by an average of 6% in courses using active learning and that students in classes with traditional lecturing were 1.5 times more likely to fail than were students in classes with active learning. The results raise questions about the continued use of traditional lecturing, with campaigners even accusing their continuance as being 'unethical' [25].

individual components of factual information which give meaning in the context of previous experience. This interpretation takes time, and, without adequate time to construct links and formulate code, information will not be assimilated into long-term memory and is discarded from short-term memory. The relationship between learning, memory, and professional performance is discussed in more detail in Chapter 3.

Another important issue in relation to lectures is the recognition that there is both a *rate limit* to the system – too fast a delivery and the mechanism of retention is overwhelmed – and a *capacity* to the system – overload of information will lead to a lack of capacity for retention. The capacity of the system relies on a full engagement of concentration generated by motivation for the learning task in hand. Some tactics can be adopted to enhance memory retention – for example, providing the description of an event, setting memorable context, or assisting in the development of relevant connections or anchors [19]. Lecturers are therefore usually advised to provide an opportunity for questioning at reasonable time intervals, enabling listeners to check their construction of knowledge and reveal and correct misunderstandings. As little is recalled after 20 minutes of uninterrupted concentration [20], introducing variety in presentation every 15–20 minutes will aid learning (Figure 8.1 and Box 8.2). A listener needs an opportunity to participate or question to avoid misinterpretation and thus incorrect integration into existing knowledge. The marked decline in attention after 15 minutes might offer the opportunity for a brief anecdote or real-world example illustrating the principles described by the lecture. This will act to embed the delivered ideas in the narrative component of the listener's long-term memory, complementing information previously stored in conceptual or semantic memory [22].

Of even greater concern, Bligh describes processes that may result in memory loss: those of retroactive and proactive interference [23]. The former relates to when there is a requirement to learn a new series of facts shortly after committing the first set to short-term memory, such as in successive lectures. The latter occurs where memorising the first set of facts interferes with remembering the second set. It is also relevant to note that there is a beneficial effect in repetition, which has the opposite effect of interference and tends to consolidate learning, although if learning has not taken place the first time (i.e. the listener did not understand), then repetition has no additional benefit.

In summary, useful learning occurs when the learner makes connections with pre-existing knowledge and can anchor new knowledge to previous concepts [19]. A context for information is also needed for both understanding and recall. Each listener will have personal levels of depth and breadth of understanding related to past experience and motivation for learning. What is learnt depends on knowledge of the range of the subject; old information may become obsolete or be viewed with different importance, and new information develops and finds a place. But this will only be of use in patient care if the learner has the internal connections to draw out the information when needed and put it into practice.

Disruptive Change in Medical Education

'Twentieth century strategies are unfit to tackle twenty-first century problems. … Professional education has not kept pace with these challenges [the disruptive forces], largely because of fragmented, outdated, and static curricula that produce ill-equipped graduates' (Frenk et al. [8]).

Disruptive forces of change now impact strongly on health care practice and, consequently, on health professional education [8]. Major advances in medical knowledge, sciences, and technology, as well as advances in the learning sciences and information technology (IT) is one such disruptive force. Another disruptive force is rapidly changing demographics, with progressively increasing numbers of elderly patients afflicted with (usually several) long-term conditions. These demographic shifts are coupled with changing patient characteristics; patients are better informed about health matters, more involved in their own health care, and have a greater expectation from a health care delivery system.

There is now strong global consensus that current health professional education is no longer adequate, nor appropriate, for the educational preparation of today's students to become tomorrow's practitioners. Grave concern has been expressed that the education of health professional students in the twentieth century has resulted in '… ill-equipped graduates …' (end products) who are 'unfit' to take on the role of twenty-first century health care practitioners [8]. Consequently, major reforms in health professional education have been strongly advocated, with the global acceptance of the need for outcomes-based education for health professional students [26]. In an outcomes-based approach the design of instructional strategies must be closely aligned to the learning outcomes intended as the requisite end-product capability.

An important aspect to be considered in the major reform of health professional education is the significant role change which teachers themselves must undergo, as is discussed in the sections that follow.

From Informing to Involving Students

Firstly, teachers must undergo a significant paradigm shift from informing to involving students in the instructional process, a change that can be greatly facilitated today by technology. Exploiting technology to actively engage students in large groups is an important strategy teachers must use in contemporary medical education. This has enabled students to learn and study across time and space in innovative and interactive ways [8, 27].

Information technology (IT) has progressed to the point that powerful Internet search engines now make knowledge readily available. Today, at the click of a button, information galore will appear before our very eyes within seconds, including: abundant text, sophisticated images, and streaming of high quality, real-life, and well-illustrated educational videos. Such an array of 'information' can be retrieved by students practically anytime and anywhere around the globe. Most of today's students possess their own personal digital assistants (PDAs or e-devices), including smart phones, iPads, tablets, etc., all of which can easily access the Internet. Many of the students of today will be quite comfortable and efficient at searching the Internet for relevant and reliable information (knowledge) required for their education. In fact, The Lancet Global Independent Commission Report strongly emphasised that 'IT-empowered learning is already a reality for the younger generation in most countries …' [8]. Teachers should not hesitate to exploit this situation in the design of new instructional strategies. In view of such progress already made in IT, the role of teachers has changed to provide guidance or academic exercises involving the process of independent information search, as well as the evaluation, integration, analysis of and meaningful or logical conclusion to the information obtained during the search.

From Teacher-centred Instruction to Student-centred Learning

The notion of student-centred learning was already considered by Payne in 1883 when he observed the interaction between teacher and students (the teacher–student relationship) in various teaching situations, and then concluded that learning '… can be performed by no one but the learner, … he is in fact his own teacher, and … learning is self-teaching' [28]. This was subsequently developed further by Barr and Tagg [29]. 'To build the colleges we need for the 21st century … we must consciously reject the instructional paradigm and restructure what we do [as teachers] on the basis of the learning paradigm' [29].

The need for the paradigm shift from teacher-centred instruction to student-centred learning implies that instructional (learning) strategies should now be based on the intended outcomes of student learning, rather than just on the delivery of instruction by the teacher on the assumption that once the instruction is delivered student learning will occur. The paradigm shift should, therefore, encourage students to participate actively in the instructional (learning) process with the teacher functioning as a facilitator (or guide-by-the-side) through questioning of and responding by students.

The main reason for the active participation of students in an instructional process is to ensure that students will not only acquire a broad knowledge base (i.e. foundational knowledge) from various disciplines, but also undergo the development of intellectual skills, such as critical thinking and the power of reasoning, which can be expected to facilitate future medical problem solving and decision-making. In order to achieve this, teachers of today must design instructional strategies that also guide, or facilitate, student learning that involves knowledge processing and knowledge application. In other words, in twenty-first century education of health professional students, teachers need to design instructional strategies (including large-group lectures) that involve students in analysing, integrating, evaluating, and applying knowledge to resolve medical problems or issues. The outcome of student learning in the twenty-first century must therefore ensure not only what students learn, but also how they learn.

Involving students actively in the instructional process will also facilitate the acquisition of higher-order learning outcomes in the cognitive domain, whereas fact memorisation (achieved mainly through rote-learning), results largely in the acquisition of lower-order learning outcomes.

When students are actively engaged in the instructional process, teachers may also be creating opportunities through guidance and facilitation for students in large lecture classes to socially co-construct their own meaning and understanding of what needs to be learned. Meyers and Jones in 1993 elaborated this concept of developing one's mental schemas and have pointed out that 'Students learn not by just absorbing content (taking copious notes and studying for exams), but by critically analysing, discussing, and using content in meaningful ways' [30].

Context-based or situational-based learning in which knowledge is to be used in their future practice (contextual learning) is also an important value-added component in instructional design as well as delivery. According to Cole and Wilson [31]: 'Perhaps the most important feature of contextual learning is the establishment of an appropriate context in which learning can take place' [31]. Moreover, in the Editorials section of the *New England Journal of Medicine*, McMahon and Drazen [32] also drew attention to the fact '… that clinicians learn best when [solving] problems that mirror real-world situations …' [32].

Promoting collaboration with fellow students whenever the opportunity to learn together arises, i.e. learning with-from each other in an 'All teach-All learn' mode and developing their self-directed learning activities, such as completing assigned homework before attending class the following day as practised in a flipped classroom situation, can be highlighted as other value-added attributes for active learning among students.

What then should be the main role of the teacher in the delivery of instruction to students in twenty-first century health professional education? Firstly, teachers need to take on the role of a guide or facilitator of the learning process (rather than that of information provider) in any instruction delivered to students; other related roles which the teacher will find useful to undertake include that of designer and planner of the learning strategy (especially in large-group lectures), as well as that of manager of the learning environment '… to expedite the intellectual and interpersonal process' [33].

The main role change for teachers in twenty-first century education will enable students to be actively involved (engaged) in the instructional process; such a role change also demands greater accountability from teachers. Thus, the role change can also be expected to create ample opportunities for teachers to deliver high-quality instruction through the application of educational scholarship [34]. A scholarly approach to teaching and learning should also enhance the status of teachers themselves to that of educational scholars, rather than that of teachers just performing the same 'boring' routine of delivering factual content knowledge to students. The enhanced status of teachers should provide the impetus for teaching with scholarship in mind [35].

Implications for Lectures and their Redesign

However, as discussed earlier, the major shortcomings of lectures and, therefore, the limitations which such shortcomings impose on the outcomes of student learning, have rendered the design of the 'traditional' didactic lecture rather obsolete and no longer appropriate, nor adequate, for the educational preparation of today's health professional students.

In contemporary health professional education, students undergo a period of learning during which they acquire relevant knowledge, skills, and attitudes from a diversity of disciplines which will equip them with professional competencies that will transform the students into the desired twenty-first century health care practitioners. Thus, the learning process which students (learners) undergo is expected to progressively, transform (shape) their heads, hands, and hearts to think, act, and behave as the new generation of health care practitioners. However, it should also be borne in mind that in the education and training of health professional students, the process of professional socialisation progressively leads to the formation of professional identity that facilitates the transformation of students to health care practitioners. It is through transformative learning (and, of course, the process of professional socialisation as well) that students progressively develop the capability to deliver twenty-first century health care that can match the needs, demands, and challenges of patients and the community. All instructional strategies must, therefore, contribute to the intended institutional goal of progressively transforming (shaping) the way students think, act, and behave as future health care practitioners. This has already been emphasised by Cooke et al. [36] in their statement: 'The future demands approaches to shaping the minds, hands and hearts of physicians. Fundamental change in medical education will require new curricula, new pedagogies, and new forms of assessment' [36].

Today, the design of lectures used for the delivery of instruction to students in large groups must apply the principles of transformative learning. In particular, the key design features of such large-group lectures must ensure that students will be able not only to acquire a pool of broad-based foundational knowledge garnered from a diversity of disciplines, but also develop intellectual skills to enhance students' critical thinking and power of reasoning for future medical problem solving and decision-making.

The significant role change which teachers must undertake in health professional education, from informing to involving students in the teaching and learning process and from teacher-directed instruction to student-centred learning, will fit well with the role of teachers who apply transformative learning principles to deliver instruction to students.

However, an additional and important role change that would enhance the ability of teachers to apply the principles of transformative learning should also be considered: this relates to the mindset change required in the teacher/student relationship; i.e. from the 'traditional' master–apprentice type of relationship to partnership-and-bonding in learning, a practice referred to by Christensen [37] as discussion pedagogy. Christensen [37], in his wisdom, has clearly expressed the sentiment that: 'In discussion teaching, partnership – a collegial sharing of power, accountability, and tasks – supplants hierarchy and asymmetry in the teacher-student relationship. The discussion process itself requires students to become profoundly and actively involved in their own learning, to discover for themselves rather than accept verbal or written pronouncements. They must explore the intellectual terrain without maps, step by step, blazing trails, struggling past obstacles, dealing with disappointments' [37].

Christensen [37] also drew attention to the fact that: 'Such creative activity cannot be ordered or imposed upon the unwilling. Teachers can police attendance and monitor the memorization of theory and fact by tests. But we cannot order our students to be committed to learning and willing to risk experimentation, error, and the uncertainty of exploration. Such attitudes are gifts from one partner to another' [37].

Large-group Teaching for Transformative Learning

Several design strategies are available for use in the delivery of lectures to students in large groups.

A key learning principle that must be applied is, of course, the involvement of students actively in the instructional process. How then can teachers ensure this? Often this is achieved through 'questioning – listening – responding' between the teacher and students or, where applicable, through interaction among students themselves (see Figure 8.2), depending on the design of the instruction; the latter is more commonly used in small group settings.

A clear understanding of how people learn [38, 39] will help teachers design instructional strategies that will optimise and facilitate student learning in an interactive learning environment aimed at having a positive impact on the outcomes of health care delivery. In this context then, it is advantageous to apply and incorporate the following key learning principles into the design of various transformative instructional strategies, including large-group lectures [40]:

- Activate students' recall of prior knowledge; this will facilitate learning in terms of building new on existing knowledge.
- Help students organise related facts and ideas into conceptual frameworks (meaningful schemas) that will facilitate knowledge retention and retrieval, as well as foster deep learning with understanding.

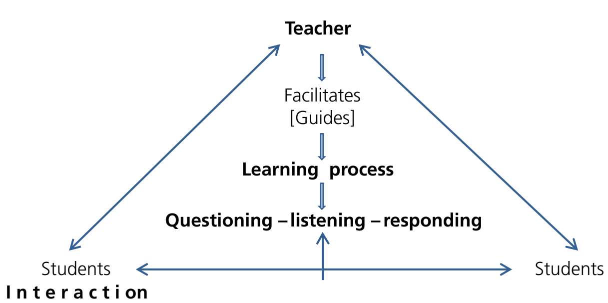


Figure 8.2 Active involvement of students in a large group interactive lecture.

- Encourage students to develop metacognition in their learning, i.e. help students develop their ability to identify their own strengths and weaknesses in learning and to take the necessary remedial action for future improvement.

Morton, in his book chapter titled 'Lecturing to Large Groups' in 2003 shares that any presentation must be informative, interesting, and engaging. The presentation can be developed in four stages [41]. The first stage is the pre-planning stage. Before starting on the actual content of the presentation, one needs to take note of the curricular blueprint and see where this activity is placed in the overall teaching and learning programme for students. If you are not familiar with the student group, it is important to know some of their attributes such as prior level of learning, learning preferences, and group dynamics as well as the class size. The time allocated to the activity and the time of the day, i.e. early morning or late in the afternoon, are also important considerations. These will assist you to develop the large-group teaching activity by aligning the content areas that could be covered during the time allocated and plan the short activities.

At the beginning of an activity/lecture, it is best to capture the attention of students. This can be done by using a short video depicting a relevant real-life event which will serve to stimulate further interest in the topic. In health professions, some teachers start the presentation with a patient narrative or even invite a patient to relate their experience at the beginning of the lecture. This catches the attention of the students since they will be anticipating what is to follow. For example, in delivering a lecture on 'Anticholinesterase Agents', it will excite students greatly and capture their attention by presenting, as the first PowerPoint slide, a picture appearing on the cover page of Time magazine (April 3, no. 13, 1995) [42] which reported on the deadly nerve gas attack in the Tokyo subway or the one-minute BBC news clip. Once you capture students' attention then outline the purpose and the key areas that would be covered during the large-group activity.

The next step involves the main body of the large-group activity. Here the teacher needs to develop the argument/s or explain the key content with real-world examples from practice or use of clinical cases. Repeating important areas or content is necessary to clarify difficult concepts or constructs. This will assist students to organise related facts and ideas into conceptual frameworks or meaningful schemas that will facilitate knowledge retention and retrieval, as well as foster deep learning with understanding for future application. To actively engage the students employing a few of the examples elaborated in the section below is important to develop deeper understanding of the subject/content matter. Use of humour is another useful technique to enliven a large-group lecture session. Some teachers have the gift of cracking impromptu jokes with good responses from students. Such teachers should not hesitate to do so at appropriate intervals during a lecture.

The final stage involves summarising the key-points. It is useful to summarise key-points at the end of a given lecture which will serve as 'recall' for students. At this stage the teacher could pose some questions for the students to discuss or answer and use an audience response system to share their responses/feedback as discussed in the following section.

A well organised and systematic large-group teaching or lecture together with handouts of key points in the large-group activity/lecture are usually much appreciated by students. A focused handout fosters deeper learning when it allows students more time to listen, think, engage and provides a framework on which students can build their understanding of a topic. Useful handouts also direct the students for in-depth learning by including exercises, questions to elaborate the topic further as well as suggesting further reading or resource materials [43].

'The difficulty lies, not in the new ideas, but in escaping from the old ones, which ramify, for those brought up as most of us have been, into every corner of our minds' (John Maynard Keynes, English economist, 1883–1946 [44]).

Perhaps the words of wisdom uttered by Keynes will resonate with senior teachers who, in the past, have had to regularly deliver instruction through 'informing' students. Such a habitual practice may be difficult '… to escape from', as the practice would be deeply ingrained in the minds of the teachers.

However, teachers must discard the 'old school' mentality and be prepared to accept and design new instructional strategies based on firm evidence obtained from advances made in the learning sciences, and also because the educational landscape has changed considerably.

Strategies to Aid Large-group Learning

Several strategies have been designed for the delivery of lectures to students in large groups. One can employ these strategies during the large-group activity to promote learning that is active and collaborative. If managed well these learning situations foster good teacher-student and student–student relationships as well as challenge them to develop their academic abilities. This section will elaborate on the various strategies that can be considered for use by teachers (who must, of course, first accept the significant role change that they themselves must undergo) in the education and training of health professional students in the twenty-first century [45].

Questioning

During the lecture, the teacher could employ the principle of the '5 P's as a strategy to engage the learners (see Figure 8.3).

The teacher first *poses* a question to students, and then *pauses* to allow students time to think about and to process the answer to the question. If no answer is forthcoming, after a short period of time, the teacher then *prompts* the students; if there is still no answer forthcoming, the teacher then has to *provide* and *process* the answer. The latter simply refers to a clear explanation of the expected answer. It is best to pose questions that relate to 'why', 'how', and 'when', rather than 'what' questions, in order to achieve higher-order learning outcomes.

Figure 8.3 Engaging learners though questioning.

The 'Think-Pair-Share' technique can be used in a more formal questioning setting. This requires good planning and incorporating into the large-group teaching session. Think-Pair-Share strategy helps to break the monotony of the large-group teaching session and engages students actively with their peers and the teacher. This is especially effective during the main body of the lecture.

- THINK: the teacher poses a question and allows students some time to think and process the question.
- PAIR: the teacher then requests the students to form pairs and, between them, to discuss what they think are the likely answers.
- SHARE: the teacher then requests a few pairs to share their answers with the whole class.

Promoting Interactivity

Evidence suggests that interactive lectures promote understanding in the audience rather than simply encouraging fact retention [46]. Interactivity also offers the opportunity for lecturers to check assumptions and for participants to feel included, it engenders collective learning, and aids retention of learning through facilitating different learning styles and empowering the audience in the learning process.

Buzz Groups

While most traditional lecturers try to discourage talking between students attending lectures, an alternative approach is to recognise that this is likely to happen anyway and will encourage interchange of ideas at appropriate intervals. The concept is to set a task for groups of two or more students to discuss at regular intervals. The task can be related to the preceding section of the lecture or a controversial issue raised, or (perhaps even braver) it can provide the opportunity for the learners to 'shape' the next section of the lecture. The challenge is always to regain the interest of the students, but the task should include a requirement for selected feedback to the rest of the audience to maximise group learning.

Snowballing

This aptly describes a process where each individual member of the audience is invited to work alone on a problem or issue for a couple of minutes, then share with their neighbour for a similar length of time, the two of them generating a discussion which they then share with another pair, and so on. This process can be time consuming, so it generally suits smaller groups, and enough time should be included to get useful feedback from all the groups before the lecture continues or comes to an end.

Nominal Group Technique

For some groups and suitable subject material it may prove desirable to divide the audience into smaller groups with a set task. This particular technique gives all group members the right to express any opinion without challenge. All the group contributions are written down by a nominated group leader and time is then given for elucidation, explanation, and challenge to the ideas raised. At this point similar items can either be aggregated if they are very similar, or disaggregated if a group's members feel they are materially different. After this period of discussion, the group members are invited to vote for the issues raised, identifying their perceived importance, and the key issues are then fed back to the wider audience.

'Flipped Classroom'

Techniques are increasingly being used to engage students to learn content with a greater engagement and analysis to promote deeper understanding. The 'flipped classroom' is an instructional strategy which uses a type of blended learning approach reversing the traditional learning environment by delivering instructional content, often online, outside of the classroom. During the face to face session, material that may have traditionally been considered homework assignments is discussed or the lecturer may get the class to discuss related exam questions such as MCQs in the lecture theatre – using maybe some of the methods mentioned above. Some lecturers also get the students to watch streaming lectures or collaborate in online discussions as classroom activities with careful supervision of a facilitator [47–49]. The strategy has two broad steps:

- Homework assignment: relevant reading material (online or hardcopy) relating to a lecture topic is provided to students as homework; alternatively, students may be required to review a video clip for subsequent discussion in class.
- In-class activity: often, during the actual lecture period itself, a question/answer session is conducted based on the homework assignment; the session is facilitated by the teacher. However, sometimes the in-class activity may involve a problem-solving session based on the lecture topic [50].

Audience Response Systems (ARS)

ARS are best used for polling student responses to a question posed by the teacher. Several systems can be used, including colour-coded cards: students simply raise their preferred colour cards in response to a question posed by the teacher to match the colour representing the likely answer(s). The use of electronic 'clickers' is popular with many institutions since the responses can be analysed immediately and shared with students. The analysis of the students' response patterns can be the basis for discussion. The students opt for their preferred answers by 'clicking' on various options available on a key pad supplied to each student and operated through radio frequency; such a system is now less commonly used because of costs involved in purchasing the system from a vendor and due to the availability of online apps which could be downloaded to student mobile phones.

Due to the wide use of electronic devices (smart phones, iPads, tablets, PCs, etc.) by learners themselves, and that many ARS applications are low cost or free, these are now widely used by institutions as effective interactive tools during large-group teaching sessions (e.g. Poll Everywhere).

Conclusion

Today, instructional learning strategies are designed to empower students to progressively take greater initiative and responsibility to direct and to manage their own learning, as well as their professional and personal development. Transformative learning principles will contribute to such an aim. Lectures to students in large groups should adopt and apply transformative learning principles in their instructional design. Only then will lectures be able to contribute fully to the overall institutional goal of transforming students to think, act, and behave as future practitioners who are fit to deliver twenty-first century health care to meet the needs, demands, and challenges of patients and the community.

Acknowledgements

The editors acknowledge Andrew Long and Bridget Lock, some of whose contribution to the first two editions of *Understanding Medical Education* has been retained as a complement to the new material in this chapter.

References

1 Dunkin, M.J. (1983). A review of research on lecturing. *Higher Education Research and Development* 2: 63–78.
2 Charlton, B.G. (2006). Lectures are an effective learning method because they exploit human nature to improve learning. *Medical Hypotheses* 67: 1261–1265.
3 Brown, G. and Manogue, M. (2001). AMEE Medical Education Guide No. 22: refreshing lecturing: guide for lecturers. *Medical Teacher* 23 (3): 231–244.
4 Smith, B. and Staff and Educational Development Association (1997). *Lecturing to Large Groups*. Birmingham: SEDA.
5 Smith, J.T. (2016). *Key Questions in Education: Historical and Contemporary Perspectives*. London: Bloomsbury Publishing.
6 Zepke, N. and Leach, L. (2010). Improving student engagement: ten proposals for action. *Active Learning in Higher Education* 11 (3): 167–177.
7 Schmidt, H.G., Wagener, S.L., Smeets, G.A.C.M. et al. (2015). On the use and misuse of lectures in higher education. *Health Professions Education* 1 (1): 12–18. doi: 10.1016/j.hpe.2015.11.010.
8 Frenk, J., Chen, L., Bhutta, Z.A. et al. (2010). *Health Professionals for a New Century: Transforming Education to Strengthen Health Systems in an Interdependent World*. The Lancet Commissions.
9 Khosa, M., Ncube, A., and Tshabalala, T. (2015). Effective teaching of form four history at Zimbabwe's Dakamela secondary school: a comparative study of group discussion and lecture methods. *British Journal of Education, Society & Behavioural Science* 8 (1): 34–36.
10 Christensen, C.R. (1991). *Education for Judgment: The Artistry of Discussion Leadership*, 312. Boston: Harvard Business School Press.
11 Lave, J. and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.
12 Swanwick, T. (2005). Informal learning in postgraduate medical education: from cognitivism to 'culturism'. *Medical Education* 39: 859–865.
13 General Medical Council (1993). *Tomorrow's Doctors. Recommendations on Undergraduate Medical Education*. London: GMC.
14 Entwistle, N.J. (1988). *Styles of Learning and Teaching*. London: David Fulton.

15 Tulving, E. (1972). Episodic and semantic memory. In: *Organisation of Memory* (ed. E. Tulving and W. Donaldson), 381–403. London: Academic Press.
16 Neisser, U. (1976). *Cognition and Reality*. San Francisco, CA: Freeman.
17 Verner, C. and Dickinson, G. (1967). The lecture: an analysis and review of research. *Adult Education* 17: 85–100.
18 Gardiner, L. (1994). Redesigning Higher Education: Producing Dramatic Gains in Student Learning. George Washington University, Washington, DC. ASHE-ERIC Higher Education Report No. 7. http://bern.library.nenu.edu.cn/upload/soft/0-article/021/21052.pdf (accessed 15 November 2012).
19 Ausubel, D.P. (1968). *Educational Psychology: A Cognitive View*. Austin, TX: Holt, Reinehart and Winston.
20 Stuart, J. and Rutherford, R.J.D. (1978). Medical student concentration during lectures. *Lancet* 2: 514–516.
21 Higher Education Academy Engineering Subject Centre (2005). Guide to Lecturing. https://www.heacademy.ac.uk/knowledge-hub/engineering-subject-centre-guide-lecturing (accessed 25 June 2018).
22 Ellis, A.W. and Young, A.W. (1988). *Human Cognitive Neuropsychology*. Hillsdale, NJ: Erlbaum.
23 Bligh, D.A. (2000). *What's the Use of Lectures?* San Francisco, CA: Jossey-Bass.
24 Freeman, S., Eddy, S.L., McDonough, M. et al. (2014). Active learning increases student performance in science, engineering, and mathematics social sciences – psychological and cognitive sciences. *Proceedings of the National Academy of Sciences of the United States of America* 111 (23): 8410–8415; published ahead of print May 12, 2014. doi: 10.1073/pnas.1319030111.
25 Bajak, A. (2014). Lectures aren't just boring, they're ineffective, too, study finds. http://www.sciencemag.org/news/2014/05/lectures-arent-just-boring-theyre-ineffective-too-study-finds (accessed 25 June 2018).
26 Frank, J.R., Snell, L.S., Cate, O.T. et al. (2010). Competency-based medical education: theory to practice. *Medical Teacher* 32 (8): 638–645. doi: 10.3109/0142159X.2010.501190.
27 Nordquist, J. and Laing, A. (2014). Spaces for learning – a neglected area in curriculum change and strategic educational leadership. *Medical Teacher* 36 (7): 555–556.
28 Payne, J. (1883). *Lectures on the Science and Art of Education*, 313. Thoemmes Press.
29 Barr, R.B. and Tagg, J. (1995). From teaching to learning – a new paradigm for undergraduate education. *Change* 27 (6): 13–25.
30 Meyers, C. and Jones, T.B. (1993). *Promoting Active Learning. Strategies for the College Classroom*. San Francisco, CA: Jossey-Bass Inc.
31 Cole, P. and Wilson, B. (1991). *Cognitive Apprenticeships: An Instructional Design Review*. Institute of Education Sciences.
32 McMahon, G.T. and Drazen, J.M. (2014). Introducing NEJM knowledge+ and its adaptive personalized learning. *New England Journal of Medicine* 370 (17): 1648–1649. doi: 10.1056/NEJMe1401777.
33 Gresham, C.L. and Philip, J.R. (1996). Problem–based learning in clinical medicine. *Teaching and Learning in Medicine* 8: 111–115.
34 Bernstein, D. and Bass, R. (2005). The scholarship of teaching and learning. *Bulletin of the American Association of University Professors* 91 (4): 37.
35 Austin, A.E. and Baldwin, R.G. (1991). Faculty Collaboration: Enhancing the Quality of Scholarship and Teaching. ASHE-ERIC Higher Education, Report No. 7, 138.
36 Cooke, M., Irby, D.M., and O'Brien, B.C. (2010). *Educating Physicians: A Call for Reform of Medical School and Residency, The Carnegie Foundation for the Advancement of Teaching*. San Francisco, CA: Jossey-Bass.
37 Christensen, C.R. (1991). Premises and practices of discussion teaching. In: *Education for Judgment: The Artistry of Discussion Leadership* (ed. C.R. Christensen, D. Gavin and A. Sweet), 15–34. Boston: Harvard Business School.
38 Bransford, D.J., Brown, A., and Cocking, R. (2000). *How People Learn: Brain, Mind, Experience, and School: Expanded Edition*. National Academies Press.
39 Ambrose, S., Bridges, M., Lovett, M. et al. (2010). *How Learning Works: 7 Research-Based Principles for Smart Teaching*. Wiley.
40 Samarasekera, D.D., Nayak, D., Yeo, S.P., and Gwee, M.C.E. (2014). Teaching for effective learning in higher education: focusing on learning and moving from a 'based learning' mind-set. *Asian Journal of the Scholarship of Teaching and Learning* 4 (2): 113–123.
41 Morton, A. (2003). Lecturing to large groups. In: *A Handbook for Teaching and Learning in Higher Education, Enhancing Academic Practice*, 3e (ed. H. Fry, S. Ketteridge and S. Marshall), 58–71. London and New York: Routledge.
42 Front cover of Time magazine, April 3, no. 13, 1995. http://content.time.com/time/magazine/0,9263,7601950403,00.html?iid=sr-link3 (accessed 27 June 2018).
43 Cantillon, P. (2003). ABC of learning and teaching in medicine: teaching large groups. *BMJ [British Medical Journal]* 326 (7386): 437.
44 Keynes, J.M. (1936). *The General Theory of Employment, Interest and Money*. London: Macmillan.
45 Zepke, N. and Leach, L. (2010). Improving student engagement: ten proposals for action. *Active Learning in Higher Education* 11 (3): 167–177.
46 Social Policy and Social Work Subject Centre. Activities to try in lectures. The Higher Education Academy Social Policy and Social Work (SWAP) website. www.swap.ac.uk/resourcespubls/digests.html (accessed 1 August 2013).
47 Bergmann, J. and Sams, A. (2012). *Flip Your Classroom: Reach Every Student in Every Class Every Day*. U.S.A: International Society for Technology in Education.
48 Prober, C.G. and Khan, S. (2013). Medical education reimagined: a call to action. *Academic Medicine* 88 (11): 1407–1410.
49 McLaughlin, J.E., Roth, M.T., Glatt, D.M. et al. (2014). The flipped classroom: a course redesign to foster learning and engagement in a health professions school. *Academic Medicine* 89 (2): 236–243.
50 Moffett, J. (2015). Twelve tips for 'flipping' the classroom. *Medical Teacher* 37 (4): 331–336.

Further Reading

Bligh, D.A. (2000). *What's the Use of Lectures?* San Francisco, CA: Jossey-Bass.

Brown, G. and Manogue, M. (2001). AMEE Medical Education Guide No. 22: refreshing lecturing: guide for lecturers. *Medical Teacher* 23 (3): 231–244.

9 Learning in Small Groups

Peter McCrorie

Department of Medical Education, University of Nicosia Medical School, Nicosia, Cyprus
Institute of Medical and Biomedical Education, St George's University of London, London, UK

KEY MESSAGES

- Always consider whether a session can better be run in small groups.
- Always plan the session well ahead.
- Vary the session plan according to the size of the group, the venue, and the purpose of the session.
- For large groups (around 30), run the session in workshop format, by splitting the group into four or five smaller groups.
- Include a range of formats in any one session, recognising that active learning is likely to be more successful than passive learning.
- Consider your own role in the group process; remember that being the group leader is only one of several options.
- Deal with group issues as and when they arise, bearing in mind that it is often better to let the group address the issues themselves.

Introduction

Two people give a lecture on a medical topic. One is charismatic and inspirational. He holds the audience in his hands – they are on the edge of their seats throughout the lecture. He does not use notes, illustrates his talk with a few well-chosen photos, and uses the traditional 'talk and chalk' approach. He regularly wins the 'best lecturer' prize. Contrast this with the second lecturer – a very dry and serious speaker, who has carefully prepared every word he intends to use, uses a myriad of PowerPoint slides, each laden with detailed information, and who whips each slide away before his audience has had a chance to get down the half of what he says. Now imagine that each lecturer sets an examination question based on his lecture. Which question do you think the students would perform better in?

You may be surprised to find out that they would get higher marks in the second, less charismatic, lecturer's question. Why? Because in order to make any sense of his lecture, the students would have made a beeline for their textbooks and this self-directed learning would have enabled them to learn and retain the information more effectively. While the first lecture was stunning at the time, the students would have assumed they would remember every word of it several weeks later and therefore would not have bothered reading around the subject – an approach that would let them down in their examination.

The conclusion from the above is that an active, self-directed approach is likely to have a much greater impact on a student's learning than passive, lecture-based learning. Lectures should be used sparingly and then for broad overviews, summaries, and difficult topics. Sadly, only about 5% of what is taught in lectures is actually retained, and too many people use lectures for imparting large quantities of detailed information that could easily be picked up, more effectively, by reading a textbook. (See Chapter 8 for more on lectures and large groups.)

Learning in small groups is a sort of halfway house, involving active learning, or more precisely *interactive* learning. But here the direction of learning is determined by the group as a whole, rather than the individual. Group learning can be a most rewarding and effective experience at all stages in a medical career – undergraduate, postgraduate, and throughout years of continuing professional development. Too many people still consider that unless material is passed on from professor to student or consultant to trainee in a formal lecture, it would not be learnt properly. This chapter aims to challenge that assumption and to provide some ideas and suggestions, backed up by theory and the available evidence, for getting the most out of the small group learning experience.

The Learner Experience

If you were to ask a group of medical graduates to relate their personal experiences of small group work in their undergraduate course or postgraduate training, their responses would vary considerably. But the two factors most likely to underpin their accounts would probably relate to *tutor variability* and *curriculum philosophy*.

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

A *tutor* who creates a relaxed atmosphere, who keeps the group focused on the task in hand, who deals effectively with group dynamics, who allows the learners to take ownership of their learning, and who helps make the process an effective, yet enjoyable, experience will produce strongly positive and enthusiastic responses. On the other hand, a tutor who talks all the time, who does not encourage group participation, who belittles learners who answer questions incorrectly, who has clear favourites in the group, and who creates an atmosphere where one is frightened to open one's mouth does an extreme disservice to education.

In a university which has a traditional curriculum heavily dependent on imparting information through lectures, where students or trainees are seen as an inconvenience and a hindrance to research and their clinical practice, rather than as a benefit and an opportunity to mould and influence the future medical workforce, and where time spent by large numbers of tutors on small group teaching is perceived as an ineffective use of staff time, such small group teaching is likely to be received unfavourably. A progressive university or postgraduate team that values its students, believes in active learning, encourages small group teaching, and trains its staff in the art of facilitation of learning will undoubtedly invoke a positive response.

At the postgraduate level, lectures are of even less value than at the undergraduate level, yet they are used just as frequently. Trainees gain far more from small group teaching sessions and on-the-job training than from, for instance, revision courses for passing membership exams. It is perfectly possible for consultants to carry out effective education while at the same time maintaining their clinical practice load. Bedside teaching and teaching ward rounds are examples of this – effectively small group learning for junior doctors and medical students (and indeed other health care professionals) together. The effectiveness of bedside teaching is discussed by Jolly in his book entitled *Bedside Manners* [1]. With the explosion of e-learning and distance learning, online training has proved a useful addition to the ever-increasing toolkit available for training. An example of how this can be used for small group learning is *The Virtual Ward Round* [2], an ingenious but simple device for students and junior doctors to explore ward scenarios, evolving over a period of several days.

De Villiers et al. [3] in a report of an evaluation of a continuing professional development programme for primary care medical practitioners in South Africa, suggest that the following aspects should be incorporated into the design of small group activities to make them effective:

- build on prior knowledge and experience
- relate to the perceived learning needs of the participants
- involve active learning
- be focused on problems
- be immediately applicable to practice
- involve cycles of action–reflection
- allow the acquisition of technical skills.

Steinert [4] researched medical students' perceptions of small group teaching in a Canadian medical school. Her conclusions at the undergraduate level were not a million miles away from de Villiers' findings at the postgraduate level, concluding that positive student perceptions of small group teaching were related to:

- effective small group facilitation
- a positive group atmosphere
- active student participation and group interaction
- adherence to small group goals
- clinical relevance and integration
- cases that promote thinking and problem solving.

The inefficiency argument is often used by those opposed to small group teaching. Twenty-five tutors spending two-to-three hours working with students in small groups of eight is clearly less efficient than one lecturer talking to 200 students all at once – but only in terms of delivery of material. Efficiency does not just take account of the delivery of *teaching*; what the students *learn* is what matters. I return to the point made earlier, that the taught curriculum is not the same as the learnt curriculum. Through small group discussion, much more is retained, especially if the learning is contextualised, for example, in case scenarios. In terms of efficiency of learning, small group work wins hands down. More particularly, small group work encourages critical thinking, which, although certainly not impossible, is less common in lectures. See Box 9.1.

What Constitutes a Small Group?

The ideal size of a small group is probably around seven or eight. If the group is smaller, it becomes too threatening, the synergistic effect – the collective knowledge of the group being greater than the sum of the knowledge of each member of the group – is reduced, and the interaction is less successful. If the number increases above eight, some learners can get by without fully participating, or without joining in at all, and others are less able to get their voice heard, because the size of the group deters them from

BOX 9.1 WHERE'S THE EVIDENCE: Does small group work encourage critical thinking?

Most of the research around small group work relates specifically to problem-based learning (PBL). The nature of courses structured around PBL is such that small group work, and the subsequent self-directed learning, is the main vehicle by which students gain information, with lectures contributing relatively little to students' knowledge acquisition. This is in contrast to courses where lectures and other forms of didactic teaching provide the main source of information for students, and where small group teaching is really an adjunct to their learning. Tiwari et al. [5] have compared the effects of these two course styles on the development of students' critical thinking. They found that PBL students had significantly higher levels of critical thinking (measured using the California Critical Thinking Disposition Inventory [6]) compared with students on a predominantly lecture-based course. Furthermore, they continued to have higher scores for two years afterwards. Similar conclusions have emerged from a variety of other sources [7–9].

expressing their point of view. For PBL, the size of the tutorial group rarely exceeds eight – the process would simply fall apart were there any more in the group. G.R. Norman (personal communication) carried out an email survey of a cross-section of medical schools around the world using PBL, asking about the size of group they felt was appropriate, and eight turned out to be the norm. Peters [10] cites academic evidence that supports his own findings that the optimum size of a group is somewhere between 5 and 10.

Some medical schools are so short of willing and able teachers that they operate small groups of 20 or more. This can still be successful, however, if the larger group is run in workshop format, where the whole group is given a task to carry out in smaller subgroups of seven or eight. Here, one facilitator handles several groups. Many postgraduate training courses are run this way. Team-based learning is another approach used in some medical schools to get round the issue of insufficient facilitators. This is discussed later in this chapter.

Group size is probably less important than what the group actually does. The purist view of small group teaching is that it must be learner-centred, with all students joining in free discussion of a particular topic. Some teaching may indeed take place in small groups but sits outside this definition. The seminar is a case in point, where invited speakers present on a topic about which they are passionate. The seminar has its place in a university – particularly at level of the Bachelor of Science or Masters – but is invariably teacher-centred, with any discussion taking the form of questions and answers.

Even within the confines of our working definition, a wide range of styles of small group work exists and many of these are discussed later.

Housekeeping

Before embarking on small group teaching, some thought needs to be given to the environmental arrangements. It seems obvious to say it, but the first requirement is to hold the session in an appropriately sized room. What is appropriate depends on the number of learners participating. Consideration should also be given to heating, lighting, and temperature control in the room, all of which are the unique responsibility of the group leader.

For a group of eight or so undertaking PBL or some similar activity, a room with a table in it, preferably round or oval, and big enough for all to sit around it comfortably – including the tutor, who should be part of the group and not outside it – is ideal. The walls of the room should be lined with whiteboards or flipcharts for students to write and draw on.

For a session where argument is more important than capturing information, for example, an ethical debate, then a circle of chairs is all that is required. Sitting in a circle has two advantages:

- everyone in the circle is equal
- everyone has eye contact with each other.

There are different perspectives on where the tutor should be positioned. Personally, when I am facilitating a group, I prefer to be *in* the group rather than *separate* from it, particularly if I have a guiding role (e.g. in the progressive release version of PBL). Others believe the opposite – that the tutor should deliberately sit *outside* the group to enable the students to interact with each other, rather than direct their discussion towards the tutor (e.g. in the short case version of PBL). It's really down to personal preference and the extent of tutor involvement in the learning process.

For larger groups, the workshop format is the method of choice, and hence the room has to be larger and laid out in cabaret style (several round tables with chairs round them to allow for working in separate groups). There needs to be space at the front of the room for the tutor, PowerPoint, and/or overhead projector facilities (including an accessible power source), and whiteboard(s) that everyone can see. Ideally, each group should have a flipchart as well. Again, if there is no requirement for a table, then several circles of chairs will suffice.

Positioning of the tutor is different for a workshop. Plenary sessions are commonly part of a workshop, usually at the beginning, to introduce the topic or task, and at the end, to take feedback from the groups and summarise what has been achieved. Since the tutor needs to be able to engage everyone together, he or she needs to stand or be seated at the front of the class for these plenaries. However, for the rest of the session, the tutor's job is to go round the groups to check how they are getting on.

The Role of the Tutor

The tutor can adopt a range of roles, depending on the nature of the small group session. Rudduck [11] suggests that tutors can have four differing roles:

- the *instructor*, who is there to impart information to the students
- the *devil's advocate*, who intentionally adopts a controversial view in order to stimulate discussion
- the *neutral chair*, who literally chairs the discussion but expresses no strong opinions
- the *consultant*, who is not part of the group, but is there for the students to ask questions.

To these roles, I would add a fifth:

- the *facilitator*, similar to the neutral chair, but with more of a guiding role, for example, asking the group open-ended questions to facilitate their progress with the task in hand. The facilitator need not be the chair of the group – the role of chair might fall to a student.

Tutors may adopt any or all of the above roles during the course of a small group teaching session. They may begin the session in instructor mode, defining the task for the students. During the rest of the session, they may adopt a more facilitatory role, at times prompting discussion by playing the role of the devil's advocate or taking up a chairing role. If the students get stuck, they may even take on the role of expert to allow them to move on (not in PBL, of course, where the tutor must reflect any questions back to the group and encourage them to research the answers for themselves).

Richmond [12] defines more specific roles for tutors, which he refers to as 'strategic interventions':

- to start and finish group discussion by outlining the group task, summarising the group's or groups' achievements and conclusions, and setting further learning activities
- to maintain the flow of content, for example, by preventing sidetracking and keeping the group(s) focused on the task
- to manage group dynamics by encouraging the shy or bored student(s) and by handling the dominant, aggressive, offensive, or nuisance student(s)
- to facilitate goal achievement through open questioning, making suggestions, and checking group understanding
- to manage the group environment by keeping an eye on the time and dealing with any distractions (e.g. noise, insufficient flipchart paper, pens running dry, heating).

According to Brown [13], tutors need a range of skills in order to make a success of small group teaching. These include questioning, listening, reinforcing, reacting, summarising, and leadership. But the real skill, one that Brown refers to as a *super skill*, is the skill of knowing when to use which skill.

Getting Started

The first time the tutor meets with a new group of students or trainees is always exciting. Is it going to be the group from heaven or the group from hell? Often, the first encounter is the defining moment for the group. How the tutor handles the opening small group teaching session may establish the atmosphere for the sessions that follow.

The group members may or may not know each other. At the start of the course they almost certainly would not; but even later on, particularly in undergraduate curricula, because of the trend to have large numbers of students on courses, group members may still not know every other person on the course. *Ice-breakers* have a useful role here – ways of introducing strangers, trying to relax the group, and getting group members to interact with each other.

There are many ice-breakers in common use – some simple, some elaborate. A standard technique is to ask the group members to pair up and talk to someone they do not know. Their task is to learn something about their partner and report it back to the group. This exercise works better if the task is fairly specific, and in addition to their partner's name and brief biographical details, participants may be asked to report back on the most interesting place their partner has visited, something unusual that they have done in their life, or something they have done of which they are proud. Participants could also be asked to imagine themselves as a musical instrument or a vehicle or a colour, and to describe which most closely matches their character – all of which can be quite revealing!

An example of a more complex ice-breaker, which needs a bit of time, is to divide the group into two teams which have to compete with each other to complete a task, e.g. build a bridge of specific dimensions and requirements out of Lego bricks in a specified time. The dimensions and rules they have to adhere to make this a hard task, but it enables the tutor to observe how each member of the team behaves under pressure. It also provides an opportunity for discussing people's behaviours at the end of the exercise, and seeing how each team worked. Such exercises or games are common on off-campus postgraduate training courses.

Having got the group talking, the next thing to do is set the ground rules. This is a really important activity and should never be omitted. It is essential that the group itself comes up with the ground rules – they must not be imposed by the tutor. Box 9.2 shows examples of ground rules that my own undergraduate students have come up with. Probably the most important ground rule is to do with valuing each other's contribution. That way a relaxed atmosphere is created where no one is embarrassed about saying something stupid. Humiliation must be avoided at all costs.

Techniques to Use in Small Groups

Following on from the ice-breaker, a good technique to get a group engaged in the topic under discussion is known as *snowballing*. The students are given a question, for example, 'How might cystic fibrosis affect the life of a 15-year-old young man?', and asked to think about it individually, without conferring with anyone else. After five minutes, the students are asked to pair up with another student and discuss their thoughts with each other. After another five minutes, the pairs are invited to join with another pair, and all four students continue their discussion. Two sets of fours then compare notes, and so on. The process is called 'snowballing' because of its resemblance to a snowball rolling down a hill, gathering more and more snow, and getting bigger and bigger in size. The big advantage of this technique is that everyone has to participate, even the most reticent of students. By allowing time for them to gather their own thoughts and then share these thoughts with one other, it gives even the shyest student confidence to speak.

BOX 9.2 Examples of ground rules for small groups

- Turn up punctually.
- Finish on time.
- Do not talk over each other.
- Do not interrupt.
- Value each person's contribution.
- Respect each other's viewpoint.
- Turn off mobile phones.
- Turn up prepared.
- Join in the discussion.
- Keep personal issues outside.
- Maintain confidentiality within the group.

A variation on snowballing is called *jigsaw groups*, where the participants divide into small groups and discuss an issue. After a period, the groups re-form into new groups, with each of the new groups containing one member of each of the old groups, thereby maximising the mix of participants (see Figure 9.1).

Brainstorming is another widely used technique. Again, learners are given something to think about (e.g. possible diagnoses for a patient who presents with lower back pain). One group member acts as scribe and writes all the suggestions from the group – in this case possible diagnoses – on a whiteboard. Absolutely everything is written up, no matter how unlikely the suggestion may be. No one is allowed to make any value judgements on the suggestions at this stage. A reflective analysis of what has been written up follows, where items are prioritised, grouped together, or removed altogether.

A third introductory technique involves the use of *buzz groups*. This technique is more generally used in lectures, but can also be used in small groups if the group becomes stuck in its thinking. The tutor who senses an impasse can interrupt the discussion and throw in a question for the group to ponder in pairs or threes, to help the members get back on track. If students are struggling to understand respiratory acidosis, for instance, the tutor might ask them to discuss in pairs what the role of the kidney is in maintaining pH or what blood buffers exist. When this technique is used in lectures, there is a loud 'buzz', as members of the audience start talking with their neighbours – hence the name.

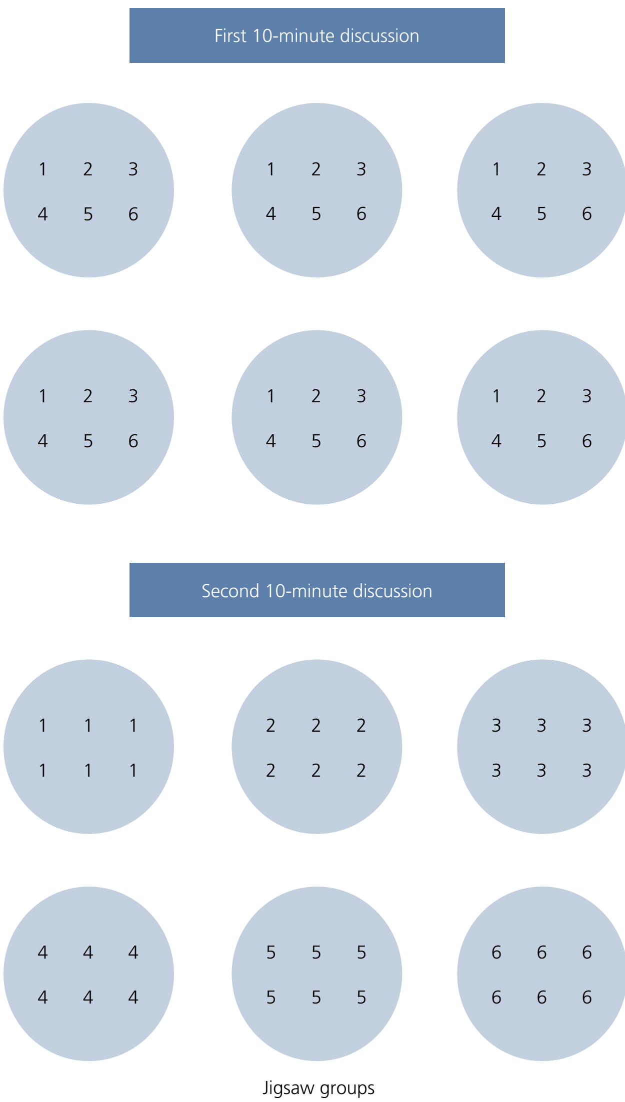


Figure 9.1 How jigsaw groups work.

Further on in the life of the learning group, a number of other techniques may be used.

Simply *chairing a discussion* – or, better still, getting one of the group members to do so – works well when there is more than one perspective about an issue, such as a genuine ethical dilemma like abortion, animal experimentation, or euthanasia. Occasionally with moral and ethical arguments, the tutor may need to step in to clarify matters of fact, such as primary legislation, case law, or professional guidelines – an example of the tutor acting as a consultant or expert. A more imaginative and hands-off approach to such issues would be to get the students to set up a *formal debate*, with students primed to speak for or against a particular motion, such as 'This house believes that patients with coronary artery disease, who continue to smoke after they have been repeatedly counselled about the dangers of smoking, should be refused heart bypass surgery'. The advantage here is that the relevant factual elements to the discussion are prepared in advance.

A variation of the discussion group is called *line-ups*. The tutor makes a controversial statement, such as 'Doctors should be allowed to hasten the death of old people who have a terminal illness'. The tutor identifies a point in one corner of the room where everyone who strongly agrees with the statement should stand. Another point is identified, as far away as possible, where those who strongly disagree should stand. The rest of the group members have to line up somewhere between the two points, according to how much they agree or disagree with the statement. The participants have to talk to the others in the line and argue their point of view. The tutors, and indeed the participants, get an immediate view of the spread of participants' opinions on the topic. If the line-up is carried out both at the beginning of a session and at the end, the tutor can gauge if there has been a change of opinion after the session has been run and the participants are better informed.

There are a number of variations on the discussion approach to small group teaching. Some of these are less than satisfactory, however. One approach used frequently in postgraduate medical education is the *journal club*. Staff are asked to present their comments on recent papers in the medical literature. This works well, provided the topic is of direct relevance to everyone and there is plenty of opportunity for discussion. It is less successful at the undergraduate level, where it is more usual to run the session along the lines of a *syndicate* presentation, where a topic such as diabetes is subdivided into several subtopics (e.g. mechanism of action of insulin, diabetic ketoacidosis, clinical presentation of diabetes, treatment of diabetes) and each of these subtopics is given to a student or pair of students to research. They then all report back at the next tutorial. The problem with both methods is that the presentations become mini-lectures, are frequently delivered in an uninspiring manner and, frankly, bore the pants off everyone. Discussion is limited because only the chosen few have researched the topic, and then only the subtopic they were allocated. Unless everyone researches everything, the only people who join in the discussion are those who researched the topic and the tutor!

Another old favourite at the undergraduate level is the *post-lecture tutorial*. This is another disaster site. Most students come unprepared and have not read up around the lecture topic (indeed, they may not even have attended the lecture or may be blissfully unaware what lecture topic is being discussed). Many students see it as an opportunity to listen to a rerun of the original lecture, which is what it often turns into, particularly if the tutor gets monosyllabic responses to the questions they ask to try to stimulate the students. There is really only one solution to this – ensure that students are given a self-directed learning exercise to prepare *before* the tutorial. A good example would be to give them a detailed case study to read up, with some questions attached for them to research, for example, a case history of a patient with a peptic ulcer, followed by a series of questions on, for instance, medical imaging, regulation of intestinal pH, *Helicobacter pylori*, drug treatment, and dietary advice. In this way, the students know what the topic of the tutorial is, have done their homework, and therefore feel able to join in the discussion and get a lot more out of the whole exercise.

The *use of triggers* provides an ideal springboard for a tutorial. The tutor can design a whole session around one or more of these. They are commonly used when running sessions in the workshop format. At the beginning of the session, after a brief introduction, the tutor hands out a trigger. A trigger is simply a tool to get the discussion started. They might take various forms, such as the following:

- an electrocardiogram (ECG) strip from a patient with ventricular fibrillation – to stimulate a discussion about cardiac muscle, the sinoatrial node, ECGs, arrythmias, defibrillation
- a photomicrograph showing neoplastic growth changes to stimulate discussion about the appearance of cancerous tissue, the characteristics of cancer cells, metastasis
- a chest X-ray showing a pneumothorax to stimulate discussion about X-rays, pleura of the lungs, pneumothorax – its causes, presentation, and treatment
- an audiogram from a patient with age- and noise-related hearing loss to stimulate discussion about the anatomy of the ear, the hearing mechanism, sensorineural hearing loss, audiometry, Rinne and Weber's tests
- a copy of *Good Medical Practice* to stimulate discussion on what constitutes a good standard of practice and patient care, professionalism, working with colleagues, and probity
- an anonymised or mock-up of a patient record for students to discuss record-keeping in the context of a particular case
- a photograph of a patient with an obvious goitre and exophthalmos to stimulate discussion around the thyroid gland and thyroid disorders, including clinical signs, symptoms, and treatment
- a photograph of a man in a wheelchair playing with a young child to stimulate discussion around disability, psychosocial sequelae of chronic illness, social care, child care, single-parent families
- a family pedigree, showing the distribution of a genetic disorder such as haemophilia to trigger a discussion on genes and patterns of inheritance
- a video showing, for example, a doctor explaining to a couple that their baby has Down syndrome
- a paper, or excerpts from a paper, showing some statistical data, in order to get the students talking about P-values or odds ratios or randomised controlled trials.

Case-based Learning

There are a variety of ways of running small group learning sessions built around cases. The term 'case-based learning' covers a wide range of learning activities. The most obvious of these is teaching on *real* patients on the wards or in clinics. By its nature, this is opportunistic and dependent on whoever happens to be on the ward or in the clinic (see later). Teaching can, however, sometimes be planned in advance, especially in the general practice setting where particular patients can be invited to attend at specific times. That way, sessions can be organised that relate to a teaching module, e.g. patients with arthritis are invited to attend during teaching on the musculoskeletal system.

Problem-based Learning

In the early years, case-based learning may not relate to real patients but to paper-based patients or virtual patients. The most sophisticated of this type of case-based learning is *problem-based learning* [14–18].

In PBL, the patient is either paper-based or computer-based (so-called 'virtual patient'). There are two main variants of this: one where the case is presented in the form of a short summary (1–2 pages maximum); the other much longer version (20–30 pages) where the case is released in stages, a page at a time, by the tutor (presentation, history, examination findings, investigations, results [lab results, imaging, histopathology are all presented in detail], treatment, patient progress, complications, and outcome). The former is generally used for school leavers, the latter for graduates. The tutor has rather different roles in the two versions. In *short-case* PBL tutorials, the students take turns at being chair of the group and the tutor is passive, intervening only to ask the odd question to stimulate discussion or to keep the group on track. In the long, *progressive-release* version of PBL, the tutor acts as the chair, moves the group along when their discussion has come to a halt and occasionally asks specific, programmed questions. Both styles of PBL are student-centred and produce deep and high-level discussions, particularly with graduate students. In the virtual patient PBLs [19], the case does not develop linearly, but is branched. At certain points in the case, students are provided with a choice of actions (e.g. what to do next; which drug to prescribe, which tests to order) and, dependent on their response, the case can follow different paths, some of which may actually lead to the death of the patient. As you can imagine, this is a very strong learning point for the group of students; they never make the same mistake again!

PBL is generally reserved for the early years of the medical curriculum, but it can be adapted for use in the clinical years, using real patients [20], and there is no reason why it can't be used in the postgraduate years as well. Whatever the design of the PBL process, there are basically two steps involved:

- co-operative learning in the group
- self-directed learning outside of the group.

During the group process, students come up with *learning issues* – what they don't know or have to check up on. This defines their self-directed learning. All students research *all* the learning issues and leads to high level of discussion in the group when they feed back. In addition, they acquire a number of very useful skills, including:

- communication skills: active listening, presenting, questioning, responding, clarifying, empathising
- team work: contributing to/collaborating with/learning from others
- testing and applying knowledge, constructing/defending an argument
- giving and receiving constructive feedback.

Team-based Learning

Another small group learning approach is *team-based learning*. This is a five-stage process [21].

- Stage 1: students are given material to study on their own in advance of the team-based learning session, examples include a podcast of a lecture, or a DVD.
- Stage 2: students complete a test (usually MCQs) on their own (Individual Readiness Assurance Test) at the beginning of the group session.
- Stage 3: in their various teams, students redo the test and reach a team consensus about their answers (Team Readiness Assurance Test).
- Stage 4: the facilitator explains any concepts that students found difficult, to ensure they are all up to speed with the topic.
- Stage 5: the teams are all given a challenging clinical problem to address (the trigger), which requires them to apply the concepts they have been studying to a real-life situation (i.e. to contextualise their learning). The teams then simultaneously report back to the whole body of students, justifying their reasoning.

Other small group learning approaches include narrative-based learning [22] and task-based learning [23].

Role Play

Role play is another commonly used small group learning technique, particularly for learning clinical communication skills. Students are given a scenario to act out, e.g. taking a headache history, explaining a medication regime for eczema, or breaking bad news about a cervical smear result. Professional actors are often employed to simulate real patients, and students can try out their clinical communication skills in a safe environment. If the students are video-recorded as well, they can study their communication attempts in their own time and work on improving their technique before encountering the real thing. This technique works well at the postgraduate level too, especially with doctors who have been identified as having a communication problem. Watching your own performance can be very revealing and informative. Discussing your performance, captured on videotape, with a trained communication skills expert is invaluable and always leads to improvement. The secret is to have enough insight and be brave enough to do it in the first place.

Clinical Skills Teaching

Clinical skills are also taught in small groups. Here the trainer necessarily takes more of a teaching role. The clinician firstly demonstrates the skill to the group before letting them have a go. One commonly used technique [24] takes place in four stages:

- the tutor demonstrates the skill in silence (the silent run-through)
- the tutor runs through it again, but explains the rationale behind the technique at the same time, carefully describing each step in detail
- the students then talk the tutor through the technique
- finally, a volunteer student runs through the technique without tutor intervention.

The students then practise on each other, watched by their supervisor.

The Teaching Ward Round

The classic mode of small group teaching in both undergraduate and postgraduate training is the *teaching ward round*. There are many ways of doing this, some better than others. For best results, a teaching ward round should be just that – a ward round with two purposes – patient review and student/junior doctor teaching. It needs to have more time allocated to it than just a straight consultant ward round and for best results it needs to adhere to a few ground rules. Gill and Dacre [25] have written an informative guide for bedside teaching for the London Deanery. In the guide, they discuss what *not* to do. They delightfully call this the 'Carry on Doctor' approach.

- Start by gathering students round unsuspecting patient who has good clinical material for teaching.
- Embarrass the patient by exploring the abdomen without consent.
- Pick the most timid student to perform and present findings at the bedside.
- Dissect the student's errors at the bedside.
- Quiz students about management plan and diagnosis at the bedside.

- Suggest other students come back later to 'feel the spleen'.
- Move onto the next patient and start again picking on a different student.

Fortunately, they also provide plenty of ideas of how a teaching ward round *could* be run.

In summary, consultants should:

- seek patient consent for the interaction
- encourage active participation rather than passive observation
- concentrate on teaching of applied problem solving
- integrate clinical medicine with basic science
- closely observe students during the bedside interview/examination rather than rely on subsequent side room case presentations
- provide adequate opportunity for students to actually practise their skills on the patients during the teaching ward round
- be a good role model, e.g. for interpersonal relationships with patients
- make teaching patient orientated rather than disease orientated
- demonstrate a positive attitude towards teaching.

The TOSBA

Other ideas for teaching by the bedside include a novel approach from the Royal College of Surgeons of Ireland – the Team Objective Structured Bedside Assessment or TOSBA [26]. The TOSBA involves three groups of five students rotating through three ward-based stations (each station consists of an inpatient and a facilitator). Each group spends 25 minutes at a bedside station, where the facilitator asks consecutive students to perform one of five clinical tasks (history-taking, examination, assessment, management, discussion of a related topic). Every student receives a standardised grade and is provided with educational feedback at each of the three stations. I've observed this myself and it is a really good learning exercise.

Balint Groups

The *Balint group* [27, 28] is a special kind of group activity undertaken mainly by general practitioners (GPs), although increasingly patients, and even students, are becoming involved. Michael Balint was a Hungarian psychoanalyst who worked extensively with UK GPs in helping them to understand the psychology behind the doctor–patient relationship. He set up discussion groups with GPs to allow them to share their personal experiences of specific problems or dilemmas that had arisen in their practice. A modern-day Balint group consists of a handful of GPs who meet on a regular basis, often with a psychoanalyst and sometimes with one or more patients, to discuss specific issues concerning patients arising from their daily practice. The presenting clinician brings to the group cases that have given cause for thought. The purpose of the group is to increase understanding of the doctor–patient relationship, not to find solutions to the patient's problem. Discomfort or distress in the doctor are not ignored, but are worked through in the context of the needs and problems of the patient rather than of the doctor. Balint group members find that the benefit gained by sharing their experiences far outweighs any pain they may feel as a result of the experience. Balint groups are not for everyone and should not be entered into lightly.

Action Learning Set

Another type of group activity used in the postgraduate arena is the *action learning set*. An action learning set is a group of six to eight people who meet regularly to help each other learn from their experiences. The set is not a team, since its focus is on the actions of the individuals within it, rather than on a shared set of work objectives. Sets are usually facilitated by a set adviser whose responsibility is to create a suitable learning environment by encouraging, challenging, and focusing on learning. Action learning is based on the concept of learning by reflection on experience. It is underpinned by the cycle of experiential learning, as shown in Figure 9.2 (modified from Kolb [29]), where the stages of reflection and generalisation are worked through with the members of the set.

The action learning approach was first developed by Revans [30]. Each participant works on a project or task over the life of the set (which may be a few weeks, or spread over several months). The set decides on its own way of working, but usually a meeting involves participants taking turns to present their project to the set. This will normally involve the following:

- an update of progress on actions from the last meeting
- a discussion about current issues or problems
- an agreement on actions for the future.

Participants work with the presenter, by listening and questioning, to help them decide what actions to take. This kind of group is useful for individuals working on an educational or research project largely on their own.

Group Dynamics

Before considering the role of the tutor in addressing issues of poor group dynamics, it is worthwhile taking a look at how a group evolves during its lifetime. Tuckman [31] has

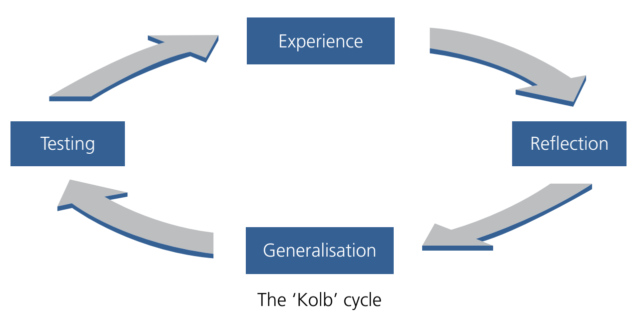


Figure 9.2 The 'Kolb' cycle (modified).

summarised it into four stages, which have been further interpreted by Walton [32] and Mulholland [33].

Forming

The participants get to know each other, form alliances, and establish themselves. The tutor must ensure they are introduced to each other.

Norming

The participants set the ground rules for working as a group. Being unfamiliar with what is expected of them can lead to some uncertainty and insecurity at this point. The tutor may need to explain the way the tutorials are to be run to alleviate anxiety at this point.

Storming

The group begins to function. Individuals adopt the roles with which they feel comfortable. One might be a leader; another good at initiating conversation; a third might be skilled at asking probing questions; a fourth might be good at clarifying and explaining; another might be good at keeping the team together, coming to the rescue of anyone who appears lost, frustrated, or angry; one or two may simply be good listeners and may only contribute to the process when they have something to say. There is the potential for a good deal of friction during this stage, while the group is sorting itself out. The tutor needs to keep a careful watch on the group members, identifying problems and trying to relieve tensions. If the participants are mature enough, they may be able to address their own issues as a group. This is to be encouraged. The less the tutor gets involved, the better. Such a group skill takes time to develop and the tutor will undoubtedly need to facilitate the process initially.

Performing

This is the position every group should aim for. Essentially the group has settled down and is functioning well. The group members are comfortable with their roles. There is a good atmosphere within the group and the goals of each tutorial are generally accomplished through successful collaboration. The tutor can now relax.

Unfortunately, some groups never achieve this state. This is usually due to one or two personality clashes and the odd difficult student. If the group cannot sort itself out the tutor must act, otherwise both participants and the tutor will dread meeting up for tutorials and attendance will fall off.

Adjourning

Also called *mourning*, this is the final stage of working in groups. The lifetime of the group has run out, and members move on to join new groups. This phase is a mixture of celebration and sadness (or relief, in the case of an unsuccessful group). A good group will look back over its achievements and reflect on each other's contribution, on lessons learnt, on what worked well, and what could have been done better.

This process of reflection can be formalised, and the tutor can help the group debrief itself in a structured way. One process that works quite well is carried out in two stages. First, the group rates itself as a whole against set criteria. The criteria depend on the nature and purpose of the small group work, but might include attendance and punctuality, preparedness for the sessions, adequacy of input into the sessions, and behaviour towards each other within each session. Having established a group rating for each criterion, the members of the group then rate themselves against the group rating. Everyone shares their scores and a discussion ensues. A good group will recognise individuals' strengths and weaknesses, will be constructively critical, and will try to be supportive and encouraging to the more self-effacing members of the group. A good group will also not shirk its responsibility to address issues arising from individuals who are lacking in insight as to their own behaviour. Often, as a result of this kind of caring and inward reflection, change does take place when these members move on to their next groups (see Box 9.3).

The Interprofessional Group

Interprofessional learning is considered important at both the undergraduate and the postgraduate levels (see Chapter 14 for a full discussion) [36]. Health care is all about working in teams, and it is sensible therefore to have regular joint educational training throughout the continuum of learning. The key to interprofessional education is that it should relate to real practice and should not be contrived just so that the box can be ticked for the next accreditation visit. Furthermore, interprofessional education does not require the participation of *all* the professions at each session – only those for whom it is appropriate.

Interprofessional team meetings are common at the postgraduate level, interprofessional training less so, although both acute and primary care trusts provide programmes of training sessions for their staff. Quite a few of these are

BOX 9.3 FOCUS ON: Discussion in small groups

Visschers-Pleijers et al. [34] have carried out an analysis of verbal interactions in tutorial groups. They have subdivided these interactions into five types:

- exploratory questioning – exchanging ideas, critical exchange of ideas
- cumulative reasoning – uncritical accumulation of information
- handling conflicts about knowledge – discussion of contradictory information, arguments and counter-arguments
- procedural interactions – conversations relating to process matters
- off task/irrelevant interactions – general asides, discussion about the weather.

Most of the interactions (about 80%) were learning-oriented, demonstrating the high task involvement of the group and confirming the findings of De Grave et al. [35].

interprofessional, although they tend to be for large numbers, rather than small groups. Training in leadership and management for senior staff is less likely to be profession-specific and often follows the workshop format.

At the undergraduate level, certainly in the UK, there is only a small amount of useful interprofessional education taking place. What there is takes a variety of forms, ranging from a common foundation term with shared PBL, lectures, and anatomy sessions [37], shared clinical skills training [38, 39], a simulated ward environment for junior medical and nursing students [40], and an interprofessional training ward [41–43] to regular joint interprofessional sessions spread over a number of years [44]. Apart from the lectures and the anatomy sessions, most of this involves small group work.

The skills training described [38, 39], involved final-year medical students with newly qualified staff nurses and took place in an interprofessional clinical skills centre. The programme was based around a developing patient scenario, which was pertinent to the participants' area of practice. Each session was led by an experienced nurse lecturer and doctor, supported by specialist contributors. The style of learning was participative, with small interprofessional groups addressing a range of patient management issues. In this way, relevant clinical and communication skills were integrated within the context of holistic patient care. This short exposure to interprofessional learning appears to have been highly effective.

The interprofessional training ward takes small group learning in the context of simulation even further. Here, final-year nurses, physiotherapists, occupational therapists, and medical students look after a small ward [41–43]. They work shifts, carry out all ward duties, work extensively with each other, and learn about each other's professions, while at the same time building up their uniprofessional skills. Handover from one shift to another is key to the learning process. This provides students with an excellent preparation for working on the wards when they graduate, and is again popular with most students.

The issues around interprofessional learning are the same at both undergraduate and postgraduate levels and revolve around group relations (see Bion in the 'Further Reading' section). A strong hierarchical culture pervades all areas of the health service. When people from different professions find themselves being trained together, there is a tendency to adopt these hierarchical roles. Even at the undergraduate level, medical students, physiotherapists, occupational therapists, and nurses all have different entry requirements – and immediately a barrier exists. Of course, one of the aims of interprofessional education is to break down such barriers. Another aim is for the members of one profession to have a clear understanding of the roles of the other health care professionals. Putting these two aims together leads to the third, overarching, aim, which is to learn how to work in teams, thereby leading to a more efficient and effective health service. Small group learning in interprofessional groups is one way to help achieve this last aim.

For an interprofessional group, good facilitation is essential. If the guidance described for any small group work activity is followed, the majority of problems will soon disappear. Setting clear ground rules and ensuring that the group adheres to them is vital – particularly the rule about valuing everyone's input. In addition, the topic or activity needs to be inclusive, that is appropriate for all professions present. The learning needs of the students or trainees must be taken into account, as they will not be the same for all professions. Where they *are* the same, then joint interprofessional learning is more likely to succeed.

Dealing with Difficult Group Members

Dealing with difficult group members is a key role of the tutor running small group teaching sessions. Box 9.4 features a list of common problems encountered, and regular group leaders will recognise them all. It is not the purpose of this chapter to provide suggestions for dealing with each form of aberrant behaviour – for a full discussion of this topic read Tiberius' excellent book *Small Group Teaching: a trouble-shooting guide* [45]. One general point though is worth emphasising; wherever possible, the group should sort out its own problems. This is liable to be much more effective in the long term than the tutor taking control, which often leads to resentment. The tutor's role here is to raise the group's awareness of the issue. One way of doing this is to say, 'Let's take time out for a minute. Is everyone happy the way the group is working? Does anyone want to make a comment about the group process?' Then stand back and watch the sparks fly!

BOX 9.4 Learners who are challenging

- The dominant learner
- The arrogant learner – the know it all
- The learner who wants to be the centre of attention all the time
- The aggressive or argumentative learner
- The offensive and rude learner
- The politically incorrect learner
- The flirtatious learner
- The joke-a-minute learner
- The garrulous learner
- The disengaged learner
- The bored learner
- The learner who relies on everyone else to do the work
- The lazy learner
- The shy learner
- The delicate, tearful learner
- The over-dependent learner
- The constantly late learner
- The frequently ill learner
- The mentally disturbed learner

Of course, this is not always appropriate, nor does it always work. In such circumstances, the group leader has to step in. Sometimes issues have to be addressed in the presence of the group; at other times the leader needs to address the issue outside the group – for example, in the case of a learner who is clearly upset, or one who has failed to respond to everything the tutor has tried in the session. Failure to address group dynamics by one means or another can be very harmful, not just for that particular group, but for future groups. The issue will rarely go away of its own accord and often gets worse and worse if nothing is done about it. No matter how difficult or painful it is, the tutor has a responsibility to sort matters out.

Here are some ideas.

The Dominant Group Member

People can dominate discussions for a number of reasons:

- they simply know a lot
- they think they know a lot – but are frequently wrong
- they like to be the centre of attention and are showing off
- they want to impress the tutor, or their girlfriend/boyfriend
- they enjoy teaching others and are keen to share their knowledge
- they feel that someone has to start the ball rolling, as no one else seems to want to.

If dominating group members are allowed to continue, the group will get really angry, the shy group members will disappear into their shells, and people will either switch off or drop out. Often the dominant member can contribute successfully to a group and does actually have some really useful information to bring. The secret is to reduce their input, while not causing offence.

Sometimes group members realise that they have a tendency to dominate and readily accept being told to keep quiet. Not many have such insight though, and one suggestion for dealing with those who do not is to give them a specific task to do, for example, acting as scribe for the rest of the group. Make sure, however, that they do put up the group's ideas on the whiteboard, and not just their own interpretation. Another technique is to ask them to discuss a particular point and let them have their say, but next time invite someone else to start the ball rolling. Seating can also play a key role. If the tutor sits opposite the dominant student, then that student has the eye of the tutor for the whole session. On the other hand, if the dominant student sits *beside* the tutor, interaction is significantly reduced and the tutor's eye catches other students much more readily. The use of the sweeping hand gesture directed at the dominant student can sometimes be effective – the held-up hand is saying, 'Hang on a minute. I want to hear contributions from other students.' Indeed, sometimes the tutor has to speak these very words aloud. If none of this works, then a word outside the session is necessary, explaining how valuable the individual's contribution is, but how it is important to give *everyone* a chance to be heard.

The Reticent Group Member

The converse of the above is the reticent group member. Again there are many reasons for this, including the following:

- they *are* innately shy
- they are upset or worried about something going on in their lives
- they are in a bad mood
- they are upset with someone in the group
- they have not done any work for the session
- they do not know anything about the topic
- they are completely out of their depth
- they are very tired
- they are very bored
- they have lost their motivation to be a doctor
- they are ill or depressed.

Some of these issues are most definitely not in the group tutor's remit to deal with, particularly the last one. Ill or depressed group members need to be seen by qualified practitioners – GPs, occupational health physicians, or student counsellors. Under no circumstances should the small group learning tutor attempt to address such issues. Other issues may need to be taken up with the learner's personal tutor, trainer, or educational supervisor. The group leader's role is to encourage the individual concerned to seek help from the appropriate person, and to help identify who that person might be.

However, dealing with a genuinely shy, or work-shy, learner is down to the tutor of the small group session. The former needs to be subtly encouraged to join in. Snowballing and buzz groups work quite well, since the learner does not have to speak in front of the whole group. Like the dominant group member, giving the shy individual a task to do, such as being scribe, works well. They are responding to the rest of the group's ideas, a role they are comfortable with. Pointing a finger at them and asking them a direct question is likely to send them deeper into their shell. Humiliating them in any way, for instance by laughing at their answers, is hugely counterproductive. Getting them to comment on something that they are almost certain to know might work, but sometimes the shy person simply needs to be left alone for a while, particularly if the reason for their non-involvement is transitory.

As to the work-shy, usually peer pressure takes care of this, but if not, then a quiet word outside the group is the solution. If a student is really out of his or her depth, this is a more serious situation and needs reporting to the appropriate member of staff responsible for academic progress, so that they can consider whether some form of remediation would help.

The Flirtatious, 'Jokey', or Offensive Group Member

These are definitely examples of situations when the tutor should get the *group* to deal with the issue. One way is to get the group to take time out from the session and ask them if everyone is comfortable with the atmosphere in the group, or with the behaviour of everyone in the group. Usually someone is heartily sick of it and says so. If this does not do the trick, the tutor should speak to the student out of hearing of the group. The tutor should be very firm with them. One thing the tutor must *not* do is collude with the student by smirking or laughing. This is effectively sanctioning the student's behaviour. Stopping it then becomes much harder.

The Late or Absent Group Member

Groups can get really angry with perpetually late or absent group members and usually make their feelings on the matter quite clear. Starting the group later to accommodate the perpetually late learner is not an option. The only possible exception to this is the person who has *genuine* problems getting in for, say, a 9 a.m. start – perhaps for child care reasons or finishing off a ward round. If they really are unable to change their arrangements, then the issue should be taken up with the group right at the beginning, when forming the ground rules.

Persistent absence is a course- or programme-attendance issue, even if this is due to illness. The learner may, as a result, be held back, and the school or postgraduate dean may refuse to sign them off or award credits. These are issues of professionalism, and unprofessional behaviour must be addressed when it first appears, not left to smoulder until it is too late to help learners amend their behaviour.

There is one other essential ground rule, particularly at the undergraduate level: if a person is unable to attend a session for whatever reason – planned or last-minute – they *must* inform the tutor prior to their absence, or as soon after as is possible. If the tutor is unavailable, then they must inform someone in the group. In the era of mobile phones, there really is no excuse for not adhering to this policy.

The Over-dependent Group Member

Here, the tutor has to learn to keep a distance from that individual and must take great care not to overindulge him or her. Some people are good at manipulating vulnerable members of staff for their own ends – another instance when the person concerned should be seated out of the sightline of the tutor.

Exploring Boundaries

Before taking charge of a small group, it is essential to be trained in the art of facilitation. Usually such training will itself take the form of a workshop in which an experienced facilitator demonstrates how to facilitate a small group by example. The facilitator is likely to cover the issues described in this chapter. The training session is likely to include some consideration about handling difficult group members and may also include some consideration about interpersonal boundaries – boundaries for both the students and the staff. Boxes 9.5–9.8 depict four scenarios that illustrate some situations that may arise for group leaders. There are no comments on these scenarios. They are presented for the reader to think about whether the tutor's response was appropriate.

BOX 9.5 Case A

Stephen is tutor to a group in which Tom and Sarah are students. Tom is a likeable but dominant member of the group, who has strong knowledge that he likes to demonstrate to the group. Sarah is a student who comes from an arts background and is not confident about her ability. During a session, Tom begins (as always) to tell the group about his understanding of the learning issues. Stephen (the tutor) stops Tom briefly, and asks everyone if they understand what Tom is describing. The group nods, apart from Sarah, who replies, 'This is all over my head, but don't worry about me. I just don't have the basic knowledge to join in this conversation. I'll catch up one day I'm sure.' Tom continues his feedback, and Sarah immerses herself in a textbook. Stephen is sympathetic to Sarah but does not want the group to get behind. He believes that it is important that the group finishes its work in the time allocated and never allows the session to run over. Anyway, he has a lecture to give immediately after the session.

BOX 9.6 Case B

Hannah is a student in Rebecca's group. Hannah is a lively and strong character, who has said quite openly that she 'wears her heart on her sleeve'. During a tutorial that focuses on lung cancer, Hannah spends a lot of time describing very emotionally how her grandmother died from lung cancer last year. As learning issues are written up on the board, Hannah says that she does not think she will 'be able to do many of these as it's all too close to home for me'. Later in the tutorial, Hannah suddenly begins to cry and says, 'It's all too much for me', and runs out of the tutorial in tears. Rebecca (the tutor) follows her and is out of the room for about 20 minutes.

BOX 9.7 Case C

Claire has been a student in Maria's group for several weeks. She is known to be struggling with the course and is open about a number of personal problems that she says are 'affecting her work'. Maria is concerned about Claire and has become something of a confidante to her. At various points in tutorials, Maria has mouthed the words, 'Are you okay?' to Claire. Maria has offered to talk to Claire 'any time' about her difficulties, and Claire has taken to staying behind after teaching sessions to chat to Maria.

Frequently Asked Questions

How Long should a Small Group Teaching Session Last?

There is no simple answer to this. Everyone's attention span is limited. The key is to vary the activities on a regular basis. This is easy in a workshop format, but in, for instance, a PBL session, students may be incarcerated in a windowless room, engaged in more or less the same activity for up to three hours. In any lengthy session, it is essential to build in a reasonable break for rest and refreshment. In summary, a session should last between 45 minutes and 3 hours, depending on the activity. The less the students are engaged in active learning, the shorter the session has to be.

BOX 9.8 Case D

Mike is a student in Hardeep's group. Mike is an extremely confident and able student, who is a lively and enthusiastic contributor to the group. At the start of the tutorial Hardeep asks the group how their week has been. Most members of the group make comments about how useful they have found the lectures so far and how they have enjoyed their GP visits. Mike adds, looking at Hardeep, 'The content of that Asian guy's session was crap. I could hardly make out one word he said'. Later, when the students are presented with the first page of a new problem, Hardeep asks them to identify its key features. The front cover to the week's problem shows a picture of a young woman in some distress. Mike says, 'Well, she's blonde, about 20 and far too skinny. Not my type at all. She's probably gay, anyway!' The group dissolves into laughter. Hardeep joins in.

Many postgraduate workshops last all day, and sometimes they are spread over two or three days, or even a week – for example, for residential training courses. All-day workshops must be split into manageable chunks, separated by frequent refreshment breaks. Week-long courses should include at least a half day of relaxation.

How Often should Group Membership be Changed?

In undergraduate education, a group should generally remain together for something between a term and a year. It is important for students to learn to mix and establish new teams. Groups that function well never want to change; groups that function poorly cannot wait to change. A term is probably a good compromise.

In postgraduate education, for example, day-release courses in vocational GP training schemes or Balint groups, it is important to maintain as much continuity as possible, and groups should stay together for their natural lifetime (welcoming newcomers as appropriate).

Should Learners who do not get on be Swapped into other Groups?

As far as medical students are concerned, absolutely not – under any circumstances! Doctors have to work in teams for the rest of their lives. These teams may not always function ideally. There may be personality clashes, jealousy, rivalry – but they still have to work as a team. The sooner they learn to find a way to get on with people they do not like, the better they will function when it really matters.

In postgraduate education, doctors know their own needs (unfortunately, not always) and are free to attend whatever courses they feel they will gain the most from. Continuing professional development is a requirement for all health care professionals, but apart from essential updating of skills (for example, resuscitation and life-support courses), few courses are compulsory and most are one-offs. Apart from GPs and psychiatrists, doctors therefore do not tend to attend regular workshops with the same people, although they do meet together as teams all the time. Instead, doctors tend to create their own 'clubs' and meet up frequently at conferences. Rarely, these national or international meetings involve small group work, but they consist mainly of hundreds of usually fairly dire presentations.

Do I Need to be a Subject Expert to Run a Small Group?

It depends. For PBL, there is some evidence that the best tutors are those who are expert facilitators, but that it helps to have knowledge expertise as well [46, 47]. Knowledge experts who have no skills of facilitation and who turn each PBL session into a question-and-answer session, or worse still, a mini-lecture, make poor tutors, as do those who do not even have the excuse of being knowledge experts and who are unable to grasp the principles of PBL (and indeed scorn that whole approach to learning). But for other small group activities, such as a seminar or workshop, and for most postgraduate teaching, where the tutor has more input, knowledge expertise is indeed necessary, again alongside the skills of small group facilitation.

Can a Group Survive with a Series of Different Facilitators?

Generally, continuity of group leader is important, although rarely is an undergraduate tutor available to take every session in a term because of other commitments. One solution to this is to have paired tutors, so that when one is absent, the other can take over.

In the case of, say, a series of small group sessions covering a range of specialist topics, such as in day-release workshops for junior doctors, it is actually desirable to have different tutors because specialist medical expertise is required. The problem then is that the group never really has the time to develop a meaningful relationship with each tutor, because they are with them for such a short time. The group may get beyond the forming stage, but the tutor rarely does. There are two solutions to this: reduce the number of guest tutors or maintain continuity through the use of a regular co-facilitator.

Is There such a Thing as a Floating Facilitator?

Yes, that is exactly what a workshop facilitator does. But the question is more to do with a single facilitator looking after more than one group simultaneously, where the groups are not located in the same space but, for example, in nearby rooms. This is rather like the consultant surgeon keeping an eye on his trainees in adjacent operating theatres – common enough practice, but is it *good* practice? Usually this kind of multi-group facilitation only happens because of tutor shortages, and is aimed at making economies. The economies are really perceived economies rather than actual economies, because they only take into account savings in staff time and not any reduction in the quality of student learning. The learning experience is unlikely to be beneficial to the student (although sharing one good tutor is undoubtedly better than being tutored by two separate bad tutors). For multi-group facilitation to work, the activities being undertaken by each group must be task based, with the tutor popping in now and again to check on group progress. Again, in the postgraduate arena, where the learners are often more committed and motivated, groups can be left to get on with the task by themselves for a while, which leads in nicely to the next frequently asked question.

Do Groups Need a Facilitator at all?

Again, the answer is that it depends. Students in the early stages of their course undoubtedly need a facilitator around. Partly, this is simply because they do not know what to do, and partly, for reassurance that they are learning what they are supposed to be learning. 'Will it come up in the exam?' is a question frequently asked of tutors. As students develop and become more mature, the need for a tutor is reduced. Such students can facilitate their own groups. Once they are let loose on the wards, students revert to Stage One again and need a bit of mollycoddling and guidance.

Postgraduate groups can survive without a tutor if the purpose of the small group session is clear. They may then appoint a chair from within the group and get on with the task in hand.

Conclusion

Learning in small groups can be very productive, if at times challenging. Not everyone enjoys the process, particularly if the group is dysfunctional. Here, the skills of the facilitator are put to the greatest test, although there are guidelines for handling difficult group members, which are touched on here and elsewhere. A poor facilitator can jeopardise the success of even the best of groups, through bullying, humiliating, patronising, prejudicial, or over-didactic behaviour. A good facilitator can help the group to achieve a high standard of learning, by encouraging active learning and reflective thinking, by questioning and challenging, and by setting a good example. The most successful groups are tutored by people who are good at facilitation, and not necessarily those who have subject expertise.

Small group learning is particularly successful if the facilitator adopts a variety of techniques during sessions. Workshops, which often last a morning or afternoon, must be split up into a range of activities – discussion, role play, debate, exercises using triggers, watching videos, practising skills, observing demonstrations, problem solving, question-and-answer sessions, presentations – the list is endless.

Facilitation expertise does not grow on trees. People who are brilliant teachers do not necessarily make good facilitators, because they may find it hard to get out of the information-giving mode in which they excel. Group facilitation should never be forced on such people. It is more sensible to recognise their skills and make use of them in the more didactic elements of a course, leaving facilitation to those who are good at it. Would that life were that simple …

References

1 Jolly, B.C. (1994). *Bedside Manners: Teaching in the Hospital Setting*. Maastricht: Universitaire Pers Maastricht.
2 Homepage for a Virtual Ward Round. www.jround.co.uk/virtual_ward_round/d1_bedview.html (accessed 27 June 2018).
3 De Villiers, M., Bresick, G. and Mash, B. (2003). The value of small group learning: an evaluation of an innovative CPD programme for primary care medical practitioners. *Medical Education* 37: 815–21.
4 Steinert, Y. (2004). Student perceptions of effective small group teaching. *Medical Education* 38: 286–293.
5 Tiwari, A., Lai, P., So, M., and Yuen, K. (2006). A comparison of the effects of problem-based learning and lecturing on the development of students' critical thinking. *Medical Education* 40: 547–554.
6 Facione, N.C., Facione, P.A., and Sanchez, C.A. (1994). Critical thinking disposition as a measure of competent clinical judgement: the development of the Californian Critical Thinking Disposition Inventory. *Journal of Nursing Education* 33: 345–350.
7 Barrows, H.S. (1986). A taxonomy of problem-based learning methods. *Medical Education* 20: 481–486.
8 Engel, C.E. (1997). Not just a method but a way of learning. In: *The Challenge of Problem-Based Learning*, 2e (ed. D. Boud and G. Feletti), 17–27. London: Kogan Page.
9 Kamin, C.S., O'Sullivan, P.S., Younger, M., and Deterding, R. (2001). Measuring critical thinking in problem-based learning discourse. *Teaching and Learning in Medicine* 13: 27–35.
10 Peters, T. (1995). *In Search of Excellence: Lessons from America's Best-Run Companies*. London: Harper Collins Business.
11 Rudduck, J. (1979). Learning to Teach Through Discussion. Centre for Applied Research in Education occasional publications no. 8. University of East Anglia, Norwich.
12 Richmond, D.E. (1984). Improving the effectiveness of small-group learning with strategic intervention. *Medical Teacher* 6: 138–145.
13 Brown, G. (1982). How to improve small group teaching in medicine. In: *The Medical Teacher* (ed. K.R. Cox and C.E. Ewan), 70–78. Edinburgh: Churchill Livingstone.
14 Neville, A.J. and Norman, G.R. (2007). PBL in the undergraduate MD program at McMaster University: three iterations in three decades. *Academic Medicine* 82 (4): 370–374.
15 Barrows, H.S. (1985). *How to Design a Problem-Based Curriculum for the Preclinical Years*. New York: Springer.
16 Wood, D.F. (2003). ABC of learning and teaching in medicine: problem based learning. *British Medical Journal* 326: 328–330.
17 Dolmans, D.H.J.M., De Grave, W., Wolfhagen, I.H.A.P. and van der Vleuten, C.P.M. (2005). Problem-based learning: future challenges for educational practice and research. *Medical Education* 39: 732–41.
18 Albanese, M.A. and Mitchell, S. (1993). Problem-based learning: a review of literature on its outcomes and implementation issues. *Academic Medicine* 68: 52–81.
19 Poulton, T., Conradi, E., Kavia, S. et al. (2009). The replacement of 'paper' cases by interactive online virtual patients in problem-based learning. *Medical Teacher* 31: 752–758.
20 Macallan, D.C., Kent, A., Holmes, S.C. et al. (2009). A model of clinical problem-based learning for clinical attachments in medicine. *Medical Education* 43: 799–807.
21 Parmelee, D., Michaelsen, L., Cook, S., and Hudes, P. (2012). Team-based learning: a practical guide: AMEE guide no. 65. *Medical Teacher* 34 (5): e275–e287.
22 Greenhalgh, T. and Hurwitz, B. (1999). Why study narrative? *British Medical Journal* 318 (7175): 48–59.
23 Harden, R.M., Uudlaw, J.M., Ker, J.S., and Mitchell, H.E. (1996 Jan 1). AMEE medical education guide no. 7.: task-based learning: an educational strategy for undergraduate postgraduate and continuing medical education, part 1. *Medical Teacher* 18 (1): 7–13.
24 Peyton, J.W.R. (1998). *Teaching and Learning in Medical Practice*. Rickmansworth: Manticore Europe.
25 Gill, D. and Dacre, J. (2003). Teaching and Learning 'At the Bedside'. https://faculty.londondeanery.ac.uk/e-learning/explore-further/teaching_and_learning_at_the_bedside.pdf (accessed 22 June 2018).
26 Miller, S.D.W., Butler, M.W., Meagher, F. et al. (2007). Team Objective Structured Bedside Assessment (TOSBA): a novel and feasible way

of providing formative teaching and assessment. *Medical Teacher* 29: 156–159.
27 Balint, M. (1957). *The Doctor, His Patient and the Illness*. Edinburgh: Churchill Livingstone [1964, reprinted 1986].
28 Balint, E. and Norell, J. ed. (1983). *Six Minutes for the Patient: Interaction in General Practice Consultations*. London: Tavistock Publications.
29 Kolb, D.A. (1984). *Experiential Learning: Experience as the Source of Learning and Development*. Englewood Cliffs, NJ: Prentice Hall.
30 Revans, R. (1982). *The Origins and Growth of Action Learning*. Bromley: Chartwell Bratt.
31 Tuckman, B.W. (1965). Development sequence in small groups. *Psychological Bulletin* 63: 384–399.
32 Walton, H.J. (1997). *Small Group Methods in Medical Teaching*. ASME medical education booklet no. 1. Edinburgh: Association for the Study of Medical Education.
33 Mulholland, H. (1994). Teaching small groups: facilitating learning. *Hospital Update* 20: 382–384.
34 Visschers-Pleijers, A.J.S.F., Dolmans, D.H.J., de Lang, B.A. et al. (2006). Analysis of verbal interactions in tutorial groups: a process study. *Medical Education* 40: 129–137.
35 De Grave, W.S., Boshuizen, H.P.A. and Schmidt, H.G. (1996). Problem-based learning: cognitive and metacognitive processes during problem analysis. *Instructional Science* 24: 321–41.
36 Molyneaux, J. (2001). Interprofessional teamworking: what makes teams work well? *Journal of Interprofessional Care* 15: 29–35.
37 Mitchell, B.S., McCrorie, P., and Sedgwick, P. (2004). Student attitudes towards anatomy teaching and learning in a multiprofessional context. *Medical Education* 38: 737–748.
38 Freeth, D. and Nicol, M. (1998). Learning clinical skills: an interprofessional approach. *Nurse Education Today* 18: 455–461.
39 Tucker, K., Wakefield, A., Boggis, C. et al. (2003). Learning together: clinical skills teaching for medical and nursing students. *Medical Education* 37: 630–637.
40 Ker, J., Mole, L., and Bradley, P. (2003). Early introduction to interprofessional learning: a simulated ward environment. *Medical Education* 37: 248–255.
41 Areskog, N. (1994). Multiprofessional education at the undergraduate level – the Linköping model. *Journal of Interprofessional Care* 8: 279–282.
42 Reeves, S., Freeth, D., McCrorie, P., and Perry, D. (2002). 'It teaches you what to expect in future…': interprofessional learning on a training ward for medical, nursing, occupational therapy and physiotherapy students. *Medical Education* 36: 337–344.
43 Kirk, A., Parish, K., and Buckingham, I. (2006). Interprofessional education – the benefits and drawbacks of an interprofessional training ward. Paper presented at the 12th International Ottawa Conference on Clinical Competence, New York, 20–24 May.
44 O'Halloran, C., Hean, S., Humphris, D., and Macleod-Clark, J. (2006). Developing common learning: the New Generation Project undergraduate curriculum model. *Journal of Interprofessional Care* 20: 12–28.
45 Tiberius, R.G. (1990). *Small Group Teaching: A Trouble-Shooting Guide*. Toronto: Ontario Institute for Studies in Education Press.
46 Barrows, H.S. (1988). *The Tutorial Process*, 2e. Springfield, IL: Southern Illinois University School of Medicine.
47 Schmidt, H.G., van der Arend, A., Moust, J.H.C. et al. (1993). Influence of tutors' subject-matter expertise on student effort and achievement in problem based learning. *Academic Medicine* 68: 784–791.

Further Reading

Bion, W.R. (1952). Group dynamics: a review. *The International Journal of Psycho-Analysis* 33: 235–247. [Reprinted in: M. Klein, P. Heimann, and R. Money-Kyrle ed. (1955) *New Directions in Psychoanalysis*, pp. 440–77. Tavistock Publications, London; Bion (1961), pp. 141–91].

Bion, W.R. (1961). *Experiences in Groups*. London: Tavistock Publications.

Campkin, M. (1986). Is there a place for Balint in vocational training? *Journal of the Association of Course Organizers* 1: 100–104.

Dennick, R. and Exley, K. (2004). *Small Group Teaching: Tutorials, Seminars and Beyond*. Abingdon: Routledge.

Edmunds, S. and Brown, G. (2010). Effective small group learning: AMEE Guide No. 48. *Medical Teacher* 32 (9): 715–726.

Habeshaw, S., Habeshaw, T., and Gibbs, G. (1988). *53 Interesting Things to Do in Your Seminars and Tutorials*. Bristol: Technical & Educational Services.

Jacques, D. (2000). *Learning in Groups: A Handbook for Improving Group Work*, 3e. London: Kogan Page.

Jacques, D. (2003). Teaching small groups. *British Medical Journal* 326: 492–494.

Lee, A. and Higgs, J. (1988). How to help students learn in small groups. In: *The Medical Teacher*, 2e (ed. K.R. Cox and C.E. Ewan), 37–47. Edinburgh: Churchill Livingstone.

McGill, I. and Beaty, L. (1995). *Action Learning, a Guide for Professional, Management & Educational Development*, 2e. London: Kogan Page.

Newble, D. and Cannon, R. (2001). Chapter 3: teaching in small groups. In: *A Handbook for Medical Teachers*, 4e, 39–54. Dordrecht: Kluwer Academic Publishers.

Norman, G.R. and Schmidt, H.G. (2000). Effectiveness of problem-based learning curricula: theory, practice and paper darts. *Medical Education* 34: 721–728.

Weinstein, K. (1995). *Action Learning: A Journey in Discovery and Development*. London: Harper Collins.

Westberg, J. and Jason, H. (1996). *Fostering Learning in Small Groups*. New York: Springer Publishing Company.

10 Technology-enhanced Learning

Rachel H. Ellaway
Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada

KEY MESSAGES

- It is the use of technology rather than the technology itself that should be appraised in the context of medical education.
- Technologies do not teach or learn they mediate and augment teaching and learning.
- Learner and teacher technology preferences and abilities often vary significantly.
- Technologies are often used for economic and convenience reasons rather than for their instructional qualities.
- Technology use introduces many risks and challenges to medical education as well as many advantages and opportunities.
- Technology use intersects with almost every other aspect of medical education and it should be considered in terms of these intersections.

Introduction

While there are many technologies (such as textbooks, models, and illustrations) that have a long pedigree in medical education, the contemporary use of the term 'technology' is generally understood as referring to digital tools and systems. Indeed, there has been a proliferation of personal digital devices in recent years, to the point that (digital) technologies have become a near-ubiquitous presence in the training of tomorrow's doctors. I invite readers to consider this chapter in the context of *all* technology use in medical education, not just digital technologies. To fully address the multitude of intersections between technology use and medical education practices is beyond the scope of a single chapter so I would also encourage readers to consider technology use as it is reflected in the material presented throughout this volume.

I use the term 'technology-enhanced learning' (TEL) as my main frame of reference rather than the more popular but rather problematic concept of 'e-learning'. Central to the concept of TEL is a focus on the mediating role of technology. Teachers teach and learners learn; it is *how* they do these things (and what follows) that is changed by their use of technology. In exploring this topic, I will consider a range of theoretical frames for appraising the use of technology in medical education, and from these I will set out a series of techniques for developing, using, and evaluating technology-enhanced learning in medical education. My goal in doing so is to provide a critical review of the positions, practices, opportunities, and challenges associated with using technology in contemporary medical education.

Technology Use in Medical Education

There are a great many technologies being used in contemporary medical education, including devices (such as laptop computers, tablets, and mobile devices) and their peripherals (such as printers, cameras, and keyboards), software (such as the word processing and illustration tools that I used to write this chapter) and apps (for mobile devices and tablets), services (such as those offered by Google and Skype), digital content (such as that offered by YouTube and Wikipedia), and social media (such as Twitter and Facebook). While some of these technologies are provided by institutions of medical education (such as learning management systems), many are provided by learners (such as laptops and mobile devices), or by third parties with little or no direct focus on medical education (such as Google and Wikipedia). It can therefore be challenging to discuss technology as something specific to medical education or medical educators, and difficult to appraise or direct given its diverse provenance. Although a student or faculty laptop or mobile device can be used for teaching and learning purposes, it is also used for many other purposes and in many other contexts. It is important to think about this intersectionality in the context of medical education, exemplified for instance by teacher and learner use of social media where the social and instructional benefits can become rather blurred.

We tend to pay more attention to the technologies with which we interact directly and often oblivious to the critical role of the many supporting technologies they depend on. Essential if unexciting factors such as availability of networks and electrical power, the provision of security and sign-in services, and the ability to host systems in a robust

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

way are necessary to provide a foundation for the use of technology in medical education. Medical schools that do not have these foundational technologies readily available (for instance because of cost, conflict, or natural disasters) tend to make more tentative use of technologies. Drawing loosely from Maslow, the dependence of medical teachers and learners on their technologies can be understood in terms of a 'hierarchy of needs' (Figure 10.1) [1]. The key lesson here is that the use of technology in medical education can be compromised or undermined if its dependent factors are uncertain or absent. Not only does this include functional and safety factors, it also depends on the legitimacy and acceptability of using technology, (or particular uses of technology such as students having smartphones at the bedside) in medical education.

It is also important to note the division of labour regarding the use of technology in medical education. Some technologies are used by both learners and teachers but in different ways. For instance, online portfolios and learning management systems present different tools and views depending on the user role. Other technologies tend to be more or less learner, or teacher, specific. For instance, teachers may make extensive use of tools like PowerPoint, while learners tend to make much greater use of reference materials.

Drawing these preliminary observations together, it becomes clear that we cannot say that technology use is, or is not, effective in any absolute way in the context of medical education. Its utility depends on the task, the technology, the individual, and their circumstances. Given the plurality and diversity of technologies and their uses in medical education it may therefore seem like an overwhelming task to guide practice in both a succinct and generalisable way. However, there are several theoretical frames we can draw on to set out an inclusive model of issues in and around technology use in medical education.

Technologies for Learning

Medical education can be understood as sequences of interconnected activities (e.g. lectures, skills labs, problem-based learning, exams, bedside learning, etc.). We can analyse these activities using concepts from activity theory [2], a central tenet of which is that activities are *mediated*; they both depend on mediating artefacts and are shaped by them. For example, whereas 10 or more years ago it was normal practice for many medical schools to provide their learners with paper handouts for lectures, labs, and other learning sessions, it is now much more common for notes to be provided electronically or even that there are no notes, simply access to the teacher's PowerPoint slides. The once ubiquitous ring binder full of annotated handouts and paper notes has been largely replaced by electronic files and documents. We can say, therefore, that computers and the various tools they support now mediate learners' creation and use of their personal knowledgebase.

Rather than listing all activities in medical education and describing specific mediating technologies within each, we can instead consider different types of activity in terms of their functional role in medical education and different mediating technologies associated with them. See Box 10.1.

Clearly there are many activity types in medical education and many technologies that can be used to mediate them. Moreover, some technologies can be used to mediate more than one kind of activity (such as learning management systems), and any given session (class, rotation, lab, etc.) may employ many activity types and be mediated by many technologies. We should not therefore be concerned solely with single instants of activity *plus* mediating technology, we should also consider the entangled clusters of activities and mediating technologies that are the norm in medical education. This leads us to the concept of 'blended learning' where technology mediation is entangled with other kinds of mediation. For instance, using digital technologies to look things up or create materials within a face-to-face class setting. We should be clear that it is not the learning that is blended, rather it is the use of different mediating components in support of learning that is blended. There is evidence to suggest that blended uses of TEL rather than isolated uses of TEL lead to better learning outcomes than using TEL on its own [4].

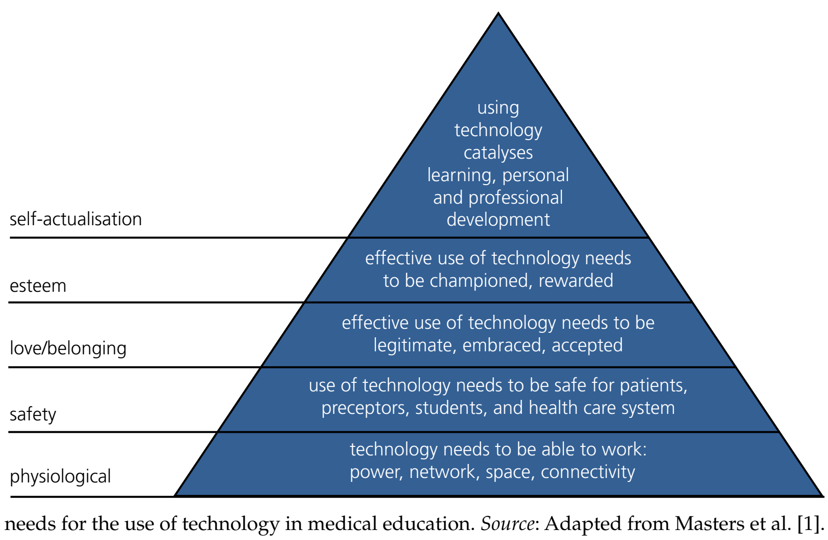


Figure 10.1 A hierarchy of needs for the use of technology in medical education. *Source*: Adapted from Masters et al. [1].

BOX 10.1 Learning activity types and associated mediating technologies

	Example activities	Examples of mediating technologies
Content activities	Read, explore, reflect, remember, synthesise	Knowledgebases such as Wikipedia, eBooks, note taking and mind mapping tools
Absorb-type activities	Presentations, storytelling, readings, tours	PowerPoint, YouTube, podcasts, vodcasts
Do-type activities	Practice, discovery, games, simulations	Onscreen simulations, virtual patients, games, e-mannequins, virtual task trainers
Connect-type activities	Ponder, job aids, research, original work	eHealth systems and tools, bibliographic databases, decision support tools
Social activities	Discuss, collaborate, observe others, reflect on own ability	Discussion boards, Twitter, blogs, wikis, social media
Test activities	Challenge, perform, assess, evaluate, provide feedback	Quizzes, test banks, summative games, simulations
Collection activities	Logging, tracking, storing, organising, reporting	Encounter logging, portfolios, CVs, analytics, professional development management systems

Source: Adapted from Horton [3].

Although devices are usually designed to be used by one person at a time, technology use is enacted in a social context; it is seen and interpreted by others according to circumstance. Learner and practitioner uses of technologies can therefore be seen as professional behaviours. Indeed, learner use of computers, tablets, and smartphones has been a focus of professionalism concerns reflected in debates over whether learners should be allowed or encouraged to use their devices in lectures or at the bedside. While the use of technology in lectures is now commonplace, learners' use of digital technologies at the bedside is still a matter of contention. The intersections between TEL and professionalism can be considered as a matter of 'digital professionalism' [5]. This is not just about *where* technologies should and should not be used, it also focuses on *how* technologies are used and how the use of technologies can blur personal and professional aspects of individuals' lives. A key example of this is learner and practitioner use of social media. Tools such as Facebook and Twitter afford an unparalleled ability to communicate with others, but they also create the means by which individuals can misstep by making aspects of their lives public, particularly as professionals, that should have been kept private. Whether concerned with disclosing confidential information, sharing images of oneself or others in potentially compromising situations, or sharing unprofessional opinions and comments, the ability to use digital media in a professional way has become part of the training of tomorrow's doctors. See Box 10.2.

Digital professionalism is not the only intersection of technology, medical education, and practice that we need to consider. Digital technologies such as electronic health records, picture archiving and communication systems (PACS), formularies, and order entry systems are increasingly mediating the practice of medicine, which means that preparing learners to practice in a digital environment is a growing issue for medical education. These technologies can mediate learning as well as clinical practice, and to that end, have called them 'medium-as-message technologies'. [6] The implication for medical education is that

BOX 10.2 FOCUS ON: Digital professionalism

Health professional learners, teachers, and practitioners are exposed to new opportunities and risks through their use of digital technologies. Opportunities include the ability to form and sustain professional networks and the ability to access and share information on demand. Risks include a blurring of personal and professional boundaries and a greater exposure to threats to professional reputations. The institutional response has often been to proscribe or limit the use of technologies, which limits the benefits of technology use as well as the risks. Digital professionalism builds on the principles of medical professionalism to frame technology use in a more equable light:

> Digital media are not an intrinsic threat to medical professionalism. Professionals should use digital media for positive purposes in ways that support principles of patient care, compassion, altruism, and trustworthiness. Professionals should be aware of the shaping nature of their relationships with digital media and they should maintain the capacity for deliberate, ethical, and accountable practice. [5]

Digital professionalism can be taught, modelled, and assessed by attending to:

- Proficiency: the ability to use technologies effectively, safely, and responsibly.
- Reputation: the ability to maintain one's professional reputation both through their own actions and in responding to the actions of others.
- Responsibility: developing and sustaining professionally sound ways of using technologies, and teaching, modelling, and encouraging this in others.

developing skills in the use of these technologies needs to be combined with their use for mediating learning. To be clear, this is more than developing operator skills, but rather about understanding the principles of using digital tools and information within clinical activities and workflows. See Box 10.3.

Before closing this section, it is important to add a word of caution regarding the widespread assumption that contemporary learners are all keen on using technology in their studies. Terms such as the 'Net Generation' and 'digital natives' have been coined to allude to some intrinsic change in the nature of learners and learning in the digital age [8]. There is, however, little evidence that there are significant differences in the learners themselves but rather in the opportunities and risks afforded by the changing range of technologies that they and others, such as their teachers, use. The reality is that, in any given cohort or class, there will be those who are interested in using and exploring technologies for learning, those who would rather use traditional media such as print and handwritten resources, and many who like some technologies but prefer not to use others, or those who are indifferent to the media with which they are using to learn. TEL practices need to encompass the diversity of learner attitudes to technology as well as the diverse uses they may make of different technologies.

BOX 10.3 eHealth competences [7]

The following are a selection of eHealth competences for postgraduate trainees along with suggested learning activities to facilitate the development of those competences.

Competence	Learning activity
Uses information and communication technology (ICT) to provide patient-centred care, and to monitor and audit practice	Teaching using a student electronic health record
Communicates effectively using ICT	Learners practice communication skills using a range of media
Collaborates in using eHealth systems and techniques with other health care professionals	Interprofessional learning mediated by using a student electronic health record
Advocates for balance between privacy and the needs of health care system	Problem-based learning involving technology use in health care
Speaks out against harmful medical misinformation	Critical appraisal and redesign of patient information sources
Uses ICT to enhance knowledge, skill, and judgement in the provision of evidence-informed patient care	Decision support and digital professionalism

Technologies for Teaching

In the same way that learners are still the ones doing the learning, teachers still teach. Using technologies in medical education is not just about mediating learning, it is also about mediating teaching. For example, preparing a lecture now means building a PowerPoint presentation, giving a lecture usually means talking to or around one's PowerPoint slides, lecture theatres are typically configured around the screens and data projectors needed to show PowerPoint presentations, and learners follow up by downloading and reviewing their lecturers' PowerPoint slides. We can say therefore that the use of PowerPoint (and its associated technologies) mediates many of the activities associated with the contemporary lecture.

Different teaching activities may be mediated in different ways. We can build on the activity types outlined in Box 10.1 to consider teaching activity types and their associated mediating technologies. See Box 10.4.

Although technologies do not do the learning for learners, some technologies mediate how teachers teach, for instance using a multimedia teaching resource or a YouTube video to learn how to execute a clinical procedure rather than being instructed by a human teacher. This is not to say that there is no teacher present in TEL, only that the teacher is mediated by the use of TEL. We can therefore consider the concept of teaching presence in TEL. We can consider a continuum of teacher presence from minimally mediated to fully mediated as follows:

- Minimal mediation: digital technologies augment embodied interactions with teachers. For instance, learners in a problem-based learning group may research and reference materials online to contribute to their facilitated in-person discussions.
- Synchronous mediation: the teacher is not co-present with their learners but interacts with them using one or more technologies. For instance, a teacher presenting to a group in a webinar or massive open online course (MOOC) with all participants interacting through their computers. A range of technologies, such as teleconferencing, web-conferencing, and virtual worlds, can all afford variations on synchronous mediation.
- Asynchronous mediation: the teacher interacts with their learners but not in real time. For instance, a teacher uses a discussion board or a shared blog to discuss conceptual issues in professional development. Technologies such as discussion boards, email, social media, wikis, and blogs can all afford variations on asynchronous mediation.
- Full mediation: interactions between a teacher and their learners are replaced by interactions between technologies and learners; for example learners view videos, use multimedia teaching packages, or even online

BOX 10.4 Teaching activities and associated mediating technologies

	Example activities	Example mediating technologies
Content activities	Presentations, lecturing, instructing, directing, providing course content, reference materials	PowerPoint and Prezi, video and YouTube, audio and iTunes U, eBooks, learning management systems
Guiding activities	Demonstrations, directions, exemplars	YouTube, email, learning management systems
Discussion activities	Seminars, problem solving, learning issues support	Discussion boards, email, social media, learning management systems
Tracking and feedback activities	Monitoring learner progress, providing targeted feedback on progress and need for corrective actions	Email, learning management systems, evaluation systems such as One45
Assessing activities	Setting and marking assignments, exams, coursework	Assessment management systems

textbooks. In this case the teacher (or at least a teacher) has recorded their instruction in the design of the materials the learners use. Given that the learners cannot interact directly with a teacher, the instructional utility of these materials is in part dependent on the instructional interactivity built in to these technologies. Interactivity in the context of TEL is not so much a matter of having things to interact with but rather how learners can interact with a mediated teaching presence.

Some forms of technology used for teaching can be deployed fairly spontaneously (such as looking up facts using third-party online reference materials), but most need careful design and configuration to be effective. 'Design' then is a key issue in using technologies for teaching. There are many different aspects of technology use for teaching that need to be designed: content (presentations, course materials), activities (simulations, scenarios, labs), communication (discussion, collaboration), and training and support (guides and manuals). Although medical teachers may undertake these design tasks themselves, schools often engage instructional designers as specialists to support teachers in their design activities. There is a rich literature on instructional design [9, 10] to support teachers in using technology in their teaching. For instance, Mayer's 'multimedia principles' sets out evidence-based guidance for designing technology-mediated teaching materials (Box 10.5) [11]. Other relevant educational principles informing instructional design are discussed elsewhere in this book (e.g. Chapter 6).

Although there are robust deductive principles that can be used to inform the design of technology-mediated teaching activities, much of the design and use of technology in support of teaching is shaped by the practical constraints that teachers face, such as what technologies are available to them. For instance, it is common for universities and medical schools to have single institutional e-learning platforms, such as Blackboard or Moodle, which instructors are required to use. Even for schools with a more flexible posture, the costs and other risks for a programme supporting multiple systems will likely to prove unsustainable. Whether or not they are working with instructional designers, medical teachers need to know what technologies can and should be used in their teaching, what technologies are available to them, what technologies would work best in the particular situation envisaged, and how the technology should be configured and presented [12]. They should also appreciate the local culture of technology use (what technology use is and is not acceptable or popular, etc.), and the capabilities of the participants (how experienced they are, what tools and devices they have available to them, etc.).

Clearly the use of technologies for teaching in medicine can be a rich and at times complex undertaking, and this

BOX 10.5 HOW TO: Design for technology-enhanced learning

Mayer and colleagues have identified a number of robust evidence-based principles that should inform the design and organisation of effective TEL materials [11]. Some of these key principles are as follows:

Multimedia principle	Words and pictures are better than words alone.
Redundancy principle	Students learn better from animation and narration than from animation, narration, *and* on-screen text.
Coherence principle	Students learn better when extraneous material is excluded.
Pre-training principle	Learn key concepts before tackling a complex aggregate of concepts.
Segmenting principle	User-paced segments are more effective than a single bolus of instruction.
Signalling principle	Add cues to the organisation of material.
Personalisation principle	Conversational style is better than a formal style.

complexity may be intriguing to some teachers and off-putting to others. As with learners, for any given group of teachers there are likely to be those who are keen to use different technologies, those who do not like using technologies of any kind, and those who like using some technologies but not others. Matching teacher and learner preferences as regards technology use can be difficult to negotiate and some flexibility in how much technology use is involved in any given medical education activity is advisable.

Technologies for Assessment

Having considered technologies for learning and teaching, we turn now to technologies for assessment. As with assessment in general, uses of technologies from a learner point of view tend to fall into formative and summative.

Formative uses of technology tend to focus on online tests, quizzes, and other ways of gauging current knowledge and identifying gaps or weaknesses. These may be standalone tests (such as exam preparation Apps for smartphones) or they may be embedded in multimedia teaching materials. Either way, although technologies allow for the presentation, automation, and scoring of tests, the value of a question or test still depends on teachers' skills in writing good assessment materials: technologies can mediate formative assessments but they do not in and of themselves make for better assessment.

Uses of technology for summative assessment purposes include presenting exams online (rather than in a paper-based format), and reporting on continuing medical education activities (such as using some kind of e-portfolio system).

Assessment technologies themselves may be designed to support either formative or summative processes but many of them can be used for both purposes. A question is a question, a test is a test, and a portfolio is a portfolio: it is the intent and the procedural rigour involved that makes the difference. Technologies for summative assessment purposes, particularly for high stakes assessment activities, need higher levels of access control and security, and greater loading capacity: a proctored (supervised) computer-mediated high stakes examination must be secure against cheating and misrepresentation and be able to handle simultaneous mass system access.

As much as technologies can be useful in mediating formative and summative assessment activities, technology use in support of assessment has arguably had the greatest uptake and the greatest impact on the logistics and workflows associated with managing and deploying assessments, and analysing assessment data. For instance, the use of online systems for building and managing assessments (such as ExamSoft and Respondus) is increasingly common in medical schools around the world, not least because programmes have to manage and evaluate large quantities of questions. The question banks that databases in these systems offer, from which exams can be assembled and refined before delivery, track previous uses of questions and their performance metrics to support exam designers' in assembling exams. These systems can often also process and report on learner performance in exams much faster than paper-based systems can. However, there are also limitations to the ability of these systems to support all forms of assessment. For instance, these systems tend to be limited in their ability to deal with written answers (such as essays or long-answer test questions), their strength is in handling tabulated or structured responses such as multiple-choice questions or response grids (such as assessor forms in OSCEs). Thus, while technologies can play a role in mediating and supporting assessment processes, the core skills and roles of faculty and others involved in assessment are still an essential part of delivering robust assessments in support of medical education.

Mediation of assessment can also introduce artefacts that can compromise the results of testing learners. For instance, if a test favours learners with more experience using technology (such as being able to use a computer or to manipulate onscreen objects) then the test will to some extent measure and report on this ability. Given that complete normalisation of learner attitudes and capabilities in using technology is unrealistic, technology-mediated tests and assessments should be designed to minimalise the intrusion of artefacts arising from technology-mediation in to assessment processes and data.

Principles of effective assessment design are described in Chapter 20 and issues concerning e-portfolios considered in Chapter 18, but in relation to the general use of technology-enabled assessment (TEA), Amin et al. [13] provide helpful guidance (Box 10.6).

Technologies for Managing Medical Education

Building on the observation that much of the value of assessment technologies is to be found in the management

BOX 10.6 HOW TO: Use technology-enabled assessment

Guidelines for the use of technology-enabled assessment (TEA) [13].

- Despite the novelty of TEA, it is still dependent on robust general assessment principles.
- Schools should only use TEA where it confers a clear advantage.
- Schools should appraise the strengths, weaknesses, opportunities, and threats of using TEA, particularly for high-stakes purposes.
- Technologies for TEA should be integrated within programme-level assessment systems and practices.
- TEA requires particular technical and assessment expertise, which should be built and sustained within the programme.
- Test developers should ensure the validity of TEA, in particular the constructs being measured.

of assessment processes, the last major area of application for technologies in medical education we will consider is their use in support of systems of medical education (programmes, courses, etc.), typically in the role of managing, tracking, and reporting on activities within those systems. As with any other information management system, their value often depends on economies of scale. While tracking the progress of a few learners over a short period of time may be better done through direct interaction and observation, tracking the progress of hundreds of learners over multi-year programmes of study requires robust information tracking and management systems. It is now common practice for programmes of study to make use of learning management systems, also sometimes called Virtual Learning Environments (VLEs), as an online support system. These systems, both generic (e.g. Moodle and Blackboard) and medicine specific (e.g. Entrada), provide an integrated array of tools for learners, teachers, and programme managers. It is the latter group who may benefit the most from using these systems as these systems usually provide tools for monitoring, tracking, recording, and reporting on what learners and teachers have done in the system. Activity logging is therefore a key function of these tools. Some activity logs are generated automatically (such as logins, downloading files, or contributing to discussion boards) while others are more explicitly about tracking and reporting on activity (such as clinical encounter tracking).

Log data has no intrinsic value in and of itself, it needs to be analysed and conclusions and inferences drawn from this analysis for it to become useful. To this end, learning analytics is a growing area of interest in higher education and in medical education. For instance, the adoption of competency-based medical education (CBME) is in part based on longitudinal continuous assessment and the comparison of individual learner progress with expected developmental milestones, both of which can benefit from, and may indeed require, online tracking and reporting. Similarly, continuing medical education and maintenance of competence programmes also require integrated systems given the quantities of data and number of individuals involved.

However, learning analytics does have a number of challenges and limitations, not least of which are what data can be collected and what these data represent. Electronic systems can only record electronic events undertaken within them, primarily mouse clicks (or trackpad taps), and keypresses. These are usually associated with a particular person (user), particular context (on this page or in this system), and a particular date and time. These systems do not record why the event took place, whether the identity of the person in the system is the same as the individual actually causing the events, or what impact those events had on that individual (such as what they learned as a result of this event. Measures such as the time between a user loading a page and moving to the next page may reflect whether or not they read or reflected on the material, but they could also depend on whether the user was doing something else at the same time, or many other factors not associated with the quality of learning [14]. The utility of learning analytics is limited by the events that can be recorded and the semantic precision of what those events mean. It should also be noted that medical education systems may be used to track teacher behaviours as well as those of their learners. This panoptic ability to observe and scrutinise all users and to infer value judgments about them (often without the individuals concerned being aware of this scrutiny) requires careful consideration of the ethical and hidden curriculum implications of doing so [15].

The principle of economies of scale also reflects how technologies are used in managing educational programmes, exemplified by the use of curriculum mapping tools and systems. Although there is no standard approach to curriculum mapping, a comprehensive and interrogatable database of all the sessions, objectives, tests, outcomes, and other curriculum components can be an invaluable asset in managing, revising, and reporting on medical education curricula. Further benefits can accrue when curriculum maps can be aggregated and compared in the interests of evaluation and research [16]. However, the time and effort typically required to set up and maintain curriculum mapping databases tends to limit medical schools' commitment to using tools of this kind. Indeed, economics and return on investment is a critical limiting factor in the use of technologies for managing medical education systems.

Technologies as Educational Prostheses

Having primarily considered technology as a medium for medical education activities, I would also like to consider technologies as 'educational prostheses'. This raises one of the key paradoxes around the use of technology-enhanced medical education, which is that approaches to teaching and learning that work best typically involve little mediating technology. Indeed, one-to-one in-person interactions between teachers and learners, both in the classroom and at the bedside are some of the most effective and valuable either party can experience. To re-mediate these interactions through technology is often to diminish their intimacy and immediacy. Using TEL is rarely the optimal individual instructional approach to take. But circumstances are rarely optimal, and technologies can work as prostheses to compensate for suboptimal conditions.

Although technologies do not learn for learners, they can augment learning by helping them to find, record, organise, structure, and recall knowledge, by reminding them of things and helping to organise their time, by connecting them with others, and by helping them to rehearse, test, analyse, and report on their learning. Technologies can therefore function as learning prostheses; they can extend, structure, and support the ways in which learners learn and interact with the world around them.

Technologies can also act as teaching prostheses by helping teachers to record, organise, structure, remember, remind, connect, find, rehearse, test, analyse, and report on their teaching. Some of the ways in which digital technologies can augment the way we act and interact are set out in Box 10.7.

Prosthetic technologies allow us to work around many of the practical limitations we face in day-to-day medical education practice. Changing the 'rules' in this way may confer certain advantages, but technology mediation tends also to shift perceptions and values around the mediated activities. As Gordon Graham observed: 'technological innovation cannot and should not be regarded merely as an improved means to a pre-selected end, because, while some technology merely modifies, other technology transforms' [17, p. 168]. Augmentation through technological mediation often involves a series of trade-offs and a shifting of values. For instance, the use of online discussion boards means that intra-learner and teacher–learner interactions are now permanently recorded, which in turn means that they can be used as data for evaluation and assessment purposes, something that had no place in medical education until it became possible. Similarly, when teacher and learner interactions can be mediated in many different ways we begin to think about economies of presence (the value of, say, face-to-face as opposed to online interactions) and economies of attention (the ability to attract and retain learner or teacher attention) in ways that previously had no place in the training of health professionals.

Evaluating and Researching TEL in Medical Education

BOX 10.7 Prosthetic properties of digital technologies

Digital technologies can be used to augment teaching and learning through:

- Saving time and effort by helping their users to remember, repeat, find, record, and organise their work, and by accelerating the speed at which they can act and respond to others or emerging events.
- Going beyond physical limitations by helping their users to interact with things that are big (such as visualising datasets), far (such as defeating geography by allowing individuals in different places to work together), or time shifted (such as defeating temporality by allowing individuals to work together asynchronously).
- Organising and connecting with others by helping users to create, broadcast, share, and comment on materials and thoughts, and to discuss, debate, challenge, or lobby for or against particular ideas and positions.
- Changing, blurring, or flattening social conventions by helping their users to act and interact in ways where social structures that shape face to face interactions would have inhibited their ability to do so.
- Heightening the visibility and accountability of individuals and their actions by helping their users to scrutinise, record, and track what they or others do and say.

As with any aspect of medical education practice, efficient and effective practice needs to be based on robust research and evaluation evidence. Although there has been no shortage of studies into technology use in medical education, the quality of the evidence and its currency and relevance to guiding practice is somewhat limited (Box 10.8).

Technology-mediated aspects of medical education (at least initially) tend to be new and unfamiliar and as such they often face a higher burden of proof, scepticism, and over confidence than more traditional approaches. Not only do we need to appraise whether a technology can do what it is supposed to do in a reliable and consistent way, we also need to consider whether it will work in the contexts within which it is going to be used, and whether it will do so consistently for all of those who are going to use it. Both evaluation and research into TEL may draw on needs analyses and environmental scans (what need or opportunity is the technology to address), documentation (of development and application), usability (does it function as expected, can its users use it easily and consistently?), observation of implementation (what happens when people do use it?), participant experience and satisfaction (what is it like for users), learning outcomes (what impacts does using technology have?), and cost, reusability, and sustainability (is using technology a good return on investment?) [21].

BOX 10.8 WHERE'S THE EVIDENCE: Technology-enhanced learning

There have been a number of systematic reviews into whether TEL does or does not work. For instance, one of the largest reviews (focused on primary and secondary education) concluded that technology can confer significant educational advantage when used in a blended form [4]. However, given the diversity of technologies used in medical education and the many different ways in which they are used, simply asking 'does TEL work?' is impossible to answer meaningfully let alone conclusively. We should instead consider in what situations specific kinds of TEL can work and what features of TEL afford different levels and qualities of learning. This argument, advanced by Cook and colleagues [18, 19], responds to the large number of studies that have compared a technology-mediated intervention to no intervention or a technology-mediated intervention to a non-technological intervention, neither of which substantially add to the field. Other germane factors such as cost, sustainability, and risk are seldom considered at all [20]. Contextual factors also need to be considered as there are so many circumstantial variables associated with the efficacy and effectiveness of TEL [14]. All of this may be moot, however, as technological mediation is already becoming the norm in medical education, partly for logistical and convenience reasons rather than for instructional superiority, and partly in response to the broader switch to digital technologies in society as a whole. More research may be needed to evaluate and guide practice and innovation, but the 'should we use technology at all?' train left the station some time ago.

We also need clarity as to whether studies are focusing on evaluating the technologies themselves (such as ‘what properties and capabilities does this tool have?’), on activities that use technology (such as ‘how well does online PBL work?’), or on the use of technology within particular activities (such as ‘what impact does the use of mobile devices have on learner–patient interactions’?). To summarise, rather than proposing an agenda regarding what we should research and evaluate in and around TEL, I argue that it is more important to establish and maintain rigour in the ways we research and evaluate TEL.

A Moving Target

Nothing lasts for ever and digital technologies tend to be more transient than most other aspects of medical education. Although the ‘e-learning revolution’ would seem to be over [22] technologies continue to change. Some new technologies become part of mainstream medical education (such as PowerPoint), others wither on the vine. Predicting which ones will succeed or fail is notoriously hard to predict. For instance, virtual reality (VR) in medical education [23] but, although there are some uses (such as VR laparoscopic surgery simulators), at least so far, the use of VR technologies remains peripheral to mainstream medical education.

Digital professionalism is not the only emerging area of concern in medical education, there are other legal, reputational, and security issues to consider when using TEL. One of the most common challenges for medical teachers is what and how materials from the Web can be used or reused in teaching and learning. Digital technologies can make it very easy to ‘copy and paste’ or download materials and to share them with others. However easy it is to do this, the legality of doing so is rarely so simple. Not only do different countries tend to have different and shifting laws regarding the reuse of third-party materials for educational purposes, there are differing standards regarding liability for misuse of someone else’s intellectual property. This was one of the main drivers behind the development of licensing models such as Creative Commons. See Box 10.9.

BOX 10.9 FOCUS ON: Reusing materials for teaching and learning

Rights for using third-party materials tend to fall between ‘public domain’ (no restrictions on use) and ‘all rights reserved’ (materials cannot be used without the copyright holder’s permission, the default in the absence of any other arrangement). Between these two extremes, licences set out the specific conditions as to how materials can be used. Licences may be commercial (such as with publishers), or open source, (such as Creative Commons). The licensing system developed by Creative Commons (http://creativecommons.org) is based on statements regarding acknowledging the creator(s) of the work (attribution), whether it can be included as part of something else or whether components of the original work can be reused (adaptation), whether or not the work can be reused for commercial purposes, and whether any derivative works need to follow the same licensing model as the source. Creative Commons licences have been used to publish Reusable Learning Objects (RLOs), OpenCourseware, and Free Open Access Meducation (FOAM) materials, as well as many of the resources in online repositories such as MedEdPortal (www.mededportal.org), MedEdWorld (www.mededworld.org), and Wikimedia (commons.wikimedia.org). In the absence of an explicit licensing agreement or permissions for the use of someone else’s material you should always assume that all rights are reserved and that you cannot reuse it. Copyright breaches may be defended (depending on jurisdiction) under principles set out in Fair Use (USA) or Fair Dealing (UK, Canada), but these tend to be expressed as a defence rather than a right. It should also be noted that not only is unlicensed use of third-party materials risky for the individuals and institutions involved, it also sends inappropriate professionalism messages to their learners and colleagues. Using public domain and Creative Commons licensed materials can go a long way to addressing these issues.

The safety and viability of medical educators and learners using TEL is fundamentally challenged by the growing problem of identity theft, cybercrime, and hacking. Assuring and maintaining the security of electronic systems is a constant battle, even if it is one that educators rarely consider. While one way of addressing this is to harden electronic systems against attack, human rather than technical weaknesses (such as compromised passwords) are often the bigger threat problem, and training and policing user mistakes and misdemeanours can be expensive. An alternative and somewhat radical approach that has been gaining a degree of traction is the concept of a ‘post privacy society’ where all information is made publicly available [24] with a focus on security rather than privacy. While, at the time of writing, there is no imminent prospect of such a major cultural shift, it is undeniable that the widespread uptake of digital technologies around the world has impacted both medicine and education at a fundamental level and that, as a result, medical education is changing – whether it wishes to or not.

Whatever use they make of TEL, medical educators need to plan for change and instability including system failures, upgrades, replacements, or changing requirements. They need to be mindful of their dependence on the infrastructure they use (such as networks and security), and they need to monitor the viability of their technology use and its resilience in the face of attack.

Conclusions

Technologies go out of date so quickly; indeed, it is quite possible that some of the technologies mentioned in this chapter will have already become obsolete by the time this book is published. My focus has therefore been on principles and concepts that should transcend the specific and somewhat ephemeral technologies that are being used in medical education at any given time. Nevertheless, I have tried to provide concrete examples to illustrate my arguments. While these examples will inevitably become dated, I hope that the principles they illustrate will stand. I suggest therefore that readers reinterpret the specifics of their own time and circumstances in light of these principles rather than (except perhaps in the sense of historical curiosity) focusing on the specifics.

The use of TEL in medical education can be an enabler or a catalyst for positive change, and it can also be a disruptor, a distractor, and a liability. The value of technology use is rarely based on purely educational issues and tends instead to be a matter of convenience, mediation, and augmentation. To that extent, arguments for and against the use of TEL are as much, if not more, about educational economics rather than instruction. The challenge is to find more efficient and convenient ways to teach and learn, and these may or may not involve the use of technologies. TEL is also constrained by what is available. Teachers and learners tend to make do with and adapt to the affordances of the technologies they have at hand. Moreover, technologies tend to be used in an additive way rather than displacing existing practices. Displacement only happens when the convenience and utility of one technique so outweighs the other that it effectively withers away.

Given that the quality of learning and teaching depends on so many things (considerations such as who is involved, what they are learning, where they are learning, and how much time and effort can they put into their learning and so on), then, rather than asking whether a particular learning technology is better or worse than an alternative (technological or otherwise), we should ask whether it is useful and effective in supporting particular needs of particular learners and teachers in a particular place and time. I therefore return to a recurring theme in this chapter; that technological mediation of medical education is rapidly becoming the norm and as such educators should consider the intersectionalities of technical and educational issues rather than one or the other in isolation. The ideas and principles I have set out in this chapter should therefore be understood as intersecting with and informing the ideas and principles set out in every other chapter in this book. In the same way that technologies mediate and augment teaching and learning in medicine, TEL concepts should be seen as mediating and augmenting medical education in all its many forms and practices.

References

1 Masters, K., Ellaway, R.H., Topps, D. et al. (2016). AMEE Guide 105: mobile technologies in medical education. *Medical Teacher* 38 (6): 537–549.

2 Engeström, Y. (2009). Expansive learning: toward an activity-theoretical conceptualization. In: *Contemporary Theories of Learning* (ed. K. Illeris), 55–73. Abingdon, UK: Routledge.

3 Horton, W. (2006). *E-Learning by Design*. San Francisco: Pfeiffer.

4 Means, B., Toyama, Y., Murphy, R. et al. (2009). *Evaluation of Evidence-Based Practices in Online Learning: A Meta-Analysis and Review of Online Learning Studies*. Washington DC: U.S. Department of Education.

5 Ellaway, R.H., Coral, J., Topps, D., and Topps, M.H. (2015). Exploring digital professionalism. *Medical Teacher* 37 (9): 844–849.

6 Ellaway, R.H. (2016). Commentary. Medium, message, panopticon: the electronic health record in residency education. *Journal of Graduate Medical Education* 8 (1): 104–105.

7 Ho, K., Ellaway, R., Littleford, J. et al. (2014). eHealth Competencies for Postgraduate Medical Education: CanMEDS 2015 eHealth Expert Working Group Report. Royal College of Physicians and Surgeons of Canada, Ottawa, Ontario.

8 Prensky, M. (2001). Digital natives, digital immigrants. *On the Horizon* 9 (5): 1–6.

9 Reigeluth, C.M. (1983). *Instructional Design Theories and Models: An Overview of Their Current Status*. Hillsdale, NJ: Lawrence Erlbaum Associates.

10 Sweller, J. and van Merriënboer, J. (2013). Instructional design for medical education. In: *Oxford Textbook of Medical Education* (ed. K. Walsh), 74–85. Oxford, UK: Oxford University Press.

11 Mayer, R.E. (2001). *Multimedia Learning*. New York, NY: Cambridge University Press.

12 Richey, R.C., Klein, J.D., and Tracey, M.W. (2011). *The Instructional Design Knowledgebase: Theory, Research, and Practice*. New York, NY: Routledge.

13 Amin, Z., Boulet, J.R., Cook, D.A. et al. (2011). Technology-enabled assessment of health professions education: consensus statement and recommendations from the Ottawa 2010 conference. *Medical Teacher* 33 (5): 364–369.

14 Ellaway, R.H., Pusic, M., Yavner, S., and Kalet, A. (2014). Context matters: emergent variability in an effectiveness trial of online teaching modules. *Medical Education* 48 (4): 386–396.

15 Land, R. and Bayne, S. (2005). Screen or monitor? Surveillance and disciplinary power in online environments. In: *Education in Cyberspace* (ed. R. Land and S. Bayne), 165–177. Abingdon, UK: Routledge.

16 Ellaway, R.H., Albright, S., Smothers, V. et al. (2014). Curriculum inventory: modeling, sharing and comparing medical education programs. *Medical Teacher* 36 (3): 208–215.

17 Graham, G. (1999). *The Internet://a Philosophical Enquiry*. Abingdon, UK: Routledge.

18 Cook, D.A. (2009). The failure of e-learning research to inform educational practice, and what we can do about it. *Medical Teacher* 31 (2): 158–162.

19 Cook, D.A. and Triola, M.M. (2014). What is the role of e-learning? Looking past the hype. *Medical Education* 48 (9): 930–937.

20 Huang, G., Reynolds, R., and Candler, C. (2007). Virtual patient simulation at US and Canadian medical schools. *Academic Medicine* 82 (9): 446–451.

21 Cook, D.A. and Ellaway, R.H. (2015). Evaluating technology-enhanced learning: a comprehensive framework. *Medical Teacher* 37 (10): 961–970.

22 Ellaway, R. (2011). E-learning: is the revolution over? *Medical Teacher* 33 (4): 297–302.

23 Gorman, P.J., Meier, A.H., and Krummel, T.M. (1999). Simulation and virtual reality in surgical education: real or unreal? *Arch Surg.* 134 (11): 1203–1208.

24 Younge, G. (2012). Social media and the post-privacy society. *The Guardian*, 2 April 2012. https://www.theguardian.com/commentisfree/cifamerica/2012/apr/02/social-media-and-post-privacy-society (accessed 18 March 2017).

Further Reading and Resources

Clark, R.C. and Mayer, R.E. (2016). *e-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning*, 3e. Hoboken, NJ: Wiley.

Selwyn, N. (2016). *Is technology good for education?* Cambridge, UK: Polity.

Clark, R.C. (2012). *Scenario-Based e-Learning: Evidence-Based Guidelines for Online Workforce Learning*. San Francisco, CA: Pfeiffer.

Ellaway, R. and Masters, K. (2008). AMEE Guide 32 : e-learning in medical education part 1: learning, teaching and assessment. *Medical Teacher* 30 (5): 455–473.

Masters, K. and Ellaway, R. (2008). AMEE Guide 32 : e-learning in medical education part 2: technology, management and design. *Medical Teacher* 30 (5): 474–489.

11 Simulation in Medical Education

Alexis Battista[1] *and Debra Nestel*[2,3]

[1]Department of Medicine & F. Edward Hébert School of Medicine, Uniformed Services University of the Health Sciences, Bethesda, MD, USA
[2]Department of Surgery, University of Melbourne, Melbourne, Victoria, Australia
[3]Monash University, Clayton, Victoria, Australia

KEY MESSAGES

- Simulations and simulation-based learning (SBL) have gained acceptance and popularity among diverse health care professions, including medicine, nursing, allied health professionals, dentistry, and basic sciences.
- Some of the most common applications of simulations include, supporting health professional learning, patient safety initiatives, interprofessional collaborative (IPC) practice, and mediating clinical teaching constraints.
- As in any educational practice, theories can inform all facets of SBL and over reliance on one theory may inhibit learning opportunities.
- As SBL has come of age, debates about whether simulation 'works' have turned towards discussions about how and when to best employ simulation.
- A structured and systematic approach to designing SBL is valuable and may include the phases of preparing, briefing, simulating, debriefing, reflecting, and evaluating.
- Fidelity is a contested concept in SBL and it may be more helpful to think about learner engagement, functional task alignment, and meaningfulness for learners.
- Communities of practitioners who use simulation as an educational method have evolved all over the world, often forming professional associations and sharing valuable resources.
- As SBL matures, some future themes of discussion include research practices in simulation, and communicating leadership and scholarship about the community.

Introduction

Simulation is a technique which can be used to facilitate any learning, whether in the cognitive, psychomotor, or affective domains. Simulation is defined as 'A technique that creates a situation or environment to allow persons to experience a representation of a real event for the purpose of practice, learning, evaluation, testing, or to gain understanding of systems or human actions' [1]. Simulation-based learning (SBL) is defined as 'an array of structured activities that represent actual or potential situations in education and practice' [1]. The technique can involve a wide range of activities and approaches and is applicable to learners, from novice to expert. Simulation is not confined to interactions with people or models, physical or virtual; it could just as easily be a paper-based or table-top activity. Simulation can be used effectively in the classroom, in a designated simulation centre, or in the clinical workplace, also known as 'in situ simulations'.

Simulations and SBL continue to gain widespread acceptance and popularity among diverse health care professions, including medicine, dentistry, basic sciences, nursing, and other allied health specialists (e.g. pre-hospital care, respiratory technology) [2]. Some of the most common reasons simulations have gained popularity include, (i) supporting efforts to improve health care provider performance in crisis events [3]; (ii) its use in lieu of practising on actual patients to support patient safety efforts [4–6]; and (iii) being viewed as a solution to many of the challenges associated with contemporary medical education [7, 8]. Research into the use of simulation in the health professions has grown extensively in the past decade with findings from meta-analyses suggesting that simulations support gains in medical knowledge, decision-making, critical thinking, psychomotor skill acquisition, communication skills, self-confidence, and self-efficacy, to name a few [2, 9–12].

In the sections that follow, this chapter provides a practical overview of SBL in medical education, including:

- historical and contemporary perspectives of how medical education stakeholders employ simulation
- theoretical perspectives that guide and inform SBL practice

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

- practical strategies for designing and implementing simulation
- limitations and challenges of SBL
- emerging issues in SBL practice and research.

Historical Perspectives and Contemporary Applications of Simulation-based Learning

Simulations and SBL have been around for centuries in many areas of human endeavour [4]. Simulators in health professions learning date back as far as eighteenth century France, when Madame Du Coudray developed a SBL curriculum using a fetal model to train midwives in safe birthing practices [13]. The modern movement in SBL coincided with developments in the technology of patient simulators and the growing interest in improving patient safety. For example, the part-task trainer, *Resusci Anne*, led the way in standardising resuscitation training [14] and *Sim-One*, a patient simulator, developed by Abrahamson at the University of Southern California School of Medicine, helped novice anaesthetists develop skills in inserting endotracheal tubes [15]. The 1980s saw the development of the *Gainesville Anaesthetic Simulator* [16] and the *Comprehensive Anaesthesia Simulation Environment* (CASE) [17]. Further, Barrows introduced simulated patients (SPs), which provided learners with a controlled, staged experience of simulated practice [18].

Gaba et al.'s use of simulation to study operating room team members' responses to crisis events contributed to the use of simulations as a strategy for teaching clinicians how to respond to crisis and high-risk, low-frequency events [3]. Use of simulations for patient safety initiatives was bolstered when the Institute of Medicine published its landmark report '*To Err is Human*' in 1999, exposing medical error (both acts of commission and omission) as a leading cause of patient injury and death [19]. Ziv et al. [6] and Aron and Headrick [20] further argued that allowing untrained or minimally trained clinicians to practise on actual patients created an unacceptable risk when a patient simulator was available. These events, together with technological advances in patient simulators, helped SBL gain added popularity and acceptance in diverse domains, including emergency medicine [21], obstetrics [22], and neonatal, paediatric, and adult critical care [23–25], to name a few.

Contemporary Applications of Medical Simulation

Some of the most common applications of SBL include (i) supporting patient safety and quality programmes, (ii) skills training and competency assessment, (iii) ameliorating clinical teaching constraints, and (iv) supporting the development of IPC practice (see Box 11.1).

BOX 11.1 Common applications of simulation in medical education [4]

Patient safety:
- training of teams in crisis resource management
- patient-specific rehearsal of planned, novel, or infrequent interventions (e.g. pre-surgical rehearsal)
- design and testing of new clinical equipment
- design and testing of new patient wards or workflows

Skills training and competency assessment:
- routine learning and rehearsal of clinical and communication skills at all levels
- routine basic training of individuals and teams
- practice of serious and/or rare events (e.g. cardiac arrest management)
- induction into new clinical environments and use of equipment
- performance assessment of health professionals at all levels
- refresher training of health professionals at all levels

Ameliorating clinical teaching constraints:
- replacement of up to 25% of clinical rotations when clinical sites are limited
- ensure predictable and reliable clinical experiences for learners

Simulation-based interprofessional collaborative practice:
- explore professional identity and learn about the professional roles of other health care professionals

Simulation and Patient Safety and Health Care Quality

Simulations and SBL are commonly viewed as beneficial because they allow clinicians to learn and practise diverse skills without placing patients in harm's way [3, 6]. SBL is frequently used to train health care professionals in the development of non-technical skills (e.g. communication, situational awareness) that are vital to the successful management of crisis situations [3]. SBL is also used to support patient-specific rehearsal of planned, novel, or infrequent clinical events, such as carotid artery stenting, neurological procedures, and neonatal surgery [26–28]. Simulation is increasingly employed to evaluate latent environmental threats in the clinical setting [29, 30] and, more recently, to assess new patient wards and develop health care professional workflows [31].

Skills Training and Competency Assessment

SBL plays an important role in supporting skills training and competency assessment during undergraduate and postgraduate training and continuing professional development. SBL is regularly integrated into the undergraduate curriculum of diverse health professions education programmes where students learn diverse clinical skills, such as patient assessment, procedural skills, and teamwork [2, 32]. In addition, studies suggest that simulations used to train undergraduate or postgraduate students in advance of clinical practice may help improve self-confidence, self-efficacy, and skills performance, and may help ease the transition to the clinical setting, while also improving patient safety [33, 34]. For example, simulation-based boot camps are a common strategy used by hospitals to help postgraduate health professionals prepare for future clinical work. Boot camps offer learners opportunities to acquire critical clinical skills (e.g. resuscitation, vascular access) over short periods of time (e.g. days, weeks) before entering the clinical setting [35, 36]. In many cases, learners are required to meet pre-established minimum passing scores prior to being allowed to enter the clinical setting [23].

In addition to helping undergraduate and postgraduate students gain skills and knowledge, experienced practitioners also benefit from SBL. For example, Draycott et al. [22] conducted a simulation intervention with midwives and obstetrical staff to improve their management of neonatal injury associated with shoulder dystocia (SD). Outcome measures for this study included a retrospective review of birth records before and after the simulation training for the use of manoeuvres and neonatal outcomes. Following the introduction of SBL, clinical management of SD improved [22].

Ameliorating Clinical Teaching Constraints

SBL is an instructional strategy used to address clinical teaching constraints. For example, SBL provides students and educators with reliable, scalable, learning opportunities, during which feedback and interaction with faculty are incorporated [5]. In some cases SBL is used as a surrogate learning environment to augment or replace clinical experiences when access to clinical rotations is limited, or when competition among health care professional education programmes dilutes clinical experiences. For example, a recent multi-institutional randomised controlled trial, including 10 schools of nursing in the United States, was conducted to determine if SBL could replace 25–50% of clinical rotations, while not having a detrimental effect on commonly used outcome measures (e.g. knowledge assessments, clinical competency ratings, board pass rates) [32]. Hayden et al. [32] reported that students with 25–50% of simulated clinical experiences had similar scores on knowledge assessments, clinical competency ratings, and board exam pass rates when compared to students who did not partake in simulated clinical experiences. In physiotherapy education, a similar outcome has been reported with students from six Australian universities undertaking clinical education in ambulatory care settings with patients with musculoskeletal disorders. A SBL programme was developed to 'replicate' one week of a four-week clinical placement in musculoskeletal practice. Students' achievement of clinical competencies on the Australian Physiotherapy Practice tool was no worse in the SBL groups than in the traditional clinical placement groups [37].

Simulation-based Interprofessional Collaborative Practice

SBL is increasingly used to support the development of IPC practice. Simulation-based IPC practice is viewed as beneficial as it augments traditional interprofessional learning activities, such as table-top discussions or case studies, by enabling learners to engage more deeply in their professional role. In turn, this deeper engagement helps participants explore implications associated with social dynamics, professional hierarchy, and diversity more fully [38]. In addition, SBL for IPC practice helps learners gain a more robust understanding of their professional identity, while learning more about the roles of other professions [39]. A more in-depth discussion of interprofessional education is covered in Chapter 14.

Integrating Theory into Simulation-based Learning

In SBL, theories can inform the initial educational design such as making decisions about what simulation modality to choose and why; they can assist with resolving dilemmas such as how to manage underperforming learners; or, they can challenge accepted practices such as a particular approach to debriefing [40]. Some of the most common theories employed in SBL include experiential learning theory, adult learning principles, and cognitive apprenticeship. In this section, although theories that inform educational practice and SBL are shared throughout this book (see Chapter 4), we focus on broad theoretical traditions – behaviourism, cognitivism, and constructivism – and how they have influenced SBL. Please see the references and further readings for more detailed discussions.

In its simplest form, *behaviourism* enables learning in response to a stimulus. The appeal of this theory to SBL is the provision of a designed experience in simulation with a predetermined response in which the feedback learners receive helps shape the desired responses. For example, behaviourism's influence on SBL is apparent with the establishment of *behavioural* learning objectives and measurable outcomes. Through participation in simulation, learners are provided the opportunity to meet learning objectives and receive feedback on their performance that rewards desirable behaviours. These activities can be psychomotor skills, such as physical examination, cognitive tasks, such as pattern recognition, or communication skills, such as checking that a patient understands the information shared. Furthermore, these activities are well suited to SBL because of the ease with which simulators can provide the behavioural experience and feedback (simulator, peer, or faculty generated) to shape responses to achieve automaticity.

Linked to this tradition, Ericsson describes the concept of deliberate practice, a theory that offers insight to the development of expertise [41]. Through the observation of elite performers in many disciplines, he identified several key elements of practice that led to their elite status, including high degrees of learner motivation, engagement in frequent practice, and receipt of feedback from a coach or mentor. Among SBL, Issenberg et al.'s [7] review indicated that 43 (39%) of 109 included studies specified that the second commonly accepted benefit of SBL is that it affords learners with opportunities to engage in repeated practice. They describe repeated practice as 'focused, repetitive practice where the intent is skills improvement' [7]. In 2011, McGaghie and colleagues conducted a follow up critical review of SBL in which they extended the discussion about repeated practice. They suggested that repeated practice in SBL supports learning as these features are commonly associated with deliberate practice, including engagement in well-defined tasks tailored to the needs of the learner, access to regular practice, and the provision of feedback derived from both the simulator and faculty [11].

Of course, deliberate practice is not purely behaviourist but draws on the key elements of repeated practice, goal setting, and feedback in shaping performance. Although potentially resource intensive, simulation offers an excellent opportunity to support learning by deliberate practice.

Cognitivism gained prominence in response to behaviour being perceived as much more than a stimulus–response activity; that individuals have information processing capabilities that also influence behaviours. Cognitivism explores the individual's ways of thinking and knowing, memory capacities, and problem solving. Knowledge is seen as symbolic mental constructions or schema with learning considered as changes in the schema. Cognitive load theory is commonly cited in design considerations for SBL [42, 43]. For example, too much or too little cognitive load will inhibit learning. That is, too much information, not enough information, too difficult a task, too easy a task, and information presented in an ill-considered or unstructured way, can result in cognitive overload (or underload) for a learner and therefore negatively impact learning [43]. Effective management of cognitive load in SBL is often addressed as a component of *instructional design* [42, 44].

Any discussion of theories important for SBL must also include reference to *mastery learning*. Mastery learning is not a theory per se but an approach that draws on several theories – including deliberate practice, the notion of scaffolding, and instructional design. McGaghie and colleagues have demonstrated its application in diverse clinical skills [45, 46]. The approach is characterised by the focus on individual rather than time-based milestones, on baseline and progressive testing, on the clear description of sequenced skills development steps, and opportunities for repetitive practice with feedback. Box 11.2 lists seven essential steps in the use of simulation-based mastery learning [47].

BOX 11.2 Steps associated with mastery learning

1. Baseline, or diagnostic, testing
2. Clear learning objectives, sequenced as units usually of increasing difficulty
3. Engagement in educational activities (e.g. deliberate skills practice, calculations, data interpretation, reading) focused on reaching the objectives
4. A set minimum passing standard (e.g. test score) for each educational unit
5. Formative testing to gauge unit completion at a pre-set minimum passing standard for mastery
6. Advancement to the next educational unit given measured achievement at or above the mastery standard
7. Continued practice or study on an educational unit until the mastery standard is reached [47]

Another useful idea for SBL associated with cognitivism is the notion of 'scaffolding' [48]. Similar to Vygotsky's 'zone of proximal development' [49], scaffolding refers to the supports faculty put in place to help learners achieve more than they would be able to do without supports. In SBL, support may include the amount of information shared with learners during the pre-simulation briefing, the opportunity to pause and discuss progress during a simulation, or the presence of a 'confederate' who can direct the unfolding of events in a simulation. Knowing when to offer support, how much support to offer, and when to remove support are important decisions for faculty.

In addition to drawing on the traditions of behaviourism and cognitivism, SBL also draws on constructivism! *Constructivism* is an umbrella term for many theories that acknowledge the role of the learner in creating their own meaning from experiences rather than necessarily the teacher 'teaching'. These theories reflect the oft-cited metaphors for learning through *participation* rather than *acquisition* [50]. For example, in SBL, faculty are often referred to as facilitators, reflecting their role as an enabler of learning.

Reflective practice is commonly described as an illustration of a constructivist approach to learning. Briefly, this theory proposes that during and after an unexpected or critical event, practitioners (learners) will reflect-in-action *and* reflect-on-action [51]. In doing so, learners draw on their prior experiences, on cues in the current situation, and consider how these experiences might influence future practice.

In the last decade, SBL scholars have discussed theories that seek to explain learning in complex environments. For example, Fenwick and Dahlgren thoughtfully precis complexity theory, acknowledging its many traditions and diverse perspectives [52]. They wrote: 'most would agree that complexity theory examines how living phenomena (learning, for example) emerge in a web of relations that form among things, including both social and material things, such as bodies, instruments, desires, politics, settings, and protocols. Such things do not come together in a linear cause–effect trajectory, as so many aspects of our curricula seem to presume, nor are they ordered together through top-down authority. Instead, they become combined through myriads of non-linear interactions that continually present novel possibilities and exercise multiple causal influences on what emerges.' These theories usually vest power in all objects (human and non-human) in the environment and that learning occurs because of these interactions. Although these theories then can underpin arguments for reproducing entire clinical environments for scenarios and running in situ simulations, they also draw attention to the importance of the design of scenarios that may fall short of the linear cause–effect trajectory described above. Health care simulation offers tremendous scope for exploring *how* learning occurs in complex environments and hence shape future approaches to designing effective SBL.

Simulation-based Instructional Design

As SBL has come of age, debates about whether simulation 'works' have turned towards discussions about how and when to best employ simulation. In addition to determining how and when simulation ought to be used, there is a growing dialogue about what instructional strategies should be used to support or enhance simulation participants' learning [53]. Instructional design of effective SBL requires stakeholders to balance and consider several factors, including:

- determining and establishing learning goals and objectives
- choosing a simulation method (i.e. skills- or scenario-based)
- selecting a simulation modality (e.g. simulated/standardised patient, task model)
- considering how learners may be assessed
- deciding how and when learners will receive feedback and guidance.

Determining Learning Goals and Objectives

Establishing well-defined goals is central to the simulation instructional design process and should be done early on because it helps inform later decisions about which simulation method and modalities to use and helps inform decisions about assessment and feedback. When determining goals, consider the following:

- Why is a simulation worth doing?
- What issues do you want learners to understand?
- What practices or behaviours do you want to influence?
- Why do you want to do a simulation?
- Is the simulation being used for teaching or assessment purposes?

Setting goals early can help keep the design process focused and can be revisited when you are trying to balance decisions about what you hope to do with what is practical to do.

Choosing a Simulation Method

Designing a simulation activity or curriculum also requires considering which simulation method(s) will best support the goals and objectives outlined at the beginning. The two most common methods are skills-based or procedurally focused simulations and scenario-based simulations.

Skills-based simulations, which represent a partial system or anatomical landmark, are used to emphasise the teaching and practice of a designated skill (e.g. placement of an intravenous catheter, auscultation, ultrasound) [54]. They are viewed as beneficial because they allow educators to intentionally remove many of the complexities and distractions found in the clinical setting, thus allowing learners to focus their efforts on specific skills [22]. Skills-based simulations can support individual learners or small groups of learners and can be conducted in a dedicated simulation lab, classroom, or clinical setting and, in some cases, can be distributed to students for home-based practice (e.g. suturing and knot tying kits).

By comparison, scenario-based simulations, also described as *high-fidelity* simulations or *high-fidelity* scenarios, are often employed when the desired learning goals include learning how to work in a team, communication skills, or responding to a critical patient event [8]. Scenario-based simulations seek to incorporate the complexities associated with clinical practice, including engaging socially with the patient or support persons (e.g. simulator, simulated/standardised patient) and interacting with other health care professionals [8, 55]. In a scenario, a narrative is employed to guide learners' engagement and learners are assigned to specific clinical roles, such as that of the nurse, physician, or other health care professionals [8, 56]. Learners are also expected to conform to the behavioural and practice-oriented conventions of those roles as they would in clinical practice [56–58].

Box 11.3 presents a comparison of skills-based and scenario-based simulations, looking at reasons for use, common examples, and rules of participation that should be considered when designing a course or curriculum that integrates SBL.

Choosing a Simulator Modality

Another important consideration is determining which simulator will best support the goals and objectives of the simulation. Simulation modality refers to the type of simulator being used to support your goals and learning objectives [1]. Some of the most common modality categories include, part-task trainers, computer-based, manikin-based, simulated/standardised participants, and hybrid strategies (see Box 11.4).

Part-task Trainers

Part-task trainers (PTTs) are used to support teaching and learning of procedural and technical skills (e.g. vascular access, ultrasound, surgical procedures). PTTs may also include highly

BOX 11.3 Summary of skills-based and scenario-based simulations

	Skills-based simulations	Scenario-based simulations
Motivations educators employ	Emphasises a single focused procedural skill	Emphasises clinical complexity
Common application examples	Procedural skills (e.g. auscultation, central line insertion)	Communication, crisis or critical patient management, interprofessional teamwork, diagnostic/clinical reasoning
Rules of participation	Skill improvement or skill mastery	Treat the scenario 'as if' the experience is an actual clinical experience

BOX 11.4 Categories of simulator modalities

Simulator type	Examples
Part-task trainers	Venepuncture arms, arterial arms, male and female pelvic models, skin and tissue jigs for injection and suture practice, ultrasound models
Computer-based simulations	Emergency medicine (*Microsim*, Laerdal), anaesthesia or haemodynamic simulators (e.g. Anesoft), bronchoscopy and laparoscopic simulators
Manikin-based	SimMan, Gaumard, METI
Simulated patients	Standardised patients, simulated participant, embedded participants
Hybrid	Combined use of standardised patient and venepuncture arm

sophisticated computerised human patient simulators such as *Harvey*®, a cardiology simulator that can be used to learn non-auscultatory and auscultatory physical findings [59].

Computer-based Simulations

Computer-based simulations (CBS), also referred to as screen-based simulations, provide learners with interfaces that allow them to interact with materials relating to basic sciences, complex physiological models, or support the development of decision-making. Many CBS allow the learner to progress at their own pace and many offer the option to require learners to reach a predetermined level of skill before progressing to the next situation or level. They can also generate user data, which can be used to guide detailed feedback on performance and can be maintained as an ongoing record. CBS also include virtual, augmented, and haptic enabled simulations.

Manikin-based Simulators

Manikin-based simulators involve the use of a manikin to represent the patient that can represent many of the life-like aspects of people (e.g. heart and lung sounds, palpable pulses, bleeding) [1]. The observed clinical vital signs and the electrical readouts can be controlled and altered in response to interventions and therapies initiated by the learners as they interact with the manikin. Integrated simulators may be model- or instructor-driven. Model-driven simulators are physiological and pharmacological models that directly control the manikin's responses to intervention and treatments. Instructor-driven simulators respond to instructor intervention, either directly via the computer keyboard or via a pre-written computer algorithm.

Simulated Participants

The terms 'standardised patient' and 'simulated patient' and 'simulated participant' (SPs) are often used interchangeably [60]. Barrows defined a simulated patient as 'a well person trained to simulate a patient's illness in a standardised way' [18, 60]. SPs contribute to both learning and assessment [61]. They may support learning in several domains, including communication and consultation skills, physical examination, non-invasive procedural skills, and the assessment of professionalism. SPs are also known as clinical teaching associates and there are other variants on the name. Some SPs provide learners with opportunities to undertake male and female genital and digital rectal examination and female breast examination. SPs may also be trained to portray family or patient support persons and clinicians, such as nurses or physicians [55, 62]. SPs are also usually trained to give learners feedback on diverse aspects of their performance [57].

Hybrid Simulations

Hybrid simulations typically include the union or joining of two or more simulation modalities [1, 63]. The most common types of hybrid simulations are those that combine a PTT (e.g. urinary catheter model, central line model) with an SP. Hybrid approaches are often used when the goals of the scenario emphasise communication and patient assessment, yet may also require the learner to perform procedural skills that are more safely done on a model.

Integrating Feedback and Assessment into SBL

A central instructional design feature of SBL is that it affords learners with access to diverse opportunities to engage in reflection or receive feedback and support [7, 64–68]. Deciding who, when, and which debriefing tool will be used are all important considerations. For example, learners may receive feedback from multiple sources, including trained debriefing facilitators, faculty or other subject matter experts, peers, SPs, as well as from simulators themselves [64, 68, 69].

In addition, although simulations are most commonly known for their use of a post-simulation reflection activity, called a debriefing, learners may also receive feedback during simulation practice [70]. During post-simulation debriefing, learners, often guided by a trained facilitator, are encouraged to reflect on their simulation experience. Rudolph and colleagues indicate that this effort helps learners make sense of their actions and activities [68]. During simulation, learners often receive feedback from the simulators themselves when they respond to learners' actions or omissions (e.g. physiologic changes reflected on the patient monitor) or when learners pause during simulation practice to discuss their actions [70].

Selecting a debriefing tool can help focus and structure facilitator, faculty, or learner efforts while helping ensure that important steps are not missed, and ideally helps foster a safe and supportive setting. Examples of debriefing tools include the diamond debrief [17], plus/delta [64], debriefing with good judgement [68], debriefing for meaningful learning [71], and the structured and supported method [72], as well as others provided in the London Handbook of debriefing [73].

Similarly, debrief facilitator rating tools such as the *Objective Structured Assessment of Debriefing* [73–75] and the *Debriefing Assessment for Simulation in Health care* [76] have been developed to provide evidence-based guidelines for conducting debriefings in simulated and real clinical settings. Guidelines for video-assisted debriefing have been published [77–80] but optimal use remains unclear.

Designing a Simulation Activity

In addition to the above instructional design processes, designing the implementation processes for a simulation activity also benefits from a structured approach. Figure 11.1 illustrates six common phases of a SBL event: preparation, briefing, the simulation activity, debriefing and feedback, reflecting, and evaluation.

The *preparing* phase refers to all the activities that take place before the simulation event starts such as: identifying learners' needs; setting learning objectives; designing the simulation; sourcing simulators, medical equipment, and props; booking rooms; scheduling the learners; recruiting and identifying faculty, confederates, and SPs; and catering. It is important to keep in mind that the range of tasks will need to be adapted to the local simulation facility and its practices.

Although *briefing* is given relatively little attention in the literature when compared with debriefing, this phase is a critical component that helps set the stage for meaningful SBL experiences [66]. Briefing may include faculty, learners, and local simulation centre team members such as SPs and technologists. The briefing for faculty may include a re-statement of the learning objectives; the learners' characteristics; logistics such as time frames and cues to start, pause, and end the simulation; simulator programming; technical support; communication with the control room; audiovisual capacity; debriefing and feedback processes; reflective exercises; and evaluation forms. If the simulation involves SPs, they may be briefed separately to learners.

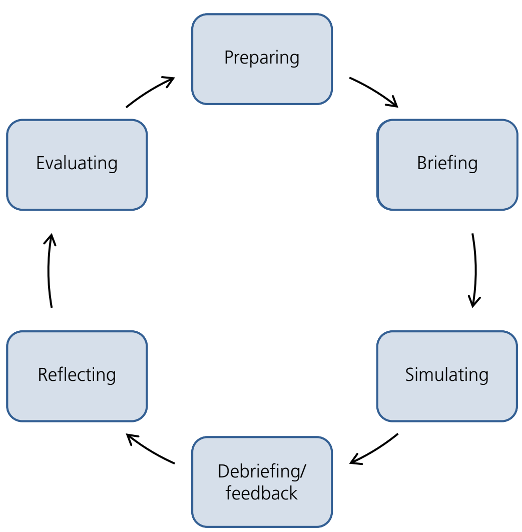


Figure 11.1 Phases in simulation implementation design. *Source*: Adapted from the NHET-Sim Program (http://www.monash.edu/medicine/nhet-sim).

Briefing the learners will include many of the items above and may also include inviting learners to set their own goals relative to those prescribed and their experiences [81]. Additionally, orientation of learners to the simulation and simulators is important. This will include explicit discussion on what is similar and what is different to real clinical practice. This is linked to what is called a *fiction contract*, where learners are invited to perform as if they were in the actual clinical setting. Some learners find simulation stressful and it may be important to normalise the experience during the briefing.

Importantly, creating a *safe* learning environment requires a truly learner-centred attitude from faculty. Learner safety can be achieved through clear explanation of the simulation phases and the learners' responsibilities, clarity over who is observing, what will happen with audiovisual recordings, confidentiality among those involved, seeking 'buy in' with respect to doing their best, and the orientation or familiarisation of the simulators and setting [66].

The *simulation activity* represents the period when learners are engaging in the simulation. During this period, faculty and simulation operations specialists should ensure that they indicate a clear start to the simulation and observe for physical and psychological safety of those within the simulation. In addition, it is also important to ensure that required cues are implemented as planned and that anything discussed during the briefing stage is enacted as stated, such as a 'pause and discuss' option for reflection. Faculty and simulation operators should also ensure that observers are encouraged to make notes to enable specific feedback during debriefing to increase value [82].

Immediately after the simulation is over it is important to help learners transition from the simulation to the debriefing and feedback phase. During this phase, faculty or debriefing facilitators should observe the learners because they often vent their frustrations, which can be helpful to the facilitator in opening the debriefing. Another common strategy that can be helpful is encouraging learners to spend a few minutes thinking about what happened or what worked well and what could have been improved. If observation tools were used, the transition time provides an opportunity for completion. It is also helpful to take a few moments to ensure that the physical space of the debriefing room is appropriately organised, paying attention to seating arrangements, and whiteboard and/or TV screen if video-assisted debriefing is used. As a facilitator, or if SPs are also attending the debriefing, having the learning objectives at hand (perhaps on a shared visual field such as a white board) helps to stay focused. Stick to the processes outlined in the briefing, although flexibility is also important to ensure learner-centredness. Use opportunities, especially for communication-based scenarios, to rehearse micro elements (segments) of the scenario.

The *debriefing and feedback* phase complements the briefing, almost as bookends to the simulation activity [81]. Facilitators explore learners' feelings, address goals and learning objectives, seek other perspectives, summarise, affirm positive behaviours, explore unplanned issues, and seek to establish new goals [83]. Although one goal of the debriefing is to promote reflection, the phase also highlights the importance of individual reflection and returning the locus of control for learning to the learner when they have left the simulation event.

For the *reflecting* phase, learners (usually individually) are encouraged to make sense of the simulation in the light of their own past and anticipated future experiences. During briefing, learners can be informed of reflecting activities and this can be reinforced after the debriefing. Of course, there is overlap between these phases and reflecting can occur before the debriefing. There are several approaches to reflecting that have been adopted in SBL [84–86]. Learners can also be directed to evidence their reflective practice following simulations by uploading and tagging digital learning resources (audio, photographs, video, and podcasts, etc.) [81].

In these simulation phases, *evaluating* refers to the successes and limitations of the simulation event in meeting its goals, rather than assessment of the individual. This phase benefits from involvement of all stakeholders. Chapter 7 addresses quality in medical education. Of course, evaluation is a crucial element to drive improvements in education, health care practice, and ultimately patient care [8, 87]. While it is essential to consider the degree to which the simulation activity has supported learning, meaningful evaluations often require sophisticated methods. Complex learning interventions require equally complex evaluations, using qualitative and quantitative methods to draw on multiple sources and triangulating data alongside exploring multiple levels of impact [28].

Simulation Fidelity

The Society for Simulation in Health care (SSH) dictionary defines *fidelity* as:

> [The] degree to which the simulation replicates the real event and/or workplace; this includes physical, psychological, and environmental elements … the ability of the simulation to reproduce the reactions, interactions, and responses of the real-world counterpart … It is not constrained to a certain type of simulation modality, and higher levels of fidelity are not required for a simulation to be successful [1].

Examples include the level of realism associated with programming physiological responses in manikins that mimic those of a real patient or the haptic sensation associated with using a laparoscopic simulator. These experiences require sophisticated technology and offer elements of the experience that are high in 'fidelity'. However, other elements of the simulator may remain low in fidelity, that is, unrealistic facial, hair, and body features of manikins or simulators. The above description thoughtfully extends fidelity to include a well-trained SP who can be utterly engaging for the learner. This description raises important ideas that are well discussed and often contested in the literature. In some ways, it is the reaction and experience of the learner that becomes more important than the predetermined level of fidelity that the simulator can offer. That is, fidelity relies to some extent on the learner interpreting the simulator/simulation as 'real'.

Following this line of argument, Hamstra et al. [88] posit that there are more important considerations in health care simulation than fidelity and even discourage use of the term [88]. They propose that *functional task alignment* and *learner engagement* have more relevance than level of fidelity. Similarly, Nestel et al. [89] suggest *meaningfulness* for the learner has greater value than the notion of fidelity or realism [89]. Both discuss the importance of instructors making choices about *fit for purpose* fidelity. There are instances when high learner engagement and/or meaningfulness is achieved while working with simulators that bear little resemblance to reality. Engagement and meaningfulness is achieved by thoughtful instructional design which includes negotiating learning goals, orientating learners to the simulation/simulator, establishing a fiction contract (inviting the learner to imagine the situation is real), offering scaffolding, ensuring feedback and/or debriefing is available, and aligning learning with learner experiences. In summary, there are no hard and fast rules about what needs to be *real*, when, and for whom. What is most important is clarity about the purpose of the simulation, from which *considered* decisions about realism (or fidelity) can be made.

Limitations and Challenges in Medical Simulation

Although simulations continue to gain widespread acceptance there are several factors that limit or challenge its use that are important to consider. These factors can be grouped into two categories – operationalization issues and barriers to participant engagement. Operational issues in SBL include factors such as the cost of simulation, the increased time and complexity that accompanies efforts to design, prepare, implement, and evaluate simulation, as well as the need for specially trained individuals who can manage and oversee these activities. Faculty development is also a critical factor in success of SBL [65].

Dieckmann and colleagues found that not all learners are willing to engage during a simulation and that participants' unwillingness or failure to engage during a simulation can have detrimental implications for learning, as well as influencing the experiences of other team members [90]. There are several factors that can contribute to learners' failure to engage effectively. Some learners find simulations or the debriefing stressful and potentially intimidating, while others may be hypervigilant in trying to anticipate what might happen next [91]. Simulation participants' engagement may also be disrupted if the simulator fails or when operators and/or SPs miss a cue.

Although further research into these areas is still needed, some of these challenges and limitations can be minimised by using a structured and systematic instructional design approach as described above. In addition, below are a few final practical tips that can help guide the process.

- *Team approach to designing simulations*. Although it is possible for a single individual to design a simulation project, designing simulation using a team approach can make the design process more manageable while also improving the design of the simulation. This team may include subject matter experts representing clinical, simulation, and technical backgrounds. A clinical expert may be able to advise about the accuracy of the clinical presentation of the patient, while those with expertise in simulation or technical operations can advise on what can be practicably simulated. If simulations are to include SPs, it is important to involve those professionals who are responsible for preparing and training them. SPs can also provide critical advice and guidance about their contribution. Involving real patients and their care-givers can help to establish learning goals, character descriptions, patient perspectives, and feedback foci [92–94].

- *Allow for adequate time to design*. Designing a simulation may take longer than developing a lecture or a case study because simulations can be very complex and require more resources. Plan to start the design process well in advance of the anticipated start of the simulation or curriculum. Doing so will help ensure there is enough time to consider each of the design steps and consult with subject matter experts. These efforts can lead to a higher quality simulation.
- *Design using an iterative perspective*. Balancing goals, learning objectives, and the other aspects of simulation-based instructional design are ideally approached in an iterative way to allow what is learned during early design stages to improve what is being developed. For example, early goals for a simulation may shift based on feedback from subject matter experts. A review of the existing literature on a simulation approach may also inform the design process by shedding light on a better way to meet certain learning objectives. Pilot testing and the early stages of the implementation of a SBL activity or curriculum will almost always reveal areas for improvement or consideration.

Conclusions and Future Issues in Medical Simulation

The landscape of health care simulation continues to evolve. A recent review of 10 years of editorials in the first health care simulation journal identified five themes: (i) embedding simulation, (ii) simulation responding to clinical practice, (iii) educational considerations for simulation, (iv) research practices, and (v) communicating leadership and scholarship about the community of simulation practitioners [95]. The first three themes have been explored in this chapter. However, it is important to acknowledge the maturation of the health care simulation professional community relevant to the fourth and fifth themes. This includes the emergence of health care simulation dedicated professional societies, of special interest groups in clinical professional societies, accreditation of SBL curriculum, certification of simulation practitioners, the establishment of a simulation academy, and at least four journals strongly oriented to health care simulation education (see Box 11.5).

By looking to the past to speculate on the future, we are offered threads of continuity. Ways to embed simulation into medical curricula will persist. Regulatory bodies and health care services are likely to identify and at times mandate specific clinical practices that must be certified in simulation prior to clinicians working in real clinical settings. New simulation modalities, especially those that are screen-based and available *in situ* or at the point of care are likely to proliferate. We will see economic analyses of SBL, including studies that link patient outcomes with SBL. Additionally, critique of existing theories and the birth of new ones will guide SBL practices. SBL remains a technique to support learning with many shared principles with other educational practices within and outside clinical settings.

BOX 11.5 Suggested peer-reviewed health care simulation journals

Simulation in Health care http://journals.lww.com/simulationinhealthcare/Pages/default.aspx
Journal of the Society for Simulation in Health care (SSH)
Advances in Simulation https://advancesinsimulation.biomedcentral.com
Journal of the Society in Europe for Simulation Applied to Medicine (SESAM)
Clinical Simulation in Nursing http://www.nursingsimulation.org
Journal of the International Nursing Association for Clinical Simulation and Learning (INACSL)
BMJ STEL http://stel.bmj.com
Journal of the Association for Simulated Practice in Health care (ASPiH)
Simulation and Gaming http://journals.sagepub.com/home/sag published in association with the International Simulation and Gaming Association (ISAGA)

Acknowledgements

We gratefully acknowledge the support from the editors who have provided guidance and support throughout the development of this chapter. We thank the authors of previous versions of this chapter, including Jean Ker and Paul Bradley. The composition of this chapter would not have been possible without their support.

References

1 Lopreiato, J.O.E., Downing, D., Gammon, W., et al. (Associate Eds.) and the Terminology & ConceptsWorking Group (2016). Healthcare Simulation Dictionary, p. 50.
2 Cook, D.A., Hatala, R., Brydges, R. et al. (2011). Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. *JAMA* 306 (9): 978–988.
3 Gaba, D.M. et al. (2001). Simulation-based training in anesthesia crisis resource management (ACRM): a decade of experience. *Simulation and Gaming* 32 (2): 175–193.
4 Bradley, P. (2006). The history of simulation in medical education and possible future directions. *Medical Education* 40 (3): 254–262.
5 Ziv, A. et al. (2007). Credentialing and certifying with simulation. *Anesthesiology Clinics* 25 (2): 261–269.

6 Ziv, A. et al. (2003). Simulation-based medical education: an ethical imperative. *Academic Medicine* 78 (8): 783–788.
7 Issenberg, S.B. et al. (2005). Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Medical Teacher* 27 (1): 10–28.
8 Jeffries, P. (2005). A framework for designing, implementing, and evaluating simulations. *Nursing Education Perspectives* 26 (2): 97–104.
9 Norman, G., Dore, K., and Grierson, L. (2012). The minimal relationship between simulation fidelity and transfer of learning. *Medical Education* 46 (7): 636–647.
10 Meyer, M.N. et al. (2011). The effect of simulation on clinical performance: a junior nursing student clinical comparison study. *Simulation in Healthcare* 6 (5): 269–277.
11 McGaghie, W.C. et al. (2011). Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review of the evidence. *Academic Medicine* 86 (6): 706–711.
12 Cant, R.P. and Cooper, S.J. (2010). Simulation-based learning in nurse education: systematic review. *Journal of Advanced Nursing* 66 (1): 3–15.
13 Sa Couto, C.D. et al. (2006). A model for educational simulation of neonatal cardiovascular pathophysiology. *Simulation in Healthcare* 1 Spec no.: 4–9.
14 Tjomsland, N. and Baskett, P. (2002). Asmund S. Laerdal. *Resuscitation* 53 (2): 115–119.
15 Abrahamson, S., Denson, J.S., and Wolf, R.M. (2004). Effectiveness of a simulator in training anesthesiology residents. 1969. *Quality & Safety in Health Care* 13 (5): 395–397.
16 Good, M.L. and Gravenstein, J.S. (1989). Anesthesia simulators and training devices. *International Anesthesiology Clinics* 27 (3): 161–168.
17 Gaba, D.M. and DeAnda, A. (1988). A comprehensive anesthesia simulation environment: re-creating the operating room for research and training. *Anesthesiology* 69 (3): 387–394.
18 Barrows, H.S. (1993). An overview of the uses of standardized patients for teaching and evaluating clinical skills. AAMC. *Academic Medicine* 68 (6): 443–451. discussion 451-3.
19 Institute of Medicine committee on quality of health care (2000). *To Err Is Human: Building a Safer Health System* (ed. L.T. Kohn, J.M. Corrigan and M.S. Donaldson). National Academies Press (US). Copyright 2000 by the National Academy of Sciences. All rights reserved. Washington (DC). https://www.ncbi.nlm.nih.gov/pubmed/25077248 (accessed 22 June 2018).
20 Aron, D.C. and Headrick, L.A. (2002). Educating physicians prepared to improve care and safety is no accident: it requires a systematic approach. *Quality & Safety in Health Care* 11 (2): 168–173.
21 Mayo, P.H. et al. (2004). Achieving house staff competence in emergency airway management: results of a teaching program using a computerized patient simulator. *Critical Care Medicine* 32 (12): 2422–2427.
22 Draycott, T.J. et al. (2008). Improving neonatal outcome through practical shoulder dystocia training. *Obstetrics and Gynecology* 112 (1): 14–20.
23 Barsuk, J.H. et al. (2009). Use of simulation-based mastery learning to improve the quality of central venous catheter placement in a medical intensive care unit. *Journal of Hospital Medicine* 4 (7): 397–403.
24 Eppich, W.J., Brannen, M., and Hunt, E.A. (2008). Team training: implications for emergency and critical care pediatrics. *Current Opinion in Pediatrics* 20 (3): 255–260.
25 Wayne, D.B. et al. (2008). Simulation-based education improves quality of care during cardiac arrest team responses at an academic teaching hospital: a case-control study. *Chest* 133 (1): 56–61.
26 Willaert, W.I.M. et al. (2012). Role of patient-specific virtual reality rehearsal in carotid artery stenting. *British Journal of Surgery* 99 (9): 9.
27 Qiu, T.-M. et al. (2010). Virtual reality presurgical planning for cerebral gliomas adjacent to motor pathways in an integrated 3-D stereoscopic visualization of structural MRI and DTI tractography. *Acta Neurochirurgica* 152 (11): 1847–1857.
28 Auguste, T.C. et al. (2011). The simulation of an ex utero intrapartum procedure to extracorporeal membrane oxygenation. *Journal of Pediatric Surgery* 46 (2): 395–398.
29 Hamman, W.R. et al. (2009). Using in situ simulation to identify and resolve latent environmental threats to patient safety: case study involving a labor and delivery ward. *Journal of Patient Safety* 5 (3): 184–187.
30 Patterson, M.D. et al. (2013). In situ simulation: detection of safety threats and teamwork training in a high risk emergency department. *BMJ Quality and Safety* 22 (6): 468–477.
31 Geis, G.L. et al. (2011). Simulation to assess the safety of new healthcare teams and new facilities. *Simulation in Healthcare* 6 (3): 125–133.
32 Hayden, J. et al. (2014). The National Council of State Boards of Nursing national simulation study: a longitudinal, randomized, controlled study replacing clinical hours with simulation in prelicensure nursing education. *Journal of Nursing Regulation* 5 (2): 1–66.
33 Laack, T.A. et al. (2010). A 1-week simulated internship course helps prepare medical students for transition to residency. *Simulation in Healthcare* 5 (3): 127–132.
34 Malekzadeh, S. et al. (2011). ORL emergencies boot camp: using simulation to onboard residents. *Laryngoscope* 121 (10): 2114–2121.
35 Cohen, E.R. et al. (2013). Making July safer: simulation-based mastery learning during intern boot camp. *Academic Medicine* 88 (2): 233–239.
36 Fernandez, G.L. et al. (2012). Boot camp: educational outcomes after 4 successive years of preparatory simulation-based training at onset of internship. *Journal of Surgical Education* 69 (2): 242–248.
37 Watson, K. et al. (2012). Can simulation replace part of clinical time? Two parallel randomised controlled trials. *Medical Education* 46 (7): 657–667.
38 Palaganas, J.C., Epps, C., and Raemer, D.B. (2014). A history of simulation-enhanced interprofessional education. *Journal of Interprofessional Care* 28 (2): 110–115.
39 Bridges, D.R. et al. (2011). Interprofessional collaboration: three best practice models of interprofessional education. *Medical Education Online* 16.
40 Bearman, M., McNaughton, N., and Nestel, D. (2018). Theories informing healthcare simulation practice. In: *Healthcare Simulation Education: Evidence, Theory and Practice* (ed. D. Nestel et al.), 9–15. Chichester: Wiley.
41 Ericsson, K.A. (2015). Acquisition and maintenance of medical expertise: a perspective from the expert-performance approach with deliberate practice. *Academic Medicine* 90 (11): 1471–1486.
42 van Merriënboer, J. and Sweller, J. (2010). Cognitive load theory in health professional education: design principles and strategies. *Medical Education* 44 (1): 85–93.
43 Reedy, G. (2015). Using cognitive load theory to inform simulation design and practice. *Clinical Simulation in Nursing* 11: 350–360.
44 van Merriënboer, J. and Kirschner, P. (2007). *Ten Steps to Complex Learning: A Systematic Approach to Four Component Instructional Design*. New Jersey: Lawrence Erlbaum Associates.
45 McGaghie, W.C. (2015). Mastery learning: it is time for medical education to join the 21st century. *Academic Medicine* 90 (11): 1438–1441.
46 McGaghie, W.C. et al. (2014). A critical review of simulation-based mastery learning with translational outcomes. *Medical Education* 48 (4): 375–385.
47 McGaghie, W.C. (2015). When I say … mastery learning. *Medical Education* 49 (6): 558–559.
48 Bruner, J. (1966). *Toward a Theory of Instruction*. Cambridge: Harvard University Press.
49 Wertsch, J. and Sohmer, R. (1995). Vygotsky on learning and development. *Human Development* 38: 332–337.

50 Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational Reasearcher* 27 (2): 4–13.
51 Schon, D. (1983). *The Reflective Practitioner: How Professionals Think in Action*. London: Temple Smith.
52 Fenwick, T. and Dahlgren, M.A. (2015). Towards socio-material approaches in simulation-based education: lessons from complexity theory. *Medical Education* 49 (4): 359–367.
53 Cook, D.A. et al. (2013). Comparative effectiveness of instructional design features in simulation-based education: systematic review and meta-analysis. *Medical Teacher* 35 (1): e867–e898.
54 Reznick, R.K. and MacRae, H. (2006). Teaching surgical skills — changes in the wind. *New England Journal of Medicine* 355 (25): 2664–2669.
55 Sanko, J.S. et al. (2013). Establishing a convention for acting in healthcare simulation: merging art and science. *Simulation in Healthcare* 8 (4): 215–220.
56 Dieckmann, P., Gaba, D., and Rall, M. (2007). Deepening the theoretical foundations of patient simulation as social practice. *Simulation in Healthcare* 2 (3): 183–193.
57 Cleland, J.A., Abe, K. and Rethans, J.J. (2009). The use of simulated patients in medical education: AMEE Guide No. 42. *Med Teach.* 31 (6): 477–86.
58 Kassab, E., Kyaw Tun, J., and Kneebone, R.L. (2012). A novel approach to contextualized surgical simulation training. *Simulation in Healthcare* 7 (3): 155–161.
59 Gordon, M.S. et al. (1980). 'Harvey,' the cardiology patient simulator: pilot studies on teaching effectiveness. *The American Journal of Cardiology* 45 (4): 791–796.
60 Bearman, M. and Nestel, D. (2014). The future of simulated patient methodology. In: *Simulated Patient Methodology*, (ed. D. Nestel and M. Bearman) 145–149. Wiley.
61 Cantillon, P. et al. (2010). Simulated patient programmes in Europe: collegiality or separate development? *Medical Teacher* 32 (3): e106–e110.
62 Nestel, D. et al. (2014). Confederates in health care simulations: not as simple as it seems. *Clinical Simulation in Nursing* 10 (12): 611–616.
63 Kneebone, R. et al. (2002). An innovative model for teaching and learning clinical procedures. *Medical Education* 36 (7): 628–634.
64 Fanning, R.M. and Gaba, D.M. (2007). The role of debriefing in simulation-based learning. *Simulation in Healthcare* 2 (2): 115–125.
65 McGaghie, W.C. et al. (2010). A critical review of simulation-based medical education research: 2003–2009. *Medical Education* 44 (1): 50–63.
66 Rudolph, J.W., Raemer, D.B., and Simon, R. (2014). Establishing a safe container for learning in simulation: the role of the presimulation briefing. *Simulation in Healthcare* 9 (6): 339–349.
67 Rudolph, J.W. et al. (2006). There's no such thing as 'nonjudgmental' debriefing: a theory and method for debriefing with good judgment. *Simulation in Healthcare* 1 (1): 49–55.
68 Rudolph, J.W. et al. (2007). Debriefing with good judgment: combining rigorous feedback with genuine inquiry. *Anesthesiology Clinics* 25 (2): 361–376.
69 Shanks, D. et al. (2013). Are two heads better than one? Comparing dyad and self-regulated learning in simulation training. *Medical Education* 47 (12): 1215–1222.
70 Palaganas, J.C., Fey, M., and Simon, R. (2016). Structured debriefing in simulation-based education. *AACN Advanced Critical Care* 27 (1): 78–85.
71 Dreifuerst, K.T. Getting started with debriefing for meaningful learning. *Clinical Simulation in Nursing* 11 (5): 268–275.
72 Phrampus, P.E. and O'Donnell, J.M. (2013). Debriefing using a structured and supported approach. In: *The Comprehensive Textbook of Healthcare Simulation* (ed. A.I. Levine et al.), 73–84. New York, NY: Springer New York.
73 Imperial College London (2012). *The London Handbook for Debriefing: Enhancing Performance Debriefing in Clinical and Simulated Settings*. London: London Deanery.
74 Arora, S. et al. (2012). Objective structured assessment of debriefing (OSAD): bringing science to the art of debriefing in surgery. *Annals of Surgery* 256 (6): 982–988.
75 Runnacles, J. et al. (2014). Development of a tool to improve performance debriefing and learning: the paediatric objective structured assessment of debriefing (OSAD) tool. *Postgraduate Medical Journal* 90 (1069): 613–621.
76 Centre for Medical Simulation 2011. Debriefing Assessment for Simulation in Healthcare. http://www.harvardmedsim.org/debriefing-assesment-simulation-healthcare.php (accessed 27 June 2018).
77 Krogh, K., Bearman, M., and Nestel, D. (2015). Expert practice of video-assisted debriefing. *Clinical Simulation in Nursing* 11: 180–187.
78 Grant, D.J. and Marriage, S.C. (2012). Training using medical simulation. *Archives of Disease in Childhood* 97 (3): 255–259.
79 Grant, J. et al. (2010). Using video-facilitated feedback to improve student performance following high-fidelity simulation. *Clinical Simulation in Nursing* 6 (5): e177–e184.
80 Levett-Jones, T. and Lapkin, S. (2013). A systematic review of the effectiveness of simulation debriefing in health professional education. *Nurse Education Today* 34 (6): e58–e63.
81 Nestel, D. and Gough, S. (2018). Designing simulation-based learning activities. In: *Healthcare Simulation Education: Evidence, Theory and Practice* (ed. D. Nestel et al.), 135–142. Chichester: Wiley.
82 O'Regan, S. et al. (2016). Observer roles that optimise learning in healthcare simulation education: a systematic review. *Advances in Simulation* 1: 4.
83 Decker, S. et al. (2013). Standards of best practice: simulation standard VI: the debriefing process. *Clinical Simulation in Nursing* 9 (6): S26–S29.
84 Husebo, S., O'Regan, S., and Nestel, D. (2015). Reflective practice and its role in simulation. *Clinical Simulation in Nursing* 11 (8): 368–375.
85 Schon, D. (1987). *Educating the Reflective Practitioner*. San Francisco: Jossey-Bass.
86 Kolb, D. and Fry, R. (1975). Toward an applied theory of experiential learning. In: *Theories of Group Process* (ed. C. Cooper). London: Wiley.
87 Gough, S. (2016). The use of simulation-based education in cardio-respiratory physiotherapy. PhD thesis. Manchester Metropolitan University. https://e-space.mmu.ac.uk/617694/ (accessed 27 June 2018).
88 Hamstra, S.J. et al. (2014). Reconsidering fidelity in simulation-based training. *Academic Medicine* 89 (3): 387–392.
89 Nestel, D., Krogh, K., and Kolbe, M. (2018). Exploring realism in healthcare simulations. In: *Healthcare Simulation Education: Evidence, Theory and Practice* (ed. D. Nestel et al.), 23–28. Wiley.
90 Dieckmann, P. et al. (2012). Goals, success factors, and barriers for simulation-based learning. *Simulation and Gaming* 43 (5): 627–647.
91 Horcik, Z. et al. (2014). A phenomenological approach to novice nurse anesthetists' experience during simulation-based training sessions. *Simulation in Healthcare* 9 (2): 94–101.
92 Nestel, D. et al. (2008). Real patient involvement in role development evaluating patient focused resources for clinical procedural skills. *Medical Teacher* 30: 795–801.
93 Snow, R. (2015). Real patient participation in simulations. In: *Simulated Patient Methodology: Theory, Evidence and Practice* (ed. D. Nestel and M. Bearman), 105–109. Wiley.
94 Nestel, D. (2015). Expert's corner: standardized (simulated) patients in health professions education: a proxy for real patients? In: *Defining Excellence in Simulation Programs* (ed. J. Palaganas et al.), 394. Philadelphia: Wolters Kluwer.
95 Nestel, D. (2017). Ten years of simulation in healthcare: a thematic analysis of editorials. *Simulation in Healthcare* 12 (5): 326–331.

Further Reading

History of Simulation

Owen, H. (2016). *Simulation in Healthcare Education: An Extensive History*. Switzerland: Springer.

Theories Informing SBL

Battista, A. (2015). Activity theory and analysing learning in simulations. *Simulation and Gaming* 46 (2): 187–196.

Bearman, M., McNaughton, N., and Nestel, D. (2018). Theories informing healthcare simulation practice. In: *Healthcare Simulation Education: Evidence, Theory and Practice* (ed. D. Nestel et al.), 9–15. Chichester: Wiley.

Eppich, W. and Cheng, A. (2015). Cultural historical activity theory (CHAT) – informed debriefing for interprofessional teams. *Clinical Simulation in Nursing* 11: 383–389.

Fenwick, T. and Dahlgren, M.A. (2015). Towards socio-material approaches in simulation-based education: lessons from complexity theory. *Medical Education* 49 (4): 359–367.

Husebo, S., O'Regan, S., and Nestel, D. (2015). Reflective practice and its role in simulation. *Clinical Simulation in Nursing* 11 (8): 368–375.

Kneebone, R. (2003). Simulation in surgical training: educational issues and practical implications. *Medical Education* 37 (3): 267–277.

Nestel, D. and Bearman, M. (2015). Theory and simulation-based education: definitions, worldviews and applications. *Clinical Simulation in Nursing* 11: 349–354.

Reedy, G. (2015). Using cognitive load theory to inform simulation design and practice. *Clinical Simulation in Nursing* 11: 350–360.

Reviews in Simulation

Issenberg, S.B. et al. (2005). Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Medical Teacher* 27 (1): 10–28.

McGaghie, W.C. et al. (2010). A critical review of simulation-based medical education research: 2003–2009. *Medical Education* 44 (1): 50–63.

Cook, D.A., Hatala, R., Brydges, R. et al. (2011). Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. *JAMA* 306 (9): 978–988.

Cook, D.A., Hamstra, S.J., Brydges, R. et al. (2013). Comparative effectiveness of instructional design features in simulation-based education: systematic review and meta-analysis. *Medical Teacher.* 35 (1): e867–e898.

Design

Jeffries, P. (2005). A framework for designing, implementing, and evaluating simulations. *Nurse Education Perspectives* 26 (2): 97–104.

Nestel, D. and Gough, S. (2018). Designing simulation-based learning activities. In: *Healthcare Simulation Education: Evidence, Theory and Practice* (ed. D. Nestel et al.), 135–142. Chichester: Wiley.

Reflection and Debriefing

Cheng, A., Eppich, W., Sawyer, T., and Grant, V. (2018). Debriefing: the state of the art and science in healthcare simulation. In: *Healthcare Simulation Education: Evidence, Theory and Practice* (ed. D. Nestel, M. Kelly, B. Jolly and M. Watson). West Sussex: Wiley.

Cheng, A., Eppich, W., Grant, V. et al. (2014). Debriefing for technology-enhanced simulation: a systematic review and meta-analysis. *Medical Education* 48 (7): 657–666.

Cheng, A., Grant, V., Robinson, T. et al. (2016). The promoting excellence and reflective learning in simulation (PEARLS) approach to health care debriefing: a faculty development guide. *Clinical Simulation in Nursing* 12: 419–428.

Cheng, A., Morse, K.J., Rudolph, J. et al. (2016). Learner-centered debriefing for health care simulation education: lessons for faculty development. *Simulation in Healthcare* 11 (1): 32–40.

Eppich, W.J., Hunt, E.A., Duval-Arnould, J.M. et al. (2015). Structuring feedback and debriefing to achieve mastery learning goals. *Academic Medicine* 90 (11): 1501–1508.

Fanning, R.M. and Gaba, D.M. (2007). The role of debriefing in simulation-based learning. *Simulation in Healthcare* 2 (2): 115–125.

Rudolph, J.W., Raemer, D.B., and Simon, R. (2014). Establishing a safe container for learning in simulation: the role of the presimulation briefing. *Simulation in Healthcare* 9 (6): 339–349.

Rudolph, J.W. et al. (2006). There's no such thing as 'nonjudgmental' debriefing: a theory and method for debriefing with good judgment. *Simulation in Healthcare* 1 (1): 49–55.

Fidelity in Simulation

Hamstra, S.J. et al. (2014). Reconsidering fidelity in simulation-based training. *Academic Medicine* 89 (3): 387–392.

Nestel, D., Krogh, K., and Kolbe, M. (2018). Exploring realism in healthcare simulations. In: *Healthcare Simulation Education: Evidence, Theory and Practice* (ed. D. Nestel et al.), 23–28. Wiley.

12 Work-based Learning

Clare Morris

Institute of Health Sciences Education, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK

KEY MESSAGES

- The workplace is an important site for professional learning and development throughout a medical career.
- Insight into the theoretical concepts underpinning work-based learning reveals a range of tools and approaches that can be employed to 'revive' apprenticeship and support learning in the workplace.
- In the workplace, knowledge is embedded in everyday practice, routines, and cultures.
- There are important decisions to be made about the timing, nature, and duration of workplace-based learning elements of the curriculum.
- Learning in the workplace is supported by more knowledgeable and experienced others – near peers, clinical teachers/supervisors, and the wider health care team.
- Clinical teachers have a vital role, identifying development opportunities that arise in daily practice.
- Learners may mistakenly believe learning to be the consequence of teaching alone – the role of the clinical teacher includes making work-based learning opportunities explicit.

Introduction

William Osler wisely observed that 'He who studies medicine without books sails an uncharted sea, but he who studies medicine without patients does not go to sea at all'. The workplace has been an important site for medical learning for centuries, yet approaches to clinical teaching have perhaps received rather less attention in the medical education research literatures than one might anticipate [1]. Nevertheless, work is where health care professionals spend a considerable amount of their lives, expending a great deal of physical and emotional energy. Work is how many people define themselves, be it in relation to their role – 'I am a doctor', 'I am a surgeon' – or to their place of work – 'I work at the local teaching hospital'. For doctors, work and the workplace is where their professional learning is made real, where knowledge, skills, and practice are crafted and developed. We influence our workplace, and it influences and shapes us.

In this chapter, working–learning relationships throughout a medical career are explored. Drawing on contemporary learning theory and research, it will consider the ways in which the workplace offers a curriculum for learning and development. Medical education is the chosen exemplar but there are parallels in the education of all health care professionals. Perhaps the biggest challenge faced is making the workplace a site of shared learning, within, between, and across profession-specific groupings.

The Changing Landscape of Medical Education and Training

Health care professionals' education is dependent upon the interplay between two complex systems, the formal learning environment of universities and the clinical workplaces where students and trainees learn how to put their knowledge to use [2]. Changes in one will impact upon the other, whether intended or otherwise. For example, a call for early patient contact in the curriculum [3, 4] has resource implications for those supporting such placements in practice settings. Changes in the way clinical services are organised and delivered means patients spend less time in hospitals, traditionally the main provider of work-based experiences. This increases the emphasis on community-based medical education and further increases demands on clinicians to deliver patient care and support the development of their future colleagues [5]. Furthermore, working time regulations, while beneficent in intent, have unintended consequences, reducing the availability of supervised work-based learning experiences [6, 7]. But this complex interplay can support innovations in medical

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

education and training too. Clinical communication skills training and simulation-based education have emerged in an environment where concerns continue to be articulated about patient safety and the preparedness of graduates to undertake clinical work [2, 8].

Whilst both environments have a shared concern to produce the next generation of doctors, tensions can arise between the liberal ambitions of universities, seeking to promote academic excellence and produce world-class graduates, and the more instrumental ambitions of employers wishing to recruit staff who are able to deliver care safely, efficiently, and effectively [8]. An exemplar of this was the 2005 reform of postgraduate medical training in the UK, under the banner *Modernising Medical Careers*. The changes were fundamental, moving away from (costly) time-served models of apprenticeship to more closely regulated time-measured, outcomes-based, competency-assessed training [7, 9]. This reform has led to debates about how medical practice should be conceptualised, with implications for how it is taught and assessed. This is not unique to the UK and there is, at the time of writing, a strong move towards competency-based approaches across the world (e.g. Canada, US, Australia) with arguments increasingly made for training based upon professional judgement embedded in the concept of 'entrustable professional activity' [10–13]. An analysis of the relative merits of such models is beyond the scope of this chapter, and covered in others, but how medical education and training is framed, shapes and skews the types of work activity that are deemed to have learning value and direct the gaze of those responsible for making assessment or progression decisions in particular ways.

Whilst work-based learning is recognised as a fundamental aspect of medical education and training, its perceived status has been challenged by the privileges afforded to formal teaching. This is evidenced in protected teaching time, investment in formal teaching spaces and simulation resources within clinical environments, and investment in off-site development opportunities for trainees and their trainers. The danger of this is that work-based learning is marginalised and undervalued. Yet, as argued in the following section, the curriculum of the workplace is fundamentally important to the development of future doctors throughout the continuum of medical education.

Work-based Learning Throughout a Medical Career

Irrespective of the perpetual reshaping of health care provision, clinical workplaces continue to be a significant site for learning at all stages of a medical career [14]. Workplace learning is clearly important but inherently problematic. Alongside the debates about curriculum models and trainees' ability to access sufficient work-based learning opportunities, other concerns emerge. Students undertake work-based placements throughout their undergraduate education. Trainees rotate through a range of clinical specialties and contexts in their postgraduate years, whilst more senior medical staff continue to develop, adapt, and innovate in their practice. Whilst this is often presented as the 'continuum' of medical education, it is important to acknowledge the lived realities of the multiple points of transition that occur [15]. Key points of transition arise as medical students enter the clinical environment for the first time and as they enter the profession as new graduates. Postgraduate training involves transitions between posts and roles, towards ever greater levels of medical responsibility. The transition into consultant/attending or family physician roles is perhaps underplayed in the medical education literature, but it too requires processes of adaptation and the support of near-peers. Here the concept of '*novel disruptive elements*' [16] may be helpful, signalling differences in task, role, and context from the postgraduate stage.

These points of transition offer opportunities for personal growth and development yet at the same time can lead to increased feelings of anxiety and uncertainty [17–19]. It is argued that 'Medical educators are relatively unsophisticated at distinguishing between formative struggles that advance learning and adverse struggles that distract or impede learning' [18]. It is recognised that points of transition to greater medical responsibility lead to dips in performance and can have an impact on the working of the clinical team as they accommodate and adjust to the newcomer, whatever their stage of training [20, 21]. It has been suggested that it is helpful to view transitions in terms of '*critically intensive learning periods*' requiring explicit support and attention to workplace cultures and practices. Whilst often framed in terms of 'preparedness' for practice, critics argue it is never possible to fully prepare for something in advance because 'performance occurs in the interface between the doctor and the work itself in a specific setting' [21]. It is helpful, however, to explore where new doctors feel themselves to be most (and least) prepared and to identify strategies that seem to ease these points of transition. There is evidence that a significant proportion of medical graduates do not feel suitably prepared for clinical practice and have concerns about dealing with the day-to-day realities of working life, be it dealing with acutely ill patients, prescribing, managing their workload, or being on call [22–25]. Significantly, additional challenges, such as understanding their role and boundaries, may only become evident in their first posts [23].

The reasons why some graduates feel unprepared are therefore complex, spanning individual and organisational dimensions [17]. There is merit in placing greater attention on orientation activities, between those that support *context* performance (i.e. understanding how things are done around here) and those that support *task* performance, related to identified training needs [26], e.g. in prescribing or particular clinical skills. In the UK, it is mandatory for newly graduating doctors to shadow an outgoing junior doctor before commencing their first role. A study of the impact of this change shows that increased time spent in a shadowing-type role, along with each additional day of organisational induction activity, reduced anxiety levels of new doctors [19]. Another study supports the view that it is important to make cultures of working explicit, with new graduates desiring access to the 'unwritten stuff' and valuing time in patient-facing activity over classroom-based induction activity [17].

The amount and nature of work-based experience that undergraduates have, and their opportunity to shadow first posts before commencing employment, appears to increase their sense of readiness for new roles. This has significant implications for curriculum design. Holmboe and colleagues, for example, adopt a critical stance to the deeply entrenched approach to medical rotations for students and trainees, noting that very little empirical evidence exists on the optimal timing and duration of rotations and how transitions should be supported [27]. They question an approach based on multiple, short rotations, noting that from a sociological perspective this undermines the capacity to understand and engage in different cultures of teams, contexts, and specialties. They argue that 'the lack of ongoing supervision and longitudinal relationships with faculty profoundly conflict with growing evidence from the literature on the development of expertise' [27].

Longer, integrated clinical attachments offer a range of potential benefits, including fostering enhanced professionalism, a more holistic appreciation of the course of illnesses, greater patient-centredness, and an appreciation of how health care systems operate [18, 28, 29]. Students undertaking these types of clerkships appear to be more actively engaged in independent patient care activity at the end of a year than those undertaking more traditional block rotations [18]. Other studies illustrate the ways in which these models support relationship building – including interprofessional relationships – and position students to make the most of serendipitous learning opportunities [30]. This is further strengthened when looking at the types of work activity that new graduates and postgraduate trainees value. Opportunities to shadow near-peers when commencing a new post, reduce anxiety and help orientate the new starter to local ways of doing things [19]. Increasingly fragmented health care systems erode the continuity that historically underpinned medical apprenticeship but opportunities to follow patients through their care pathways continue to be seen as valuable in postgraduate training [31].

There is little doubt that doctors in training value work-based learning experiences and seek opportunities to engage in work-based learning. However, reform of health care, education and training, and threats to the time-served nature of training, are placing significant challenges on those in training and those who are charged with supporting and fostering their development in the workplace. Access to a range of theoretical tools and conceptions of work-based learning may be the key to overcoming such challenges [32].

Learning and Work

This section provides an overview of the ways that working and learning relationships are conceptualised, and inevitably leads to a review of theories of learning, many of which are covered in detail in Chapter 4. Sfard argues that the views we hold about learning are significant, shaping the ways in which we engage with learners and the pedagogic practices we adopt [33]. If, for example, we see *learning-as-acquisition*, we are drawn to practices that focus on individual learners seeking to ensure they acquire necessary knowledge and skills to practise. This view of learning underpins much of formal education, often aligned with the idea of technical rationality, where knowledge, held in the mind of the learner, is 'applied' to the world of practice. Although this model has been challenged [34], it persists in the minds of many as the best way to educate, leading to a continued focus on the front-loading of theory in the curriculum and a focus on individual expertise or mastery [35, 36]. However, several commentators have argued that the tendency to compare work-based learning with formal learning, or to draw on formal models of learning in the workplace, is unhelpful [35–37]. Billet, for example, challenges the idea that workplaces are unstructured, serendipitous sites of learning, suggesting instead that 'workplaces structure and routinely provide learning experiences as part of everyday work activities and through guidance from other workers. Participation in workplace tasks assists new learning and reinforces what has been learnt through further practice' [37].

This is where Sfard's second metaphor, *learning-as-participation* comes into play, framing the purposes of learning in terms of 'taking part' and 'being part' of something. This has immediate synergies with notions of apprenticeship, focusing attention on the ways in which the workplace offers up a curriculum for learning [38]. A study of doctors supporting postgraduate training illustrates how they hold onto ideas of apprenticeship, seeking out opportunities for junior doctors to work alongside their more experienced colleagues [31]. Tensions arise when those in the learning environment hold competing views of learning. Learners more familiar with formal learning environments seek out 'teaching' where their clinical teachers scan for work-based learning opportunities [31, 39]. A study of clinical teachers' conceptions of work-based learning for medical students captures both metaphors, framing it in terms of *membership* of a professional community, *partnership* when working together to deliver care, and *ownership* in terms of individual learning goals [40].

Part of the role of the clinical teacher may be to make their framings of learning more explicit to themselves and others, articulating the learning arising from work itself [38]. This is where an examination of the working–learning relationships becomes important, helping clinical teachers make sense of their current practices and offering up new ways to support learning in the clinical environment.

Another delineation, often drawn in medical education, is between 'formal' learning (medical school) and 'informal' learning (in the clinical work environment). Formal learning is typically characterised by timetables, aims and objectives, a defined curriculum, and, often, progressive, linear teaching and examinations. In contrast, informal learning, usually in the workplace, has traditionally been less valued by teachers and students, viewed as haphazard, opportunistic, and lacking any formal educational rigour, process, or structure. As noted previously, these criticisms arise in part because work-based learning is compared with the process and pedagogy of formal learning, rather than being viewed as having a pedagogy and process of its own.

Eraut [41] proposes a move away from the use of the term 'informal learning' to that of 'non-formal learning'. In so doing he proposes a typology of non-formal learning, focusing on the learner's intention to learn. Implicit learning is characterised by learning that takes place without any prior intention to learn by the learner, with the learner being unaware of the learning at the time. He contrasts this with deliberative learning, where the learner sets time aside to learn and approaches learning in a planned and purposeful way. Between these two points, Eraut describes reactive learning, which happens almost spontaneously as a result of situation and circumstance. Whilst the learning is not consciously planned, learners recognise learning opportunities, are prepared for emergent learning opportunities, and are likely to engage in brief, almost spontaneous, reflection on learning events or experiences. This distinction is helpful when we consider ways to promote work-based learning, suggesting the possibility of explicitly recognising, responding to, and valuing the learning that arises during everyday practice, and encouraging students and trainees to do the same.

Work-based learning spans all stages of medical education and training. Boud and Solomon have explored the place of work-based learning in professional education, noting that undergraduate courses now: 'Acknowledge the workplace as a site of learning and as a source for making the curriculum more relevant. As such they are a signal of the blurring distinctions between the university and the workplace' [42].

They go on to note that this 'blurring' signals the increasing legitimisation of learning outside formal academic contexts and argue that this creates both opportunities and challenges for students and trainees: 'Learning tasks are influenced by the nature of work and, in turn, work is influenced by the nature of the learning that occurs. The two are complementary. Learners are workers; workers are learners. They need to be able to manage both their roles' [43].

This dual role is particularly striking in postgraduate training, where trainees are also employees and therefore have service as well as learning commitments. Seagraves and Boyd [43] distinguish three 'links' between work and learning, as follows:

- learning *for* work
- learning *at* work
- learning *from* work.

These semantic distinctions are important, signalling, albeit implicitly, different relationships between working and learning, and the intended purposes of that learning. The question that arises is whether, for example, the medical curriculum is designed or intended to enhance working practice or professional practice. In other words, is the learning undertaken for the benefit of the employer or the individual? To some extent, this depends on the perspective from which work is viewed and its primary purpose.

Evans et al. [44] offer up three perspectives on workplace-based learning that are helpful here; industrial relations, sociological, and social learning theory.

Viewed from the *industrial relations* perspective, work is a contested activity, with constant tension between employee and employer over rights, obligations, and the prevention or misuse of employee skills and labour. Work-based learning is driven by the needs of the workplace, rather than those within it, with access to further training and development opportunities being driven by desires to promote innovation or efficiency. Work-based learning is something employers control (e.g. study leave). For medical students and trainees this is visible in issues such as access to study leave, hours of work and rotas, and the emphasis on statutory training.

Viewed from the *sociological* perspective, however, work is more of a place of and for social interaction, socialisation, and identity formation. Clearly, this is of influence in the development of professional roles and identities, where interpersonal relationships, power, authority, and status are all part of the dynamic of the workplace. In this context, how an individual trainee or student relates to others and how they are perceived by others may have a bearing on the types of learning experiences they are offered, the training they receive, and the professional identity they ultimately develop [45, 46]. How work is perceived will have an effect on how an individual views and approaches work-based learning. Work involves professional activity, but also demands that doctors, trainees, and students take on additional duties, roles, and responsibilities. Understanding this is central to work-based learning and, as will be argued later, is important if students and trainees wish to maximise their learning at work. The difficulty is that the privileging of, and overemphasis on, formal learning in undergraduate years may influence student and trainee ability to recognise the learning that is embedded in working activity.

The value of *social learning theory* in helping make sense of medical education is specifically explored below.

Theorising Work-based Learning

Theories of learning can be seen to sit within different educational schools of thought, and distinctions between them can be drawn in a range of ways (see Chapter 2). One broad distinction might be made between theories of learning that focus on individual learning (behavioural and cognitive theories) and those that see learning as a social practice, involving interactions between individuals and the contexts in which they work, learn, and play (social learning theories). Rich accounts of work- and practice-based learning can be seen in the wider learning literatures, offering depths of insight and a level of critique that goes beyond what is possible to address here [47–49]. This chapter endeavours to highlight some helpful explanatory and analytical ways of thinking specifically about work-based learning as it relates to medicine, drawing upon behavioural, cognitive, and socio-cultural schools of educational thinking.

The Behavioural School

Behavioural orientations to learning have their traditions in psychology, tending to focus on skill acquisition of individual learners, with the influence of context relatively 'silent'. Broadly speaking, behaviourism contests that learning is manifested by changes in behaviour, these changes being the result of stimuli that are external to the individual, that is, environmental factors. Hartley argues that the learning principles arising from the behaviourist school are focused on the importance of the following components:

- learning by doing
- frequent practice in varied contexts
- reinforcement as a prime motivator
- the need to have clearly defined behavioural objectives that are communicated to the learner [50].

These learning principles are readily observed in popular models of skills-based teaching in medicine. But, however desirable the models may seem, they belie the complexity of work-based and professional learning. As Hager cautions, 'The notion that job performance can be fully specifiable in advance remains a seductively attractive one' [51], suggesting it is this thinking that has underpinned the support for competency-based training.

The Cognitive School

The cognitive orientation to learning can be seen as a shift from the behaviourist's focus on observable actions in the external world to one that focuses on the internal world of the learner and changes in their thinking. Here, the focus is on the acquisition of knowledge and skills, be it as a result of input from a more able 'other' (through processes of transmission) or through engagement with one's own experiences (constructivism). Cognitive orientations encompass constructivism, socio-cognitive, and social constructivist thinking. Whilst the latter two involve explicit consideration of how learning is shaped by engagement with others, the focus for each remains upon the individual learner. Cognitive theories have been referred to as the 'dominant' paradigm [33, 36], where learning is understood as:

- residing in individual minds
- being propositional in nature
- expressible verbally or in writing
- transparent to the mind [51].

Constructivism, within the cognitive school and closely linked to the work of Jean Piaget, posits that meaning (or learning) is generated through human engagement with experience. As Scott and Palinscar note: 'Constructivists argue there is no such thing as ready-made knowledge; regardless of what a teacher does, learners construct their own knowledge. All learning … requires reinterpreting the information to be learnt or used in light of one's existing understandings and abilities' [52, p. 29].

These constructivist framings of learning are readily found in medical education research and practice. As Swanwick notes, the contested concepts of *andragogy* [53], *experiential learning* [54], and *reflection* [50, 55] have led to the almost wholesale adoption of portfolios, appraisal, and personal development planning in all walks of medical education and training [35]. These concepts, and their implications for medical education, are further developed in Chapter 2 but what is important to note here is that they present learning as an essentially unmediated activity, which happens as a result of learner engagement with their own experiences. For example, Kolb's learning cycle of concrete experience, reflection, conceptualisation, and experimentation [54] is extracted from any social context and tells us little about the types of experience that may foster this cyclical process or the role of more expert practitioners in encouraging or supervising subsequent reflection, conceptualisation, and experimentation. In relation to workplace learning therefore, it downplays the vital role of the clinical teacher in identifying, sequencing, and supervising learners' engagement with experiences in ways that are relevant to stage of development and take into account issues of patient safety and care. Likewise, they fail to offer insights into how the clinical teacher can maximise learning arising from encounters with patients. The 'cognitive apprenticeship' model [56–58] is a helpful supplement here. See Box 12.1.

BOX 12.1 FOCUS ON: Cognitive apprenticeship

The cognitive apprenticeship model derives from traditional craft apprenticeship, but makes 'thinking visible' [58] and has been shown to have potential use in the clinical teaching environment [57, 59]. Collins et al. [56] identified the following six stages to their model, which can be readily adopted in work-based teaching and is particularly useful in teaching decision-making, ethics, communication skills, and other cognitively complex areas of professional practice.

Modelling: allow the learner to observe your practice in order to build up a conceptualisation of that practice [57, 60, 61].

Coaching: watch the learner practise [57, 62, 63], offering them guidance, critique, and feedback.

Scaffolding: offer the learner more opportunities to practise, gradually and purposefully increasing the complexity of the work undertaken while slowly fading out your input [14].

Articulation: use questioning and supervision time to encourage the learner to talk you through what they are doing, why and how, providing a rationale for the approaches taken [64].

Reflection: encourage the learner to consider his or her performance analytically and to compare it with that of the expert to identify ways to further enhance his or her own performance [57, 64].

Exploration: provide opportunities for the learner to undertake new tasks and activities, prompting the learner to become independent in his or her activity and his or her thinking.

Social Cognitive Theories

Social cognitive theory can be seen as a coming together or a bridging of the behaviourists' concern with external, environmental stimuli and the cognitive theorists' concern with the internal mind. Bandura's work has been highly influential, capturing the dynamic interplay between the personal, the cognitive, and the environmental, which, combined, determine an individual's behaviour – referred to as *reciprocal determinism* [65] (see Figure 4.1). Bandura draws attention to five fundamental human capabilities:

symbolising, forethought, self-regulation, self-reflection, and vicarious learning. It can be argued, therefore, that what people think, feel, and believe will influence how they behave, with self-belief or *self-efficacy* [66] being an important determinant of motivation and achievement (see Box 4.1).

The capability for vicarious learning, through close observation of others, has been argued to be an important element of role-modelling in medicine and a powerful means of transmitting values and ways of interacting with patients, colleagues, and the wider health care team [67, 68]. Role models have an impact on professional identity formation [69], foster professionalism [44], and influence career choices [70]. They also have an influence on clinical performance, including helping develop clinical reasoning skills [71], responding appropriately to error (disclosure behaviours) [72], and offering credible feedback that others will act upon [73]. Box 12.2 offers suggestions on how to role model professional behaviour and practice.

Social Constructivism

In constructivist models the emphasis is on how the individual learner 'constructs' knowledge, that is, how the learner makes sense of new information and experiences provided by the teacher, the environment, and their wider experience. Social constructivism goes one step further, emphasising the importance of social engagement in the learning process. In other words, learners make sense of new ideas and information by engaging with others, be it their teachers, their fellow students, or others around them. An example of social constructivism influences on medical education would be problem-based learning (in its purest form).

Vygotsky [77], who developed his theories of learning from observational studies of children interacting with adults, was a key contributor to social constructivism. He noted that children were more successful in learning tasks when they engaged with adults (a more knowledgeable 'other') than when they worked independently, arguing that learning awakens developmental processes that are able to operate only when the child is interacting with their peers and others in their environment. This is significantly different from models of experiential learning explored above, which suggest that the provision of a learning experience itself leads to learning. Vygotsky drew attention to the concepts and tools teachers use to mediate the learning of another, stressing the importance of language (or shared talk) in the developmental process. Importantly, he introduced what has been described as a fundamentally new approach to the need to match learning to the learner's developmental stage, through the construct of the *zone of proximal development* [52]. This refers to what a learner can do with the support of a more knowledgeable other (be it their teacher or their peers) and is contrasted with their *zone of actual development*, i.e. what they can do independently. Box 12.3 illustrates the ways in which Vygotsky's work can shape approaches to clinical teaching.

Social cognitive theories offer insights into how to help support the development of individual learners in the clinical workplace in a range of ways, summarised in Box 12.4.

BOX 12.2 HOW TO: Be a good role model [44, 70–72, 74–76]

- Demonstrate excellent clinical practice
- Be patient-centred and empathic in your care
- Be learner-centred in your teaching
- Show respect to colleagues and value their contributions
- Offer a range of opportunities to be observed, and debrief them
- Share your thinking, such as clinical reasoning
- Discuss values, professionalism, and responses to error
- Analyse and discuss what you are modelling with learners
- Model reflection and facilitate the reflections of others
- Make the implicit more explicit (see Box 12.6)
- Allow time for discussion and debrief
- Be enthusiastic about what you do

BOX 12.3 FOCUS ON: Vygotsky

Consider the mediated nature of learning activity
Vygotsky drew attention to the 'tools' that we use to 'mediate' a learning experience, be it the language we use to explain or guide, or the tools we use to exemplify, such as handouts, test results, X-rays, and patients-as-cases. This can help us look more purposefully at informal learning encounters, and recognise and make explicit the everyday tools and learning resources we use. Close analysis of medical learning encounters reveals the importance of artefacts and oral, visual, and gestural/haptic modes of teaching, whether in sequence or synchronously [78].

Identify learners' needs and learning potential
Vygotsky drew a distinction between what he termed the *zone of actual development* (what the learner can actually do unassisted) and the *zone of proximal development* (what the learner can do with some assistance or guidance). Learning is what takes place in the zone of proximal development, where we guide, assist, support, and coach our learners. In working with students and trainees, therefore, it is important to recognise what they can do independently and then to work out how we can add value by providing input that accelerates their learning and development to the next stage.

Engage 'more knowledgeable others' in the learning process
Vygotsky saw engagement with peers and others in the environment as an essential prerequisite for learning to take place. The value of near-peer learning in the workplace should not be overlooked, particularly when considering issues of orientation and transition to a new working environment [17, 19, 79, 80].

BOX 12.4 Cognitive theories and workplace-based learning

- Focus on the development of individual learners
- Consider learning needs, learner potential, and self-efficacy
- See experience as the basis for learning
- Foreground reflective practice
- Value vicarious learning, through observation and modelling
- Offer strategies to maximise learning through a 'cognitive apprenticeship'
- Recognise the support of 'more knowledgeable others' including peers
- Signal the importance of the tools we use to mediate learning from the conceptual to the physical

Cognitive theories focus attention on individual development, on learning through experience. Social-cognitive and socio-constructivist theories begin to consider the interplay of individuals and the environment in which they operate, acknowledging the valuable contributions of others around them. However, even these theories fail to capture the complexity of work-based learning, their contributions being increasingly called into question [21, 27, 81]. Lingard, for example, argues that the current focus on competence in medical education reflects the individualistic framings of health care education, deflecting attention from the fact that 'competent individual professionals can – and do, with some regularity – combine to create an incompetent team' [82]. Her treatise to turn attention to 'collective competence' signals the value of socio-cultural theories of learning, in order to make sense of the ways in which clinical teams work, learn, and develop together.

The Socio-cultural School

Socio-cultural theories underpin the learning-as-participation metaphor [33], where the goal of learning is seen as full participation in the work of a community, for example, the work of a health care team. The key related concepts, 'Situated Learning' and 'Communities of Practice' are introduced in Chapter 2 but further explored here as a way of understanding workplace-based learning in particular. Socio-cultural theorists see the distinction between learning and working (or practice) as being artificial. They start from the assumption that learning is an integral part of our everyday experience and practice. So, for example, when we talk to our colleagues about patients we are having difficulty with, or we 'think aloud' management options on the ward round, we are engaged in both a working activity and a learning activity. Our understanding of each other, our patients, and their illnesses is influenced by the conversations we have, and this becomes part of the learning in the workplace [83]. When we encounter a complex patient or complex situation, we draw on the 'learning resources' around us (our peers, our seniors, other members of the health care team) to consider how to move forward. We might consult other types of resource, such as internet search engines, but seldom do we immediately 'rush off to be taught' to address these issues. As students develop their practice, they are learning at the same time. Learning is therefore an everyday activity and is developed by joint participation. In other words, learning is 'situated' and collective, with a shift in emphasis from a focus on the individual learner or teacher, to one that focuses on the 'team' or 'community', in particular, a community of practice [84].

In recent years, the term 'community of practice' has been adopted by those in professional education and used, often uncritically, to capture a desire to foster collaborative working, be it face-to-face or online. However, its original use was much more specific and was to capture examples of situated learning in a range of 'apprenticeship systems' observed ethnographically by Lave and Wenger [84]. Their seminal text identifies a defining feature of learning in these contexts, which is that of *legitimate peripheral participation*, described as: 'A way to speak about the relations between newcomers and old-timers, and about activities, identities, artefacts, and communities of knowledge and practice. It concerns the process by which newcomers become part of a community of practice. A person's intentions to learn are engaged and the meaning of learning is configured through the process of becoming a full participant in a socio-cultural practice' [84].

Four key ideas emerge from Lave and Wenger's work, which have particular relevance to work-based learning in medicine and newly emerging models of apprenticeship, as follows:

- learning is part of social practice
- learning takes place in communities of practice
- learning takes place through legitimate peripheral participation
- language is a central part of practice.

Let's explore these ideas in more depth. First, *learning is part of social practice*. Every day at work we encounter new situations, new patients, and new colleagues, trainees, or students that lead us to question what we know, what we do, and how and why we do it. This is clearly a 'learning' situation, although we might not always label it as such. Second, *learning takes place in communities of practice* [85], which can be identified and defined by common expertise. The practice of a surgical team or the psychiatric outreach team demonstrates this in that their practice is effective because of the shared endeavour, the collective 'team think' that leads to successful outcomes. If we compare these two 'teams', while each contains doctors, nurses, and health care professionals, they are clearly distinct in terms of the specialist work they do, the ways they do this, and the 'cultures' of their practice (how they dress, how they talk to each other and their patients, etc.). Clearly, within medicine there are many distinct communities of practice, and students and trainees need to learn how to participate within them and, indeed, across them [2, 83]. Third, *learning has a central defining process, that of 'legitimate peripheral participation'*, a process that enables the student to develop the expertise necessary to permit full access and participation in a community. When we delegate work to students and trainees, we need to ensure it allows increasing engagement in 'real' work activity, from the periphery (e.g. scrubbing up to observe the surgical procedure) to more central core activity (e.g. leading the surgical procedure). It is important to note that the relationships described between 'newcomers' and 'old-timers' here are very different from the traditional hierarchical educational models of novice to expert. This is an important distinction as it recognises the valuable contributions students and trainees can make to shaping and developing practice and the impact they can have on the workplace. Finally, *language is a central part of practice*, not only in terms of learning from talk, but rather in terms of learning to talk – a process of talking one's way into the expertise. For example, when students and trainees 'present cases', with implicit structures and cultures of doing so – 'Mrs Smith is a 55-year-old woman, who presented to A and E with a three-day history of …' – they are learning 'to talk' medicine and therefore learning medicine itself.

Medical education researchers are increasingly drawing upon socio-cultural perspectives to develop insights into complex cultural practices that typify medical education and training. These studies span the continuum of medical education and offer new ways of understanding how health care professionals learn through work activity [30, 31, 40, 61, 83, 86, 87]. Box 12.5 explores the evidence for understanding medical learning in terms of time spent in communities of practice.

Socio-cultural perspectives clearly offer some theoretical and conceptual tools to allow researchers and practitioners to analyse work-based learning. The emphasis on learning as being something that encompasses the processes of 'belonging, becoming and identity', as well as meaning making [85], immediately encourages closer attention to what others have termed the 'hidden curriculum' of medical education and training. A powerful exemplar arises from research exploring the learning value of observation from a socio-cultural perspective [62, 89]. This work suggests that observation of trainees' practice is used selectively within specialties, focused on clinical acts that are most valued. Cultural values create dilemmas: trainees value the learning that arises from being observed, yet wish to demonstrate their autonomy to those more senior and do not wish to disrupt delivery of care by taking seniors away from their own work [89]. Furthermore, associations between observation and assessment impinge upon their performance – leading them to perform a 'textbook approach' rather than that they typically adopt. This in turn impacts on the credibility and usefulness of the feedback they receive [63]. This work has immediate implications for clinical teachers. How do we encourage a culture of observation focused on development rather than one focused on assessment of performance? How do we make observation of practice a reciprocal act, one that offers learning value to both the person being observed and the person doing the observing?

A socio-cultural perspective on learning immediately broadens out the role of the clinical teacher, leading them to consider how to engage learners in meaningful work activity, developing learning and working relationships with those around them [30, 61]. This can be liberating for those experiencing the joint demands of 'service' and 'teaching roles'. Rather than thinking 'what shall I teach today?' the emphasis becomes 'what am I doing today, and how can I involve students and trainees in that?'. A further illustration of how this works in General Practice clerkships is provided by van der Zwet and colleagues, who offer up the concept of 'developmental space' to 'denote the explicit and implicit opportunities for identity development that is afforded to

BOX 12.5 WHERE'S THE EVIDENCE: For communities of practice

Lave and Wenger [84] developed their viewpoints on learning through ethnographic studies of traditional 'apprenticeships', offering new conceptual understandings of learning as a situated, social practice. Their work around situated learning and communities of practice has influenced thinking about workplace-based learning; several commentators have highlighted its potential value for making sense of medical education [35, 36, 38, 81].

The analytic concept of *legitimate peripheral participation*, for example, focuses attention upon the extent to which newcomers are enabled to become full participants in the work of a community of practice. One study of medical student learning supports the idea that clinical attachments can be seen as time spent in 'communities of practice', albeit in a descriptive sense. Opportunities for students to engage in authentic work-based activity limits the extent to which this claim can be supported in full, as does a failure to recognise the learning opportunities that are embedded in everyday work activity [83]. A study of postgraduate medical education illustrates how senior doctors see work and learning as intertwined, leading them to organise work activity along apprenticeship lines, with 'newcomers' working alongside 'old-timers' [31]. This study notes the learning value arising from following a patient through their care pathway, a point echoed loudly in studies of longitudinal integrated clerkships [30].

Rich studies of medical student learning, theorised from a socio-cultural perspective, highlight the importance of opportunities for medical students to undertake 'independent consultations', to discuss patient cases, and to observe a full range of medical activity to gain a broader understanding of how health care systems work, adding meaning to their role and developing professional identity [61]. Meaningful working–learning relationships with others within the community of practice have also been shown to be pivotal [88]: where learning is understood as membership (of a professional community) involving shared practice (such as joint problem solving), students are more likely to access learning that arises from work itself [40].

and created by students (…) which some students summarised as "*finally feeling what it's like to be a doctor*" ' [61]. Their study draws attention to attributes of the working–learning environment and the learning culture. They argue that students' emotional energy is diverted in overly competitive or exposing learning environments: relationships of trust between clinical teacher and learner provide space for development of thinking and practice. Opportunities to undertake independent consultations, to talk about consultations, and observe the wider work activity of GPs were important participatory acts. As the researchers note 'pivotal in these activities was the meaning attached to the student's role and how that was reflected in and connected with their level of independence, with being "allowed" to be a learner and with the freedom "to really be a doctor" ' [61]. The importance of acknowledging the dual status of learners as workers is particularly poignant for those involved in postgraduate medical education. This study immediately invites consideration of the impact of workplace culture on developmental opportunities, on the need for relationships with learners built on trust, and on appropriate scaffolding of meaningful work activity. It invites consideration, too, of how to make the implicit explicit (see Box 12.6).

Socio-cultural framings of work-based learning are increasingly prominent in studies of medical learning. However, it is important to sustain a critical stance. For example, Lave and Wenger's work has been criticised for its lack of attention to a number of important issues [90, 91], including:

- individual variations in accessing learning in the workplace
- the ways in which 'old-timers' continue to learn in the workplace
- the role of formal learning opportunities for workers.

BOX 12.6 HOW TO: Make the implicit 'explicit'

- Label the learning opportunities that arise spontaneously in day-to-day work [38].
- Signal expectations in terms of culture (dress code, ways of addressing members of the team and patients), practices (preferred ways of doing things and why), and participation.
- Encourage learners to articulate and discuss observed differences in culture and practice in different settings or specialties, and consider why these may occur [45, 46].
- Be clear about the importance given to learning from work and set aside time to consider lessons learned (brief and debrief) [39].
- Prime learners for observation and shadowing (using 'advanced organisers'), making clear what it is possible to learn [39, 61].
- Adopt the cognitive apprenticeship principles of 'articulation and reflection' in your approaches to clinical teaching.
- Talk about what you are role modelling and why.

Billett, in particular, argues the need to pay attention to the *invitational qualities* of the workplace, in terms of the ways in which the workplace provides and allows access to learning activities [37, 91]. A concrete example can be seen in obstetrics and gynaecology attachments for medical students, where male students are likely to access fewer hands-on learning experiences than female students, due to patient preferences [92]. More subtle variations of opportunity may exist on the basis of 'qualities' attributed to students by staff. For example, more able students who express high levels of confidence, enthusiasm, and interest in a specialty may access more learning opportunities than a shy or struggling student or one perceived to have limited insight into their own performance. In this latter case it may be that those who most need experience to develop confidence and competence are denied these experiences [93]. It is important, too, to consider the extent to which individual learners are able to recognise and respond to the learning opportunities that arise from work itself. This again highlights the importance of making those opportunities more explicit, and using strategies such as debriefing to ensure that they are maximised.

Any single theory of learning will have its limitations. Cognitive and behavioural orientations focus attention on the individual learner and their development, elevating the relationship between learner and their expert guide seen in traditional conceptions of apprenticeship. Contemporary views of apprenticeship, building on socio-cultural theories of learning, accommodate not only the reciprocal nature of learning between 'apprentice and master', but also the contributions made by others in the professional community. New formulations of apprenticeship are emerging that enable us to look beyond the novice–expert dualism and consider a social apprenticeship that much more actively recognises the contributions made by the wider community of the workplace and decisions made about how work is organised to support learning and development [90]. Fuller and Unwin argue that expansive environments are those that treat learning as part of work activity, and see personal and organisational goals as symbiotic [90, 94]. Expansive workplaces support boundary crossing activity, for example from one workplace or team to another. They also cultivate the types of dialogue and shared problem solving between team members that in turn develop expertise. They offer an analytic framework that identifies the characteristics of expansive workplaces and therefore provides insights into how to develop those that are more restrictive. (See Box 12.7.)

Whilst socio-cultural orientations offer contextual ways to analyse and re-think medical apprenticeship, including processes of professional formation and practice, they too have their limits. They downplay the role of cognition, restrict thinking about transfer, and assume fairly stable working practices. Given the degree of reform of health care systems, the latter point is salient: traditional approaches to apprenticeship are being eroded, and teamwork is being destabilised. New ways of analysing learning within and across medical learning environments are emerging, drawing on activity theory [2, 95, 96] and actor-network theory [97].

BOX 12.7 HOW TO: Foster an 'expansive' apprenticeship

- Structure opportunities to participate in different teams or communities of practice.
- Recognise the learning status of 'newcomers' – whatever their career stage.
- Allow time for newcomers to integrate and assume greater levels of responsibility.
- Treat learning as part of work, drawing on supervision, mentoring, and coaching as appropriate.
- Structure-in time away from patient care, to think and talk about patient care.
- Support off-the-job learning and development opportunities.
- Support development opportunities that go beyond those required by the immediate job.
- Allow and build-in time for educators/educational leaders to support and develop others.
- Involve learners in problem solving and decision-making, and, where safe and appropriate to do so, offer them discretion to make their own judgements and decisions.

Source: Adapted from Fuller and Unwin [90, 94].

BOX 12.8 HOW TO: Develop clinical reasoning

- Focus learner attention on the underlying scientific concepts and principles and differing contexts in which they apply [100].
- Provide students with templates to help structure their thinking (e.g. case history templates) [101].
- Encourage students to use self-explanation, rehearsing diagnostic reasoning with paper cases as well as real patients [63].
- Model your thinking, using think aloud techniques [56, 102].
- Engage in joint problem solving and shared decision-making activity [40].
- Foster 'pattern recognition' by repeat exposure to typical cases and encouraging students to compare with what they have seen before [103].
- Be a credible role model [71].

Work-based Learning and the Medical Curriculum

These different theoretical schools of thought, highlight the ways in which we can think about work-based learning. The distinctiveness of work-based learning can also be highlighted in relation to the wider curriculum, where decisions about the timing and nature of workplace-based learning elements reflect the views we hold about the nature of theory–practice relationships.

Traditional approaches to curriculum design place knowledge before practice (in the traditional preclinical/clinical model), and focus on the curriculum as transmission of a body of knowledge (e.g. paediatrics) or on the definition of desired end points and outcomes (e.g. a competent doctor). The emphasis is on the clear delineation of specific knowledge, skills, and attitudes, which are seen as measurable outputs of learning.

The risk with this model as a framework for work-based learning is that it assumes that all worthwhile attainments are visible and quantifiable [50, 51, 55, 98], and it leads to the adoption of outcomes-based, competency-assessed curriculum.

When we turn to social theories of learning, however, additional understandings of the work-based curriculum become available. Evans and Guile remind us that 'in the workplace, knowledge is embedded in routines, protocols and artefacts, as well as organisational hierarchies and power structures' [99]. The sequencing, nature, and duration of workplace-based elements in the curriculum becomes a critical factor in designing a medical curriculum, as has been illustrated earlier in relation to design decisions including early patient contact, longitudinal clinical clerkships, or shadowing. Curriculum design decisions are also shaped by our understandings of working and learning relationships. Learning for work is very different to learning from work itself. Socio-cultural theories of learning focus on the workplace and work *as* the curriculum; in other words, there is little separation between participation in working life and learning [31, 40, 93]. The relationships between learners, their peers, their assigned 'teachers', and other workers become even more critical. The trainee's learning is to a great extent built on and derived from their workplace experiences, further developed by promoting critical dialogue and thinking, reflecting both in and on outcomes and activity. This is understood as a shared activity, where both trainer and trainee seek to critically test knowledge, and the learning is continually adapted by both trainer and trainee to make sense of experiences, making connections with propositional forms of knowledge [77]. One such example is the development of clinical reasoning skills. (See Box 12.8.)

Implications for the Clinical Teacher

What does this all mean for an individual clinical teacher, trainer, or supervisor? The direct implications can be summarised as follows.

1 Learning is part of everyday social practice [84].
Implication: We need to make learning opportunities explicit to our learners. We also need to make explicit specific workplace cultures and practices to help students and trainees 'make sense' of what they see, hear, sense, and do.

2 Teams can be seen as 'communities of practice', which are identified and defined by their shared expertise [85].
Implication: We need to recognise that the whole team (community) has an important role in supporting learning and development. This can be made more explicit and opportunities to work alongside colleagues from different disciplines can be purposefully integrated.

3 Expertise is developed through participation in communities of practice [84].
Implication: We need to consider the ways in which we can meaningfully involve our students/trainees in workplace activity, including that which extends beyond direct patient care.

4 Workplaces do not always readily invite learners in and do not always offer equal opportunities to all learners [90, 104, 105].
Implication: We need to consider how we can create the right conditions for learning to take place in our workplace and to ensure certain students or groups of students/trainees are not inadvertently disadvantaged, for example, on the basis of gender or expressed future career choices.

5 Students need help to make connections between different types of knowledge [99].
Implication: We need to understand what our students/trainees already know (where they are coming from) and help them to use it to make sense of what they see, hear, and do in the workplace.

6 'Talk' is a central part of practice – learners need to 'learn to talk their way into expertise', rather than just learn from the talk of an expert [84].
Implication: We need to find strategies to help our students and trainees talk themselves into the expertise, by using techniques such as 'thinking aloud' and case-based discussion.

7 Students and trainees learn from their entire setting [72, 73, 90, 91].
Implication: We need to be aware of the workplace climate and the effect this will have on trainees. This includes how staff relate to them and to each other, how staff are valued, and how they value their work and workplace.

8 The timing, nature, and duration of workplace-based experiences need explicit consideration [106].
Implication: A balance needs to be found between the duration of attachments – sufficient to allow students to become immersed in the workplace – and providing enough attachments to give exposure to other workplaces and specialties.

9 Cultivating a sense of belonging is important to students' engagement with workplace-learning experiences in professional identity formation [30, 61].
Implication: We need to provide trainees with increasing exposure to and involvement in workplace activities and provide opportunities to discuss these in order to maximise their learning potential.

10 Learning in the workplace is an iterative process, and trainees are active participants and can influence the workplace.
Implication: We need to allow and value feedback from students and give them the opportunity to feed back their views and impressions of the workplace.

Challenges to Work-based Learning in Medical Education

The workplace has the potential to be the central site for professional learning and development in both undergraduate and postgraduate years. However, it is important not to underestimate the potential barriers to effective work-based learning. Clearly, there is an ongoing tension between working and learning, and the intensity of work for both trainers and trainees has a significant bearing on this. Trainees need opportunities to learn through working, but they also need to be released to access the formal aspects of their training and to have time to consider and discuss the learning that arises through engagement in work activity. The sheer number of learners in the workplace is also significant, and those with responsibility for organising training need to consider how to avoid the creation of learner hierarchies, which are at risk of favouring issues of seniority, power, or status over learning need. The importance of organisational and work dynamics should not be overlooked.

The shortening and fragmentation of training has eroded time-served traditional apprenticeship models. However, the importance of the relationship between trainer and trainee continues to feature highly in studies of work-based learning in medicine. Developing safe and effective supervision strategies will be key, as will be the need to draw on new models of apprenticeship that value the relationships between and contributions of other members of the team. Finally, the rise of competency-based models of education, training, and assessment brings a risk of a 'tick-box' mentality to training that must be challenged. The focus on process aspects of the curriculum will be fundamental to the ongoing development of trainees.

Throughout this chapter we have looked at the ways in which conceptions of work-based learning can shape our practice and how a critical engagement with theoretical perspectives on learning can illuminate ways in which we can best support work-based learning in medicine. Social theories of learning in particular – those that emphasise the participatory, mediated, and context-specific aspects of learning – have been argued to be best placed to support the activity of medical educators.

Leading the Way in Work-based Learning

It is clear that clinical teachers and supervisors have a pivotal role in activating the curriculum of the workplace on a day to day basis. Those with broader roles, for example overseeing the postgraduate training of individuals (such as educational supervisors or preceptors) or groups of trainees (training programme directors) can build on the strategies for clinical teachers highlighted above. Seeing induction as an opportunity to orientate learners to ways in which to make the most of the learning that happens in the workplace will be key. These points of transition are an opportunity to communicate culture and values, including the value of learning from the wider community and not just from those with formal educational roles [15].

Re-conceptualising training as an opportunity for an 'expansive apprenticeship' [90] leads to a closer consideration of the value of time spent in a range of clinical settings and the need to balance this with time 'out' for formal training and thinking time. Formative use of the strategies underpinning workplace-based assessments adds value to training time. This includes observation of practice and case-based discussions. The latter in particular provides opportunities to extend and develop clinical thinking and reasoning through the use of 'what if' type questioning; i.e. what if the patient had been 80 not 50? What if the X-ray had shown 'y' not 'z'?

There is a growing expectation that all doctors who teach will be prepared for and developed in their educational roles: this expectation has placed increasing emphasis on faculty development within clinical workplaces [107]. D'Eon et al. argue that if we see teaching as a social practice, faculty development activity should be concerned with fostering a critical dialogue between those who share responsibility for medical education and training [108]. This is particularly pertinent given the amount of reform experienced in medical education and training and the conservative tendency of clinical communities of practice, who may seek to hold on to traditional practices even when the circumstances have changed [87]. Faculty developers themselves should offer insights from workplace-based learning theories as a way to identify teaching strategies that are aligned to 'clinic' rather than the classroom. Faculty development has the potential to model work-based learning methods; there is increasing evidence that longitudinal approaches drawing on coaching, mentoring, and peer observation of teaching are particularly valued [87, 107, 109].

Conclusion

The value and importance of work-based learning has never been clearer, nor the challenges faced greater. Changes to the ways in which health care is delivered has implications for the education and training of medical professionals. As traditional, time-served models are eroded, there is a need to re-think and develop new models of apprenticeship. This chapter has offered some ways of understanding working–learning relationships throughout a medical career. Conceptions of work-based learning, alongside a closer examination of educational theories and perspectives, may provide us with the tools to analyse and develop approaches to training. Whilst it will be important to hold on firmly to the time-honoured features of apprenticeship, rooted in a developmental relationship between trainer and trainee, there are opportunities to embrace opportunities for learning as active co-participants in patient care, supported by peers and colleagues within the communities in which we work and learn.

Acknowledgements

I would like to acknowledge the contribution made to earlier editions of this chapter by the late Professor David Blaney (1956–2015) a passionate and innovative medical educator and keen proponent of work-based learning.

References

1 Peters, M. and ten Cate, O. (2014). Bedside teaching in medical education: a literature review. *Perspectives on Medical Education* 3 (2): 76–88.

2 Morris, C. (2012). From classroom to clinic: an activity theory perspective. In: *Work-Based Learning in Clinical Settings: Insights from Socio-Cultural Perspectives* (ed. V. Cook, C. Daly and M. Neman), 85–102. London: Radcliffe.

3 Dornan, T., Littlewood, S., Margolis, S.A. et al. (2006). How can experience in clinical and community settings contribute to early medical education? A BEME systematic review. *Medical Teacher* 28 (1): 3–18.

4 Yardley, S., Littlewood, S., Margolis, S.A. et al. (2010). What has changed in the evidence for early experience? Update of a BEME systematic review. *Medical Teacher* 32 (9): 740–746.

5 Skochelak, S.E. (2010). A decade of reports calling for change in medical education: what do they say? *Academic Medicine* 85 (9): S26–S33.

6 Temple, J. (2010). *Time for Training: A Review of the Impact of the European Working Time Directive on the Quality of Training*. London: Medical Education England https://www.londonmedicine.ac.uk/wp-content/uploads/2016/11/Temple%20Report_Time%20for%20Training_%20May%202010.pdf (accessed 25 June 2018).

7 Tooke, J. (2008). *Aspiring to Excellence. Final Report of the Independent Inquiry into Modernising Medical Careers*. London: MMC Inquiry.

8 Cooke, M., Irby, D.M., Sullivan, W., and Ludmerer, K.M. (2006). American medical education 100 years after the Flexner report. *New England Journal of Medicine* 355 (13): 1339–1344.

9 Morris, C. (2011). From time-served apprenticeship to time-measured training: new challenges for postgraduate medical education. Thesis submitted in fulfilment of Doctorate of Education (EdD), Institute of Education, University of London.

10 Gwee, M.C. (2011). Medical and health care professional education in the 21st century: institutional, national and global perspectives. *Medical Education* 45 (1): 25–28.

11 Whitehead, C.R. and Kuper, A. (2015). Competency-based training for physicians: are we doing no harm? *Canadian Medical Association Journal* 187 (4): E128–E129.

12 ten Cate, T.J.O., Snell, L., and Carraccio, C. (2010). Medical competence: the interplay between individual ability and the health care environment. *Medical Teacher* 32 (8): 669–675.

13 ten Cate, O. (2013). Competency-based education, entrustable professional activities, and the power of language. *Journal of Graduate Medical Education* 5 (1): 6–7.

14 Morris, C. and Behrens, M. (2013). Work-based learning. In: *The Oxford Textbook of Medical Education* (ed. K. Walsh), 209–220. Oxford: Oxford University Press.

15 Teunissen, P.W. and Westerman, M. (2011). Opportunity or threat: the ambiguity of the consequences of transitions in medical education. *Medical Education* 45 (1): 51–59.

16 Westerman, M. and Teunissen, P. (2013). Transitions in medical education. In: *The Oxford Textbook of Medical Education* (ed. K. Walsh), 372–382. Oxford: Oxford University Press.

17 Bullock, A., Fox, F., Barnes, R. et al. (2013). Transitions in medicine: trainee doctor stress and support mechanisms. *Journal of Workplace Learning* 25 (6): 368–382.

18 O'Brien, B., Cooke, M., and Irby, D.M. (2007). Perceptions and attributions of third-year student struggles in clerkships: do students and clerkship directors agree? *Academic Medicine* 82 (10): 970–978.

19 Van Hamel, C. and Jenner, L.E. (2015). Prepared for practice? A national survey of UK foundation doctors and their supervisors. *Medical Teacher* 37 (2): 181–8.

20 Kilminster, S., Zukas, M., Quinton, N., and Roberts, T. (2010). Learning practice? Exploring the links between transitions and medical performance. *Journal of Health Organization and Management* 24 (6): 556–570.

21 Kilminster, S., Zukas, M., Quinton, N., and Roberts, T. (2011). Preparedness is not enough: understanding transitions as critically intensive learning periods. *Medical Education* 45 (10): 1006–1015.

22 Goldacre, M.J., Lambert, T.W., and Svirko, E. (2014). Foundation doctors' views on whether their medical school prepared them well for work: UK graduates of 2008 and 2009. *Postgraduate Medical Journal* 90 (1060): 63–68.

23 Illing, J.C., Morrow, G.M., nee Kergon, C.R.R. et al. (2013). Perceptions of UK medical graduates' preparedness for practice: a multi-Centre qualitative study reflecting the importance of learning on the job. *BMC Medical Education* 13 (1): 34.

24 Tallentire, V.R., Smith, S.E., Wylde, K., and Cameron, H.S. (2011). Are medical graduates ready to face the challenges of foundation training? *Postgraduate Medical Journal* 87 (1031): 590–595.

25 Tallentire, V.R., Smith, S.E., Skinner, J., and Cameron, H.S. (2012). The preparedness of UK graduates in acute care: a systematic literature review. *Postgraduate Medical Journal* 88 (1041): 365–371.

26 Wanous, J.P. and Reichers, A.E. (2000). New employee orientation programs. *Human Resource Management Review* 10 (4): 435–451.

27 Holmboe, E., Ginsburg, S., and Bernabeo, E. (2011). The rotational approach to medical education: time to confront our assumptions? *Medical Education* 45 (1): 69–80.

28 Hirsh, D., Gaufberg, E., Ogur, B. et al. (2012). Educational outcomes of the Harvard Medical School – Cambridge integrated clerkship: a way forward for medical education. *Academic Medicine* 87 (5): 643–650.

29 Walters, L., Greenhill, J., Richards, J. et al. (2012). Outcomes of longitudinal integrated clinical placements for students, clinicians and society. *Medical Education* 46 (11): 1028–1041.

30 Roberts, C., Daly, M., Held, F., and Lyle, D. (2017). Social learning in a longitudinal integrated clinical placement. *Advances in Health Sciences Education* 22 (4): 1011–1029.

31 Skipper, M., Nøhr, S.B., Jacobsen, T.K., and Musaeus, P. (2016). Organisation of workplace learning: a case study of paediatric residents' and consultants' beliefs and practices. *Advances in Health Sciences Education* 21 (3): 677–694.

32 Reeves, S. and Hean, S. (2013). Editorial. Why we need theory to help us better understand the nature of interprofessional education, practice and care. *Journal of Interprofessional Care* 27: 1–3.

33 Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational Researcher* 27 (2): 4–13.

34 Schön, D. (1983). *The Reflective Practitioner*. New York: Basic Books.

35 Swanwick, T. (2005). Informal learning in postgraduate medical education: from cognitivism to 'culturism'. *Medical Education* 39 (8): 859–865.

36 Bleakley, A. (2006). Broadening conceptions of learning in medical education: the message from teamworking. *Medical Education* 40 (2): 150–157.

37 Billett, S. (2001). *Learning in the Workplace: Strategies for Effective Practice*. Crows Nest, NSW: Allen and Unwin.

38 Swanwick, T. and Morris, C. (2010). Shifting conceptions of learning in the workplace. *Medical Education* 44 (6): 538–539.

39 Walton, J.M. and Steinert, Y. (2010). Patterns of interaction during rounds: implications for work-based learning. *Medical Education* 44 (6): 550–558.

40 Strand, P., Edgren, G., Borna, P. et al. (2015). Conceptions of how a learning or teaching curriculum, workplace culture and agency of individuals shape medical student learning and supervisory practices in the clinical workplace. *Advances in Health Sciences Education* 20 (2): 531–557.

41 Eraut, M. (2000). Non-formal learning, implicit learning and tacit knowledge. In: *The Necessity of Informal Learning* (ed. F. Coffield), 12–21. Bristol: The Policy Press.

42 Boud, D. and Solomon, N. ed. (2001). *Work-Based Learning: A New Higher Education?* Buckingham: The Society for Research into Higher Education and Open University Press.

43 Seagraves, L. and Boyd, A. (1996). *Supporting Learners in the Workplace: Guidelines for Learning Advisers in Small and Medium Sized Companies*. Stirling: University of Stirling.

44 Evans, K., Hodkinson, P., Rainbird, H., and Unwin, L. (2007). *Improving Workplace Learning*. Oxford: Routledge.

45 Monrouxe, L.V., Rees, C.E., and Hu, W. (2011). Differences in medical students' explicit discourses of professionalism: acting, representing, becoming. *Medical Education* 45 (6): 585–602.

46 Gordon, J., Markham, P., Lipworth, W. et al. (2012). The dual nature of medical enculturation in postgraduate medical training and practice. *Medical Education* 46 (9): 894–902.

47 Malloch, M., Cairns, L., Evans, K., and O'Connor, B.N. ed. (2011). *The SAGE Handbook of Workplace Learning*. London: Sage Publications.

48 Higgs, J., Barnett, R., Billett, S. et al. (2012). *Practice-Based Education: Perspectives and Strategies*. Rotterdam: Sense Publications.

49 Fenwick, T.J. and Nerland, M. (2014). *Reconceptualising Professional Learning. Sociomaterial Knowledges, Practices and Responsibilities*. Abingdon: Routledge.

50 Hartley, J. (2008). *Learning and Studying: A Research Perspective*. Abingdon: Routledge.

51 Hager, P. (2004). The conceptualization and measurement of learning at work. In: *Workplace Learning in Context* (ed. H. Rainbird, A. Fuller and A. Munro). London: Routledge.

52 Scott, S.E. and Palincsar, A.S. (2009). The influence of constructivism on teaching and learning in classrooms. In: *Knowledge, Values and Educational Policy: A Critical Perspective* (ed. H. Daniels, H. Lauder and J. Porter), 29–43. London: Routledge.

53 Knowles, M. (1973). *The Adult Learner: A Neglected Species*. Houston, TX: Gulf Publishing.

54 Kolb, D. (1984). *Experiential Learning*. Englewood Cliffs, NJ: Prentice-Hall.

55 Hager, P. (2011). Theories of workplace learning. In: *The SAGE Handbook of Workplace Learning* (ed. M. Malloch, L. Cairns, K. Evans and B.N. O'Connor), 17–31. London: Sage Publications.

56 Collins, A., Brown, J.S., and Holum, A. (1991). Cognitive apprenticeship: making thinking visible. *American Educator* 15 (3): 6–11.

57 Stalmeijer, R.E., Dolmans, D.H.J.M., Wolfhagen, I.H.A.P., and Scherpbier, A.J.J.A. (2009). Cognitive apprenticeship in clinical practice: can it stimulate learning in the opinion of students? *Advances in Health Sciences Education* 14 (4): 535–546.

58 Woolley, N.N. and Jarvis, Y. (2007). Situated cognition and cognitive apprenticeship: a model for teaching and learning clinical skills in a technologically rich and authentic learning environment. *Nurse Education Today* 27 (1): 73–79.

59 Lyons, K., McLaughlin, J.E., Khanova, J., and Roth, M.T. (2017). Cognitive apprenticeship in health sciences education: a qualitative review. *Advances in Health Sciences Education Theory and Practice* 22 (3): 723–739.

60 Liljedahl, M., Boman, L.E., Fält, C.P., and Bolander Laksov, K. (2015). What students really learn: contrasting medical and nursing students' experiences of the clinical learning environment. *Advances in Health Sciences Education* 20 (3): 765–779.

61 van der Zwet, J., Zwietering, P.J., Teunissen, P.W. et al. (2011). Workplace learning from a socio-cultural perspective: creating

developmental space during the general practice clerkship. *Advances in Health Sciences Education* 16 (3): 359–373.

62 Fromme, H.B., Karani, R., and Downing, S.M. (2009). Direct observation in medical education: a review of the literature and evidence for validity. *Mount Sinai Journal of Medicine* 76 (4): 365–371.

63 LaDonna, K.A., Hatala, R., Lingard, L. et al. (2017). Staging a performance: learners' perceptions about direct observation during residency. *Medical Education* 51 (5): 498–510.

64 Chamberland, M., Mamede, S., St-Onge, C. et al. (2015). Self-explanation in learning clinical reasoning: the added value of examples and prompts. *Medical Education* 49 (2): 193–202.

65 Bandura, A. and Walters, R.H. (1977). *Social Learning Theory*. Englewood Cliffs, NJ: Prentice-Hall.

66 Bandura, A. (1997). *Self-Efficacy: The Exercise of Control*. New York: WH Freeman.

67 Kenny, N.P., Mann, K.V., and MacLeod, H. (2003). Role modeling in physicians' professional formation: reconsidering an essential but untapped educational strategy. *Academic Medicine* 78 (12): 1203–1210.

68 Cruess, R.L., Cruess, S.R., Boudreau, J.D. et al. (2015). A schematic representation of the professional identity formation and socialization of medical students and residents: a guide for medical educators. *Academic Medicine* 90 (6): 718–725.

69 Wong, A. and Trollope-Kumar, K. (2014). Reflections: an inquiry into medical students' professional identity formation. *Medical Education* 48 (5): 489–501.

70 Passi, V. and Johnson, N. (2016). The impact of positive doctor role modeling. *Medical Teacher* 38 (11): 1139–1145.

71 Ajjawi, R. and Higgs, J. (2008). Learning to reason: a journey of professional socialisation. *Advances in Health Sciences Education* 13 (2): 133–150.

72 Martinez, W., Hickson, G.B., Miller, B.M. et al. (2014). Role-modeling and medical error disclosure: a national survey of trainees. *Academic Medicine* 89 (3): 482–489.

73 Watling, C., Driessen, E., van der Vleuten, C.P.M., and Lingard, L. (2012). Learning from clinical work: the roles of learning cues and credibility judgements. *Medical Education* 46 (2): 192–200.

74 der Leeuw HGAR, J.-v., van Dijk, N., van Etten-Jamaludin, F.S., and Wieringa-de Waard, M. (2013). The attributes of the clinical trainer as a role model: a systematic review. *Academic Medicine* 88 (1): 26–34.

75 Sternszus, R., Macdonald, M.E., and Steinert, Y. (2016). Resident role modeling: "it just happens". *Academic Medicine* 91 (3): 427–432.

76 Passi, V., Johnson, S., Peile, E. et al. (2013). Doctor role modelling in medical education: BEME Guide No. 27. *Medical Teacher* 35 (9): e1422–e1436.

77 Vygotsky, L. (1978). *Mind in Society: The Development of the Higher Psychological Processes*. Cambridge, MA: Harvard University Press.

78 Pimmer, C., Pachler, N., and Genewein, U. (2013). Reframing clinical workplace learning using the theory of distributed cognition. *Academic Medicine* 88 (9): 1239–1245.

79 Burgess, A., McGregor, D., and Mellis, C. (2014). Medical students as peer tutors: a systematic review. *BMC Medical Education* 14 (1): 115.

80 ten Cate, O. and Durning, S. (2007). Peer teaching in medical education: twelve reasons to move from theory to practice. *Medical Teacher* 29 (6): 591–599.

81 Cook, V., Daly, C., and Newman, M. ed. (2012). *Work-Based Learning in Clinical Settings – Insights from Sociocultural Perspectives*. Oxford: Radcliffe Medical.

82 Hodges, B.D. and Lingard, L. ed. (2013). *The Question of Competence: Reconsidering Medical Education in the Twenty-First Century*. Ithaca, NY: Cornell University Press.

83 Morris, C. (2012). Re-imagining 'the firm': clinical placements as time spent in communities of practice. In: *Work-Based Learning in Clinical Settings – Insights from Sociocultural Perspectives* (ed. V. Cook, C. Daly and M. Newman), 11–25. Oxford: Radcliffe Medical.

84 Lave, J. and Wenger, E. *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.

85 Wenger, E. (1998). *Communities of Practice: Learning, Meaning and Identity*. Cambridge: Cambridge University Press.

86 Cook, M.J. and Leathard, H.L. (2004). Learning for clinical leadership. *Journal of Nursing Management* 12 (6): 436–444.

87 Cantillon, P., D'Eath, M., De Grave, W. and Dornan, T. (2016). How do clinicians become teachers? A communities of practice perspective. *Advances in Health Sciences Education* 21 (5): 991–1008.

88 Berkhout, J.J., Helmich, E., Teunissen, P.W. et al. (2017). How clinical medical students perceive others to influence their self-regulated learning. *Medical Education* 51 (3): 269–279.

89 Watling, C., LaDonna, K.A., Lingard, L. et al. (2016). 'Sometimes the work just needs to be done': socio-cultural influences on direct observation in medical training. *Medical Education* 50 (10): 1054–1064.

90 Fuller, A. and Unwin, L. (2006). Expansive and restrictive learning environments. In: *Improving Workplace Learning* (ed. K. Evans, P. Hodkinson, H. Rainbird and L. Unwin), 27–48. London: Routledge.

91 Billett, S. (2002). Toward a workplace pedagogy: guidance, participation, and engagement. *Adult Education Quarterly* 53 (1): 27–43.

92 Higham, J. and Steer, P.J. (2004). Gender gap in undergraduate experience and performance in obstetrics and gynaecology: analysis of clinical experience logs. *British Medical Journal* 328 (7432): 142–143.

93 Sheehan, D., Wilkinson, T.J., and Billett, S. (2005). Interns' participation and learning in clinical environments in a New Zealand hospital. *Academic Medicine* 80 (3): 302–308.

94 Fuller, A. and Unwin, L. (2011). Workplace learning and the organization. In: *The SAGE Handbook of Workplace Learning* (ed. M. Malloch, L. Cairns, K. Evans and B.N. O'Connor), 46–59. London: Sage Publications.

95 Ajjawi, R., Rees, C., and Monrouxe, L.V. (2015). Learning clinical skills during bedside teaching encounters in general practice: a video-observational study with insights from activity theory. *Journal of Workplace Learning* 27 (4): 298–314.

96 Reid, A.-M., Ledger, A., Kilminster, S., and Fuller, R. (2015). Can the tools of activity theory help us in advancing understanding and organisational change in undergraduate medical education? *Advances in Health Sciences Education* 20 (3): 655–668.

97 Bleakley, A. (2012). The proof is in the pudding: putting actor-network-theory to work in medical education. *Medical Teacher* 34 (6): 462–467.

98 Scott, D. (2008). *Critical Essays on Major Curriculum Theorists*. Abingdon: Routledge.

99 Evans, K. and Guile, D. (2012). Putting different forms of knowledge to work in practice. In: *Practice-Based Education: Perspectives and Strategies* (ed. J. Higgs, R. Barnett, S. Billett et al.), 113–130. Rotterdam: Sense Publications.

100 Kulasegaram, K.M., Chaudhary, Z., Woods, N. et al. (2017). Contexts, concepts and cognition: principles for the transfer of basic science knowledge. *Medical Education* 51 (2): 184–195.

101 Brown, J. (2010). Transferring clinical communication skills from the classroom to the clinical environment: perceptions of a group of medical students in the United Kingdom. *Academic Medicine* 85 (6): 1052–1059.

102 Delany, C. and Golding, C. (2014). Teaching clinical reasoning by making thinking visible: an action research project with allied health clinical educators. *BMC Medical Education* 14 (1): 20.

103 Bowen, J.L. (2006). Educational strategies to promote clinical diagnostic reasoning. *New England Journal of Medicine* 355 (21): 2217–2225.

104 Billett, S. (2001). Learning through work: workplace affordances and individual engagement. *Journal of Workplace Learning* 13 (5): 209–214.

105 Billett, S. (2004). Workplace participatory practices: conceptualising workplaces as learning environments. *Journal of Workplace Learning* 16 (6): 312–324.

106 Evans, K., Guile, D., and Harris, J. (2011). Rethinking work-based learning: for education professionals and professionals who educate. In: *The SAGE Handbook of Workplace Learning* (ed. M. Malloch, L. Cairns, K. Evans and B.N. O'Connor), 149–162. London: Sage Publications.

107 Steinert, Y., Mann, K., Anderson, B. et al. (2016). A systematic review of faculty development initiatives designed to enhance teaching effectiveness: a 10-year update: BEME Guide No. 40. *Medical Teacher* 38 (8): 769–786.

108 D'Eon, M., Overgaard, V., and Harding, S.R. (2000). Teaching as a social practice: implications for faculty development. *Advances in Health Sciences Education* 5 (2): 151–162.

109 Leslie, K., Baker, L., Egan-Lee, E. et al. (2013). Advancing faculty development in medical education: a systematic review. *Academic Medicine* 88 (7): 1038–1045.

Further Reading and Resources

Cook, V., Daly, C., and Newman, M. (2012). *Work-Based Learning in Clinical Settings – Insights from Sociocultural Perspectives*. Oxford: Radcliffe Medical.

Higgs, J., Barnett, R., Billett, S. et al. (2012). *Practice-Based Education: Perspectives and Strategies*. Rotterdam: Sense Publications.

Malloch, M., Cairns, L., Evans, K., and O'Connor, B.N. ed. (2011). *The SAGE Handbook of Workplace Learning*. London: Sage Publications.

The infed encylopaedia of thinkers at http://infed.org/mobi/thinkers-and-innovators. 'infed' was established in 1995 at the YMCA George Williams College in London as an open, not-for-profit resource to provide a space for people to explore education, learning and social action – in particular the theory and practice of informal education, community learning and development, specialist education, social pedagogy, and lifelong learning.

13 Supervision, Mentoring, and Coaching

John Launer
Health Education England, London, UK

KEY MESSAGES

- Supervision and other one-to-one learning encounters occur throughout medical education and the medical career cycle. They often form the most important part of professional learning and act as the foundation of reflective practice. They may play an essential part in motivating and retaining practitioners, and in preventing stress and burnout.
- Supervision and related activities can occur formally or informally; as part of a regulatory requirement or on a voluntary basis; in a hierarchical relationship or between peers. It can take place as part of training, line management, remediation, or peer support.
- One-to-one encounters can take a variety of forms and have a variety of names, including clinical supervision, educational supervision, preceptorship, mentoring, and coaching. There is also an overlap with clinical teaching. The boundaries between these activities are sometimes unclear. Terminology can be confusing, and also varies between countries, but definitions are probably less important than understanding the context and purpose of any encounter.
- The focus of one-to-one encounters in medical education is often wider than clinical cases, and may include the contexts of these cases (e.g. professional networks) and career choices. Sometimes the focus will move around between these.
- Whatever their context and form, all one-to-one encounters contain an element of professional development and an element of performance monitoring or standard setting. These elements may be present either implicitly or explicitly. The emphasis on one or the other will vary greatly, depending on the circumstances, and the relationships and skills of the people concerned.
- Good supervision, mentoring, and coaching depend on the same set of skills. These include affirmation, emotional attunement, awareness of external requirements and standards, and ability to question and challenge people appropriately. Skills for these activities should be taught and learned.

Introduction

Traditionally, medicine has been an area where the emphasis has been on didactic training, often in groups, rather than facilitated individual learning. This is understandable. Doctors need to acquire, and to keep acquiring, tremendous amounts of factual knowledge. At the same time, most doctors can probably remember one-to-one conversations from every stage of their undergraduate and postgraduate careers that helped or influenced them. Such individual teaching, supervision, coaching, or mentorship – whatever name it was given – may have been the most important part of their learning. For those who have always provided this, it may have been one of the most gratifying parts of their work. This chapter is about such encounters.

Approaches to teaching and learning in medical careers have undergone a transformation in many places and this has brought one-to-one encounters to the fore in two ways. First, the structures within which such one-to-one support takes place are likely to be less 'ad hoc' and more organised or rigorous. In an increasing number of settings, one-to-one support is being placed at the centre of professional learning. Second, the kinds of encounter that are taking place within medicine and medical education are often different from how they might have been in the past. They are likely to be more informal and dialogical in style. They may go well beyond informational input, and involve discursive and wide-ranging consideration of cases and work issues. They are therefore coming closer to the notion of 'bringing forth' understanding. In the UK, for example, this now includes the following:

- educational supervision for doctors in specialty training [1, 2]
- tutorials for trainee general practitioners (GPs) and psychiatrists [3]
- appraisal of doctors [4]
- mentoring for hospital specialists or GPs [5]
- executive coaching for senior doctors [6]
- remedial work done within postgraduate medical education [7].

The factors that have led to such changes are many, but the following are probably the most significant:

- External requirements have put quality and performance on the medical agenda, along with clinical

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

and educational governance; this in turn has required doctors to engage more fully with their trainees and colleagues, and to address their competencies openly [8].

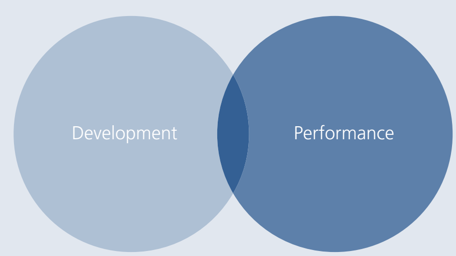
- Education in many professional fields has undergone widespread change, with an emphasis on adult learning as opposed to traditional pedagogy, and on reflective practice rather than on the acquisition of facts [9–11].
- Social developments have made both patients and learners more assertive, and less likely to accept authoritative or directive instruction without question or challenge; 'patient-centred' and 'learner-centred' styles are becoming more the norm [12].
- Doctors now work more closely with other professionals such as nurses, social workers, and psychologists, who have all established one-to-one supervision as the mainstay of their basic training and continuing professional development [13–16].
- There is an increasing need for lifelong learning in medicine, and this has led to some loss of the boundary between initial training and continuous professional development, including work-based learning [17–19].
- Many doctors have a greater awareness of medical ethics and new areas of exploration, such as complexity [20], social constructionism [21], whole-systems approaches [22], and narrative [23]. They are therefore more aware that clinical cases may involve uncertainty at many levels and be open to multiple interpretations and possible solutions.

What does Supervision Mean?

There are many definitions of supervision. Different authors take different approaches, largely according to their own professional backgrounds, experience, and agendas. In the literature, supervision is sometimes treated as entirely different to mentoring and coaching, but at other times there is an emphasis on how they share the same skills and objectives. It is easy to get bogged down in semantics and lose sight of the important principles for one-to-one learning encounters. This section attempts to cut through the jargon by looking at the history of the term and focusing on some key principles.

The term 'supervision' originated in other professions, long before it was recognised in medicine. It has been used for many years in the mental health world and nursing to mean regular, structured, extended encounters aimed at reflecting on casework [24]. Over time, people in these professions also came to take a wider view of its meaning, suggesting that it should cover any encounter that provides support in a clinical context, whether formal or informal, hierarchical or non-hierarchical, part of a training programme or outside one. Butterworth, for example, offers this very inclusive definition from a nursing perspective: 'An exchange between professionals to enable the development of professional skills' [25]. Burton and Launer, looking at primary health care, define it as 'facilitated learning in relation to live practical issues' [26]. Clark et al. have suggested that supervision should be considered an umbrella term, covering all one-to-one professional encounters and thus including mentoring and coaching, as well as activities that include an element of management, training, assessment, or remediation [27]. In line with this, the generic term 'supervision' is used in this chapter as a shorthand to cover *all* one-to-one conversations aimed at promoting competence and reflective practice – although more precise descriptions are also given below for specific types of encounter such as clinical or educational supervision. Within the medical profession, the term 'supervision' arrived relatively recently and is often used more narrowly, mainly in reference to making sure trainees perform competently. However, doctors are increasingly taking the same wider view as other non-medical colleagues, and recognising that supervision takes place – or should do – in many other contexts and guises.

An issue that often arises in discussions about supervision among doctors is whether it is principally about monitoring and standard-setting or about personal and professional development. There is an understandable emphasis in medicine on the legal and governance responsibilities carried by supervisors rather than the developmental and pastoral aspects of the role, although that too is changing. There is also an inherent ambiguity in the word supervision in English, since it can imply 'looking over someone's shoulder' as well as 'looking after someone'. The approach adopted in this chapter is to recognise that the word can carry either kind of meaning, or both, depending on the people involved and the context, and to regard the ambiguity as helpful rather than confusing. In practical terms, it may be useful to think of supervision in terms of two overlapping circles labelled 'development' and 'performance' (see Figure 13.1). On some occasions, the focus will be entirely on personal and professional development. In these circumstances, the supervisor will be able to take performance standards for granted and concentrate on extending the knowledge and reflective capacity of the other person. This occurs, for example, when a doctor is helping an experienced and proficient trainee to think through a highly complex case and consider a range of different management options. On other occasions, the supervisor has to move almost entirely towards directive teaching. This may happen when a supervisee presents a case for discussion but exposes a huge ignorance of basic knowledge in doing so.



Supervision can focus on either development or performance or both, depending on the context and the agreed relationship between supervisor and supervisee. Some supervision is entirely about development (e.g. mentoring), in other cases it is entirely about performance (e.g. remedial supervision). In most cases the circles overlap to a greater or lesser extent.

Figure 13.1 The two faces of supervision.

Even in these instances, however, it is clear that supervision will always have to pay attention to both development and performance, hence the circles overlap. Even when two experienced peers are discussing a case, both colleagues will assume certain norms that define the options that are worth considering – even if the norms remain unspoken during the supervision itself. Similarly, the supervisor who pulls up a poor trainee on the facts is still working within a framework of development and aiming to foster a relationship where standards can eventually be assumed.

What is Supervision For?

A starting point for understanding the purposes of supervision is to make a distinction between straightforward didactic teaching, which is unidirectional, and supervision, which is interactional. While didactic teaching is simply aimed at imparting facts, figures, rules, and guidelines, supervision focuses on the 'swampy lowlands' of everyday practice [10] – the complex realities that professionals face all the time and for which the textbooks may or may not have answers.

There is an inseparable connection between supervision and reflective practice: supervision is an intelligent conversation with a colleague about a case or issue, and reflective practice is an intelligent conversation with oneself [28]. Supervision should nurture reflective practice, while reflective practice in turn leads to a thirst for supervision in all its forms. The same can be said in connection with team discussion, which may or may not be present in the workplace as a regular activity, and can be regarded as a form of collective supervision. Not surprisingly, the same transferable skills operate in relation to consultations with patients, good supervision with colleagues, and effective team discussion. These include careful listening, the formation of credible hypotheses, and the construction of helpful and challenging questions. All these skills enhance critical thinking and, as such, generate a vision in the workplace that is 'open and malleable, not closed and fixed' [9]. The relationship between reflective practice, supervision, team discussion, and the learning workplace is symbolised in Figure 13.2.

One useful set of concepts to help understand the different purposes of supervision is the one provided by Proctor [29]. She regards supervision as having three aspects: 'normative', 'formative', and 'restorative'. The normative aspect is what links supervision to the 'world out there', where there are standards to meet and rules to be followed. The formative aspect is what helps supervisees to develop. The restorative aspect is what sustains colleagues in their jobs. Each of these aspects may come to the fore or remain in the background, according to the context and the circumstances.

Although supervision is generally understood as a very different concept from clinical governance, there is an argument for regarding supervision as an important means of putting governance into action. While guidelines, protocols, audits, and other tools may provide an external framework for acceptable practice, supervision can be thought of as an activity through which professionals refine and develop their own personal practice. In other words, it is still a form of regulatory activity, but involving continual mutual regulation.

Figure 13.2 The learning organisation.

One aspect of supervision that is vital but often understated is that of imagination. As with research supervision in an academic context, supervision in a clinical setting can at times (and in some hands) be a very dull affair, but it also has the potential to inspire people to perform at their best, and to go to places clinically and professionally that they may never have been able to reach otherwise. Box 13.1 covers the issue of evidence and evaluation in relation to supervision, mentoring, and coaching.

Cases, Contexts, Careers: The Three Domains of Supervision

The majority of supervision in medicine addresses clinical cases. At the same time, it is worth recognising that every case occurs within a wider set of contexts – both for the patient and the professional – and the clinicians looking after the patient are each at a particular stage in their own career. In a teaching context [32] we have found it useful to draw attention to the 'three domains' of supervision: cases, contexts, and careers.

Cases

Cases may be approached from the point of view of straightforward technical case management, for example, what is good practice or best practice, but supervision often raise issues broader than this, including the following:

- clinical uncertainty
- complex co-morbidity
- ethical issues including decision-making at the end of life
- the risk of medicalisation, over-investigation, or over-treatment.
- complaints and medical errors.

BOX 13.1 WHERE'S THE EVIDENCE: Is supervision effective?

In a systematic review of the effect of clinical supervision on patient and residency education outcomes in the US, Farnan et al. [30] found that enhanced attendance of clinical supervisors provides benefits for trainee education in both ambulatory and inpatient settings, and increased supervision during procedure-based scenarios generally results both in improved patient-related and educational outcomes.

From a literature review and a national questionnaire survey of supervision for doctors in specialty training in the UK, Kilminster et al. [31] drew up a framework for effective practice, synthesising the available evidence. This can be summarised as follows:

- **i** Supervisors must be aware of training bodies' and institutions' requirements.
- **ii** Direct supervision and working together positively affects patient outcome and trainee development.
- **iii** Constructive feedback is essential and should be frequent.
- **iv** Supervision should be structured and timetabled.
- **v** Supervision should include clinical management, teaching and research, management and administration, pastoral care, interpersonal skills, personal development, and reflection.
- **vi** The quality of the relationship strongly affects effectiveness.
- **vii** Training for supervision needs to include understanding teaching, assessment, counselling skills, appraisal, feedback, careers advice, and interpersonal skills.

In addition, *helpful supervisory behaviours* include giving direct clinical guidance, linking theory and practice, engaging in joint problem solving, offering feedback and reassurance, and providing a role model. *Ineffective behaviours* include rigidity, low empathy, failure to offer support or follow supervisees' concerns, being indirect and intolerant, and emphasising evaluation and negative feedback.

In cases like these, discussion of the technical or pragmatic options that may be available is inseparable from what has been called 'emotional work', that is, the processing of individual reactions to the case, including frustration or anxiety. This may involve a simple expression of negative feelings, or it may require something more complex and skilled, depending on the capacities of both supervisor and supervisee.

Contexts

Careful case management often depends on thinking about the work setting as much as the clinical issues involved. It is unusual to have an intelligent discussion about any case without at some point having to consider how the multiprofessional network is functioning, and whether it is supporting or hindering practitioners in their work. Similarly, formal or informal case discussions among experienced doctors will regularly address issues such as problems concerning communication, money, politics, or power relationships. Much supervision addresses difficulties in relation to roles and boundaries. Examples are the extent and limits of what patients can legitimately expect from clinicians and what teachers, trainees, colleagues, and team members can legitimately expect from each other. Supervision may also need to address other relevant contexts, including how best to conduct interactions with the patient's family, as well as cultural or faith issues that may be appropriate to the patient's care.

Careers

While formal discussions aimed at clarifying someone's career goals and choices do sometimes take place, it is probably more common for these to be addressed opportunistically at other times. For example, a clinical case might bring learning needs to light and raise issues about whether the supervisee should undertake further training to improve their level of skill. The case of appraisal is interesting in this respect. Although it is not often considered a form of either educational supervision or mentoring, appraisal does offer similar opportunities by creating a space where colleagues can reflect on their competencies, learning needs, and future aspirations [33]. However, there are clearly conflicts when appraisal is linked with professional revalidation or re-accreditation, and some have argued that a pastoral or mentoring approach cannot be combined with anything that has regulatory implications [34]. Figure 13.3 highlights the interrelationship of these three supervisory 'domains'.

Types of Supervision

This section describes some of the commonest types of supervision, with examples of each type. The typology used here is adapted from Clark et al. [27].

Informal Supervision

Informal supervision takes the form of opportunistic exchanges that are generally short and arise spontaneously in the context of everyday work. Typical examples are chats over coffee, in the corridor, or in the operating theatre

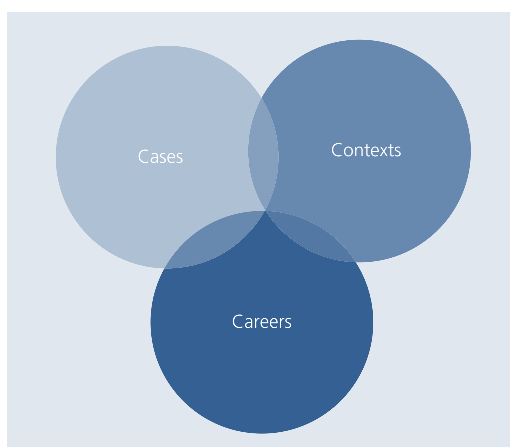


Figure 13.3 The three domains of supervision.

changing room. While formal supervision is usually given by someone more experienced, informal supervision can come from all sorts of people, including juniors and clerical staff. In many circumstances, such supervision 'on the hoof' provides the mainstay of much postgraduate learning and support, whether or not this is explicitly noticed or acknowledged. Colleagues who work well together and respect each other may be able to provide excellent and challenging supervision, even in brief moments snatched from the hurly-burly of everyday work. However, the commonest risk in informal supervision is that it takes the form of swapping anecdotes or reinforces banal or stereotypical practice. This is especially the case if it takes place in the absence of any regular or reflective supervision, or as a substitute for it.

Sometimes the most imaginative supervision that doctors ever receive comes spontaneously and unexpectedly from people outside the medical profession – for example, from lay people commenting on their own medical experiences in social encounters [26]. Some have argued that other personal experiences, including reading novels and poetry, seeing films and operas, or walks in the countryside, are among opportunities for 'self-supervision', since they may provide the conditions for processing the thoughts and emotions that arise from everyday work (see Box 13.2).

Clinical Supervision

Clinical supervision is the commonest form of supervision and is the mainstay of training in most medical settings. It consists of the oversight and day-to-day discussion of clinical cases and their management, and any issues arising from this. It may take a variety of forms, ranging from conversations on ward rounds or in clinics, to more extended and reflective case-based discussions. The degree of presence of supervisors and their availability for dialogue and teaching are crucial, and there is evidence to show that lack of supervision has a direct negative impact on patient care and can be associated with increased mortality and morbidity. Kilminster describes the duties of a clinical supervisor including [35]:

- Offering a level of supervision necessary to the competences and experience of the trainee and tailored for the individual trainee.
- Ensuring that no trainee is required to assume responsibility for or perform clinical, operative, or other techniques in which they have insufficient experience and expertise.
- Ensuring that trainees only perform tasks without direct supervision when the supervisor is satisfied that they are competent to do so; both trainee and supervisor should at all times be aware of their direct responsibilities for the safety of patients in their care.
- Considering whether it is appropriate (particularly out of hours) to delegate the role of clinical supervisor to another senior member of the health care team. In these circumstances, the individual must be clearly identified to both parties and understand the role of the clinical supervisor. The named clinical supervisor remains responsible and accountable overall for the care of the patient and the trainee.

In a training context, clinical supervision often overlaps with didactic training and assessment, particularly where case discussions reveal gaps in knowledge or skills. Even here, however, a conversational approach based on questioning is often more effective than giving advice, since this will help to establish the supervisee's existing ideas and promote independent thinking. It will also ensure that any advice that is needed is pitched at the right level. Clinical supervisors need to be familiar with the principles of effective feedback. Models for offering this include 'Pendleton's rules' [36], the 'One Minute Preceptor' [37], SNAPPs [38], and 'SHARP' [39], although it has also been suggested that supervisors should be wary of formulaic models and be more flexible and responsive in their conversational styles [40]. One of the clinical supervisor's most crucial roles is to calibrate the supervision required for a learner and to enable a gradual increase in responsibility [41, 42].

Doctors generally avoid the term 'clinical supervision' for case discussions taking place with other professions, or beyond the training years. Nevertheless, much clinical supervision does take place with team members who have other professional backgrounds, or among established practitioners. Typical forms of this include team meetings, case reviews, and phone calls or corridor conversations with colleagues to seek expert advice or an independent view of a problem. Many established doctors, particularly those working in isolation, express a wish for more regular or systematic opportunities to discuss cases than their work patterns may allow (see Box 13.3) [43, 44].

BOX 13.2 Informal supervision: An example

Dr K. was working late one evening in her practice when a new patient arrived without an advance appointment. The patient had a complicated and long-standing problem of back pain for which she had seen him many times previously. He demanded a full clinical assessment and immediate referral for another scan. Dr K. felt pressurised. She responded by insisting that he must return for a booked appointment. Later, before locking up her office, Dr K. talked to her receptionist and expressed her frustration at patients like this, and how she hated having to behave 'like a police officer'. The receptionist remarked that Dr K. would probably have handled the patient more carefully if he had come in the morning, when she was less tired. The comment led Dr K. to realise that she had probably been too harsh with the man, who may have been pressurising her because his pain was so bad. She also felt she should have carried out at least a basic examination to rule out any serious new pathology.

Educational Supervision

Educational supervision has become established in the UK and elsewhere as regular supervision taking place in the context of a recognised training, in order to establish learning needs and review progress. Arguably, this is the most complex and challenging form of supervision, since the

BOX 13.3 Clinical supervision: An example

Dr F. has recently completed her core training in general medicine and has just entered her first month of higher specialist training in cardiology. While on call, she is asked to assess a 70-year-old man in the emergency department who has a pacemaker but has been complaining of blackouts. His ECG monitor shows short runs of ventricular tachycardia in addition to his paced rhythm. She phones the on-call consultant who advises her to order a pacemaker check, echocardiogram, and blood tests, commence treatment with beta-blockers, and admit the patient to the ward for close monitoring. The next morning, before they go to see the patient, the consultant asks Dr F. to go through the test findings and explain what further imaging, medical treatment. and device upgrade she thinks the patient might need. He agrees with some of her suggestions, points her towards evidence that challenges some of her other proposals, and rehearses with her what they are going to explain and recommend to the patient.

educational supervisor has to fulfil many overlapping and (in some situations) conflicting roles. As well as facilitating learning, the supervisor has a responsibility to assess the supervisee's performance while the supervision is taking place, or by formal means at a later date. The concept and practice of educational supervision has come to the fore with the introduction of national curricula for all doctors in training grades. Educational supervisors play a key role in the delivery of such curricula. Kilminster lists their tasks, including [35]:

- providing regular review opportunities that should take place at the beginning, middle, and end of a placement
- developing a learning agreement and educational objectives with the trainee that is mutually agreed and is the point of reference for future review
- being responsible for ensuring that trainees whom they supervise maintain and develop their speciality learning portfolio and participate in the speciality assessment process
- providing regular feedback to the trainee on their progress
- ensuring that the structured report, which is a detailed review and synopsis of the trainee's learning portfolio, is returned within the necessary timescales
- contact the employer or educational regulator if the level of performance of a trainee gives rise for concern
- being able to advise the trainee about access to career management
- being responsible for their educational role to the programme director and locally to the employer's lead for postgraduate medical education.

Although educational supervision is different in many ways from mentoring and coaching, supervisors may still need to offer pastoral care at times, particularly for students and trainees who are going through crises in their personal lives or careers.

An educational supervisor may or may not take on the task of day-to-day clinical supervision. In general practice, one and the same person may be doing both jobs. On a programme of hospital training, by contrast, the organisation may assign a single skilled and trained educator to a trainee as an educational supervisor for the duration of the programme, while the day-to-day clinical supervision is carried out by clinicians working with the trainee at any given time. This is especially the case when trainees rotate through a number of different specialties and where an educational supervisor would not have the skills to offer case-based discussion across the whole range of work. When this kind of arrangement is in place, it is essential to have close coordination between all the people involved, to prevent people working at cross-purposes or giving conflicting messages to the trainee. It also makes sense to link the approval of training placements to the quality of the supervision arrangements on offer. This is particularly important because research has shown that the quality of the supervisory relationship is the key to effective supervision in medical settings [45].

When receiving educational supervision, trainees may feel their performance is being judged covertly under the guise of a supportive discussion or that their qualifications and careers may depend on what they say in apparently innocent discussions. Equally, it is hard for educators to stop themselves from being influenced by how trainees present themselves in the course of day-to-day exchanges about casework. Because of this, it is important to remain aware of the two different contexts and be willing to address this difference transparently. When real concerns arise about a trainee's competence, it is fairer and more effective to say so, to set explicit targets, with a clear timetable and (if necessary) an explanation of possible sanctions (see Box 13.4).

BOX 13.4 Educational supervision: An example

Dr P. is the educational supervisor for Alan, a doctor at the beginning of his specialty training in paediatrics. Over the past few weeks, there have been several children on the ward who are possible victims of physical abuse by their parents. Two of Alan's clinical supervisors have been concerned about his simplistic and judgemental attitude towards the problem: he seems to believe that the only desirable solution in every case is adoption outside the family. Dr P. discusses these reports transparently with Alan and arranges for him to spend some time with the child and family mental health team, seeing them do therapeutic work with parents who have abused their children but have also (in many cases) themselves been victims of abuse in the past. He also fixes up for Alan to attend a case conference, in which some of the complexities of a particular case are examined. Dr P. uses their regular education supervision sessions to review what Alan has learnt and to discuss how he might record this in a reflective piece of writing for his learning portfolio.

Remedial Supervision

Remedial supervision occurs when a regulatory agency has formally determined that there are concerns about someone's performance. This can only happen if the agency has the authority to assess performance and prescribe the supervision as a proposed remedy.

Remedial supervision takes place within a framework of assessments and reports. Essentially, remedial supervision is a type of educational supervision where the context is one of prescribed additional training, rather than basic training undertaken on a voluntary basis (see Box 13.5). Although the term 'remedial supervision' is not widely used, partly because of the stigma attached, it is useful for those practising it to be aware of the difference from other forms of supervision when supervision is being carried out in the context of remediation.

Professional Supervision

Professional supervision is a term that is currently used mainly outside medicine but is starting to be used by some doctors as well. It consists of regular, extended one-to-one meetings between established practitioners, mainly to discuss specific cases. In professions such as counselling professional supervision is generally a requirement of continuing professional practice. In the context of medicine (with the exception of psychiatry and a small number of GPs) such supervision is uncommon. Professional supervision is usually, although not always, delivered by experienced members of the same profession. However, it is reasonably common for other professionals, such as psychologists, to offer supervision to doctors. Depending on the cases being discussed, and their context, the supervisor's lack of specific technical knowledge may or may not matter (see Box 13.6).

BOX 13.5 Remedial supervision: An example

Dr L. is a GP who has had a considerable number of complaints from patients. His local health board has referred him to an agency that specialises in assessing and helping doctors whose performance is causing concern. The agency has carried out an occupational psychology assessment, some tests of Dr L.'s knowledge and skills, and a review of selected case notes. On the basis of this, they have found that Dr L. has particular difficulty managing 'grey area' cases that involve vague somatic symptoms. He tends to dismiss such patients brusquely, sometimes without adequate review or investigation. In some cases this has led to delayed diagnosis of significant illness. The agency has now assigned Dr L. a remedial supervisor who has sat in on some of his surgeries and reviewed some video records of his consultations. The supervisor has now held a series of meetings with Dr L. to go through some cases systematically and challenge him on his approach and attitude, in order to help him attain an acceptable level of practice.

Managerial Supervision

Managerial supervision is supervision carried out by someone with direct management responsibility for the supervisee. This may or may not take place within an educational framework (e.g. nurse training) and may or may not involve explicit assessment. In some of the non-medical literature, anything involving clinical line management or direct accountability in employment terms is explicitly disqualified as a form of supervision. By contrast, it is often the norm within the nursing profession (see Box 13.7).

BOX 13.6 Professional supervision: An example

Dr J., a psychiatrist, goes to an experienced psychologist colleague once a month to discuss cases that are causing him concern. His last supervision session was taken up with an account of a patient he is seeing regularly, who has made several suicide attempts. This session was spent examining Dr J.'s feelings of helplessness in relation to his patient's behaviour, and also his sense of personal responsibility for her. His supervisor helped him to understand how he was picking up the patient's own feelings of helplessness, and touched on memories of a family member who had committed suicide some years previously in spite of psychological help. They also looked at how Dr J. could help the patient understand that others did feel responsible for her and care for her, in spite of her fixed belief that nobody ever did.

BOX 13.7 Managerial supervision: An example

Ms B., a community nurse, was seeing a woman who caused her concern. The woman came from Thailand, had been married only a year, and had just had twins. She seemed cowed by her European husband, who clearly expected her to keep the house immaculate in spite of having two new babies. Ms B. had never come across a family situation like this and she felt insufficiently trained to know what to do. She took the case to her regular meeting with her nurse manager. The manager spent time with her exploring all the different issues in the case. These included the risk of domestic abuse, the cultural and age differences between the parents, homesickness and isolation, and the difference in perception between the nurse and the husband. Through the meeting with her manager, Ms B. came to realise that she needed to do a fuller risk assessment as well as building up enough trust with the woman to broach these subjects. She also came away from the discussion with a range of options in her mind. These included a joint visit with the doctor and exploring to see if there were any other Thai mothers in the locality who might be willing to make contact with the woman and help her build up a support system.

Mentoring

Mentoring is usually understood as guidance and support offered by a more experienced colleague [46, 47], although there are also descriptions of peer mentoring or co-mentoring [48]. Mentoring may be either informal or formal. Where it is informal, it may have arisen naturally between colleagues and may not even be described as mentoring, except in retrospect. Where offered as part of a formal scheme, sometimes by mentors with specific training for the role, mentoring is often wide-ranging, covering not just clinical work but professional relationships and career plans as well. Although it should not be confused with counselling, life-cycle issues such as family events will quite often come into the picture. It is often an entirely private encounter, with total confidentiality given and expected on both sides. Virtually all the literature concerning mentoring includes the assumption, implicitly or explicitly, that the arrangement is both voluntary and confidential, although agencies funding mentors may want to know in general terms that their clients are finding it useful (see Box 13.8).

Coaching

Coaching is a form of supervision that has been defined as 'unlocking a person's potential to maximise their own performance' [49]. The coaching relationship, like mentoring, is a voluntary and confidential one. A useful analogy here is with sports coaching, where the client already has an advanced degree of proficiency and the coach, who may or may not be a practitioner of the same sport, helps the client to work towards further excellence. However, even the word 'coaching' can be somewhat confusing, since it is also used colloquially when someone needs extra help (e.g. 'She needs specific coaching in how to talk to patients'). The term has come into vogue within the context of 'life coaching' [50] to help people achieve personal and career fulfilment, and in a more general sense to describe any one-to-one learning relationship that does not involve management or assessment. Some writers use the terms 'mentoring' and 'coaching' interchangeably, or have abandoned one term in favour of the other (see Box 13.9).

BOX 13.8 Mentoring: An example

Since attaining her first hospital consultant post three years ago, Dr M. has had meetings every few months with an experienced doctor who is trained as a mentor. She has used these meetings to discuss the difficulties she has had in adjusting to life not just as a newly qualified radiologist, but also as a team leader, a manager, and (most recently) a new mother. Over the past two or three mentoring sessions, the same theme has come up repeatedly: the feeling that her male colleagues are paternalistic and sexist. At her most recent session her mentor simply asked her, 'Are you really just asking my permission to get out?' Dr M. felt enormous relief at being asked the question. She realised that she did indeed want to move on, but had feared her mentor would be critical of her for leaving her first senior post so soon.

BOX 13.9 Coaching: An example

Dr T., a hepatic surgeon, was recently promoted from medical director of his hospital to chief executive. As a result, he has decided to give up most of his clinical work, except for a small private practice, and to concentrate on his management role. He is paying to see a coach from a private consultancy organisation to help him develop his new professional identity. Quite a few of his fortnightly coaching sessions have been spent examining how he can keep his trusting relationships with his medical colleagues at the hospital (some of whom have been his friends for many years), while also behaving equitably and responsibly in relation to all the other staff groups at the hospital, including nurses, technicians, non-clinical staff, and manual workers.

Summary

The types of supervision listed here are not exhaustive or mutually exclusive. Among the huge variety of learning encounters that occur in medicine, there will be many that only loosely fit any of the categories described here, while other encounters may include aspects of several types of supervision, or shift between supervision and training from moment to moment. Clark et al. have suggested that it is less important to achieve consensus on terms than to ask the following questions in relation to any supervision activity [27].

- Who is asking for this to be done?
- What do they want, and why?
- Do they know what the supervisor does and does not offer?
- Will the person be attending voluntarily?
- Is anyone expecting specific outcomes, should they be and, if so, what?
- Who is paying whom, and do all the parties know this?
- Who is reporting to whom, about what, exactly when, and do all the parties know this?
- Is everyone agreed on the terms being used, and on their meaning?

Similarly, Proctor has proposed that 'What the role relationship is called is probably unimportant in practice … The roles, responsibilities and rights need to be identified from the beginning, and to be discussed, made real and reviewed' [51].

Conceptual Frameworks for Supervision

There is no single conceptual framework that is universally recognised within the world of supervision and mentorship. Different writers use – or assume – frameworks that draw on various fields, including learning theory, with a particular emphasis on authorities such as Schon [10, 11] and Kolb [52]. Some explicitly address learning styles or personality types [53, 54]. One useful model based on learning theory comes from Proctor [29], drawing on Wackman et al. [55]. This suggests that learners move continually through a cycle from 'unconscious incompetence' through

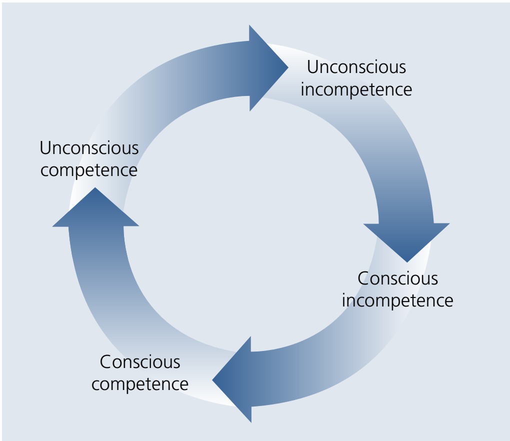


Figure 13.4 The supervision cycle (*Source*: After Proctor 2001).

'conscious incompetence' to 'conscious competence' and finally 'unconscious competence' – before once more uncovering in themselves a new area of ignorance of which they were previously unaware. In other words, supervision invites people to consider how they may be avoiding more adventurous areas of practice, and to explore these – at first tentatively, then as a matter of course – until they are ready to extend themselves even further (see Figure 13.4).

Certain themes arise again and again in the supervision literature, suggesting that there is a consensus about particular ideas that cross theoretical boundaries. These themes include the following:

- Supervision should be about enabling, empowerment, and sustaining human values.
- Supervisors may need to take on a variety of roles at different times, including guide, advisor, role model, sponsor, teacher, and facilitator.
- Supervision needs to pay attention to the personal, professional, and relational aspects of the work.
- Supervisors always have to bear in mind three separate 'clients': the supervisee, the patient, and the organisation or agency.
- Supervision needs to address the complexity and uniqueness of the problems brought, in order to generate solutions or options that are the best fit.
- Supervision is an interactional process. To be effective, it depends on emotional attunement, mutual trust, and usually an evolving relationship between supervisor and supervisee.

In keeping with the modern emphasis on narratives and their importance, a number of writers have suggested that supervision is inherently a narrative-making activity [56]. The role of the supervisor is therefore to elicit an existing narrative of 'the problem' as it is currently understood, and to question the supervisee in such a way that a new understanding emerges, in the form of a different narrative in which 'the problem' is either lessened or has dissolved. One advantage of using this approach in a medical context is that it places less emphasis on eliciting emotion and more on 'bringing forth new stories'. A narrative-based technique allows the technical and factual content of cases to be integrated into the evolving story that the supervisee brings (see Box 13.10).

BOX 13.10 FOCUS ON: Narrative-based supervision

In the context of training doctors to carry out effective supervision, I have developed a simple theoretical framework for a narrative-based approach, summarised as 'the seven Cs' [56].

1. *Conversation*. Wherever possible, supervision should aim to resolve problems through the conversation itself rather than through giving advice.
2. *Curiosity*. The best stance for the supervisor is one of curiosity: establishing what the supervisee already knows and what options have already been considered, or might be explored further.
3. *Contexts*. It is often more important to discover the contexts for a problem rather than focusing on the content. These contexts include the patient's and the supervisee's beliefs, values, and preferences, and the needs and pressures of the organisation.
4. *Complexity*. Many problems brought to supervision are inherently complex, involving many levels of difficulty or intersecting difficulties. Supervision mainly offers opportunities for supervisees to enrich their understanding of what is going on, in order to find a way forward. It rarely helps by finding a 'quick fix'.
5. *Challenge*. Supervision requires frankness and risk taking on both sides.
6. *Caution*. It also requires respect and circumspection: operating within the limits of the supervisee's capacity to tolerate anxiety, while not avoiding the challenge within those limits.
7. *Care*. Most of all, supervision requires attentiveness and positive regard.

Common Tensions in Supervision

As with any human relationship, supervision and related activities are subject to particular tensions. The ones most commonly experienced by practitioners and cited in the literature are:

- facilitation versus training and assessment
- the needs of the supervisee versus the needs of the organisation
- affirmation versus challenge.

This section addresses each of these tensions in turn.

Facilitation Versus Training and Assessment

In a technical specialty such as medicine, supervisors may have an overriding wish to bring out the best in their supervisees, and yet at the same time will be aware of the need for a secure basis of knowledge and skills, and a responsibility to assess these. There are certain facts that practitioners need to know, and competencies they need to possess. There is little point in trying to 'bring forth' such knowledge and skills where they are absent. On the other hand, there is always the risk that supervisors will be tempted to remain in a didactic, knowledge-imparting mode even when this is unhelpful. Supervisees can put pressure on their supervisors, consciously or unconsciously, to offer a quick fix. Supervisors therefore need to offer clear signposting for areas of professional certainty (e.g. the correct doses of drugs) and areas of considerable uncertainty (e.g. the wider contexts of case management). They also need to signpost when assessment is taking place, and when someone's observed performance may affect a judgement of them indirectly even when not part of a formal assessment.

The Supervisee Versus the Organisation

In many circumstances the supervisee's needs and those of the organisation are identical. For example, a trainee's clinical duties may provide excellent opportunities for training and supervision in the course of everyday work. However, this is often not the case, and medical trainees may, for example, find themselves losing out on important learning in order to maintain a clinical service. Supervisors who are employed by the organisation therefore have to manage a fine line between loyalty to the trainees and loyalty to the organisation. Openness and realism are needed to negotiate this kind of tension.

Affirmation and Challenge

A core tension in any form of supervision is the extent to which the supervisor sets out to support or challenge the supervisee. Every supervisor or mentor wants people to feel good about themselves. Equally, every supervisor wants (or should want) the supervisee to be capable of change and to develop as far as their potential will allow. Much of the skill of supervision depends on a capacity to go just beyond the 'comfort zone' without being perceived as a bully. This involves offering enough challenge to the supervisee to help in the exploration of new ideas and new possibilities, but not so much that the supervisee becomes excessively anxious, defensive, or deskilled. It may also involve fine judgement as to whether the context permits some careful exploration of personal issues that may be impeding professional performance. Like the skill of consulting with patients who may be anxious or fearful, this is not something that can be learned from books or articles. It depends on a gradual, sometimes lifelong, acquisition of the capacity to calibrate one's speech with the exact circumstances and with the wishes and needs of the supervisee.

Raising the Profile of Supervision: Changing the Culture of Medicine

Medicine has come a long way in promoting a culture of systematic supervision, but still lags behind some other professions in many respects. Regulatory changes, including accreditation of supervisors [57], have a part to play in making sure that clinical and educational supervisors have basic competencies for the job, but they cannot necessarily ensure its quality. This requires focused programmes of faculty development, including training in supervision from the early years of a medical career [58]. Pront et al. [59] have identified four roles that supervisors need to develop in order to provide learning-focused supervision:

- To partner: establish a learning relationship through communication, trust, and respect, identifying boundaries for learning and practice.
- To nurture: as a learning advocate, transition the student into the clinical setting and facilitate socialisation into the professional culture and health team.
- To enable: promote and support learning opportunities for student engagement within individual established boundaries.
- To facilitate meaning: promote understanding through problem solving, reflection, and feedback, fostering a professional way of knowing and being.

Conclusions

Doctors are always pulled in two directions. On the one hand, there is a clear need for concrete facts, explicit guidelines, and consistent policies – in other words, for certainty. On the other hand, there is also the need for practitioners to gain in confidence and in wisdom, and to be able to grapple with the complex, multidimensional problems that they encounter with an ever-increasing level of understanding, sophistication, and adventurousness. Supervision, mentoring, and coaching have the capacity to ensure that medical practice rises consistently above the uninspired, the routine, and the automatic. Medical work is not just about scientific facts. It requires practitioners to apply their knowledge in the context of individual lives and complex human systems. It crosses many domains of knowledge, including the social and interpersonal. To practise reflectively, doctors need emotional and intellectual trust, so that they can reflect frankly on their own work and learn continually without feeling overexposed or under excessive scrutiny. Activities such as supervision, mentoring, and coaching provide opportunities for doctors to examine their own work safely and effectively, based on their everyday professional experience, and in a way that complements other forms of education and training.

Acknowledgements

Many colleagues made helpful comments on earlier editions of this chapter. I particularly want to thank Dr Jonathan Burton, Dr Julie Draper, Dr Anat Gaver, Dr Helen Halpern, Dr Gareth Holsgrove, Dr Tim Swanwick, Prof. Frank Smith, and Dr Per Stensland.

References

1 NHS Modernising Medical Careers (2005). *Curriculum for the Foundation Years in Postgraduate Education and Training*. London: Department of Health.
2 COPMeD (2016). The Gold Guide: A Reference Guide for Postgraduate Speciality Training in the UK (6). Conference of Postgraduate Medical Deans, UK www.copmed.org.uk/publications/the-gold-guide (accessed 30 March 2017).
3 Gray, R. (2001). A framework for looking at the trainer/registrar tutorial in general practice. *Education for General Practice* 12: 153–162.

4 NHS England (2016). Doctor's medical appraisal checklist. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.578.606&rep=rep1&type=pdf (accessed 28 June 2018).

5 Garrett-Harris, R. and Garvey, B. (2005). *Towards a Framework for Mentoring in the NHS*. Sheffield: Mentoring and Coaching Research Unit, Faculty of Organisation and Management, Sheffield Hallam University.

6 Beecham, B., Dammers, J., and van Zwanenberg, T. (2004). Leadership coaching for general practitioners. *Education for Primary Care* 15: 579–583.

7 Bowden, R. and Schofield, J. (2004). Work based learning and poor performance. In: *Work Based Learning in Primary Care* (ed. J. Burton and N. Jackson), 103–118. Oxford: Radcliffe Publishing.

8 Department of Health (2004). *Modernising Medical Careers: The Next Steps: The Future Shape of Foundation Specialist and General Practice Training Programmes*. London: Department of Health.

9 Brookfield, S. (1987). *Developing Critical Thinkers: Challenging Adults to Explore Alternative Ways of Thinking and Acting*. Milton Keynes: Open University Press.

10 Schon, D. (1983). *The Reflective Practitioner: How Professionals Think in Action*. London: Temple Smith.

11 Schon, D. (1987). *Educating the Reflective Practitioner: Towards a New Design for Teaching and Learning in the Professions*. New York: Basic Books.

12 Stewart, M., Brown, J., Weston, W. et al. (2013). *Patient-Centred Medicine: Transforming the Clinical Method*, 3e. Oxford: Radcliffe Publishing.

13 Bishop, V. ed. (1988). *Clinical Supervision in Practice, Some Questions, Answers and Guidelines*. London: Macmillan.

14 Bond, M. and Holland, S. (1998). *Skills of Clinical Supervision for Nurses*. Buckingham: Open University Press.

15 Cutcliffe, J., Butterworth, T., and Proctor, B. (2001). *Fundamental Themes in Clinical Supervision*. London: Routledge.

16 RCN Institute (2000). *Realising Clinical Effectiveness and Clinical Governance through Clinical Supervision*. Oxford: Radcliffe Medical Press.

17 Boud, D. and Solomon, N. ed. (2001). *Work-Based Learning: A New Higher Education*. Buckingham: Open University Press.

18 Department of Health (2001). *Working Together, Learning Together: A Framework for Lifelong Learning*. London: Department of Health.

19 Burton, J. and Jackson, N. ed. (2004). *Work Based Learning in Primary Care*. Oxford: Radcliffe Publishing.

20 Steinberg, D. (2005). *Complexity in Healthcare and the Language of Consultation*. Oxford: Radcliffe Publishing.

21 Lupton, D. (1994). *Medicine as Culture: Illness, Disease and the Body in Western Societies*. London: Sage.

22 Thomas, P. (2006). *Integrating Primary Care: Leading, Managing, Facilitating*. Oxford: Radcliffe Publishing.

23 Launer, J. (2002). *Narrative-Based Primary Care: A Practical Guide*. Oxford: Radcliffe Medical Press.

24 Scaife, J. (2001). *Supervision in Mental Health Professions: A Practitioner's Guide*. Hove: Brunner Routledge.

25 Butterworth, T. (1992). Clinical supervision as an emerging idea in nursing. In: *Clinical Supervision and Mentorship in Nursing* (ed. T. Butterworth and J. Faugier), 3–17. Cheltenham: Stanley Thornes.

26 Burton, J. and Launer, J. ed. (2003). *Supervision and Support in Primary Care*. Oxford: Radcliffe Medical Press.

27 Clark, P., Jamieson, A., Launer, J. et al. (2006). Intending to be a supervisor, mentor or coach: which, what for and why? *Education for Primary Care* 17: 109–116.

28 Launer, J. (2003). Practice, supervision, consultancy and appraisal: a continuum of learning. *British Journal of General Practice* 53: 662–665.

29 Proctor, B. (2001). Training for the supervision attitude, skills and intention. In: *Fundamental Themes in Clinical Supervision* (ed. J. Cutcliffe, T. Butterworth and B. Proctor), 25–46. London: Routledge.

30 Farnan, J.M., Petty, L.A., Georgitis, E. et al. (2012). A systematic review: the effect of clinical supervision on patient and residency outcomes. *Academic Medicine* 87: 428–442.

31 Kilminster, S., Cottrell, D., Grant, J., and Jolly, B. (2007). AMEE Guide No. 27: effective educational and clinical supervision. *Medical Teacher* 29: 2–19.

32 Health Education England, Multiprofessional Faculty Development (2012). Supervision skills for clinical teachers: A world class teaching initiative. http://www.faculty.londondeanery.ac.uk/supervision-skills-for-clinical-teachers (accessed 30 March 2017).

33 Pringle, M. (2003). Re-evaluating revalidation and appraisal. *British Journal of General Practice* 53: 437–438.

34 Bulstrode, C. and Hunt, V. (2000). What is mentoring? *Lancet* 356: 178–178.

35 Kilminster, S. (2010). Understanding the nature and purpose of clinical and educational supervision. *InnovAIT* 3: 615–618.

36 Pendleton, D., Schofield, T., Tate, P., and Havelock, P. (1983). *The Consultation: An Approach to Learning and Teaching*. Oxford: Oxford University Press.

37 Neher, J.O., Gordon, K.C., Meyer, B., and Stevens, N. (1992). A five-step "microskills" model of clinical teaching. *Journal of the American Board of Family Practice* 5: 419–424.

38 Wolpaw, T.M., Wolpaw, D.R., and Papp, K. (2003). A learner centred model for outpatient education. *Academic Medicine* 78: 893–898.

39 Imperial College, London (2010). *The London Handbook for Debriefing: Enhancing Debriefing in Clinical and Simulated Settings*. London: Imperial College.

40 Launer, J. (2016). Giving feedback to medical student and trainees: rules and realities. *Postgraduate Medical Journal* 92: 627–628.

41 Peel, J.L. and Nolan, R.J. (2015). You can't start a central line? Supervising residents at different stages of the learning cycle. *Journal of Graduate Medical Education* 7: 536–538.

42 Hauer, K.E., Oza, S.K., Kogan, J.R. et al. (2015). How clinical supervisors develop trust in their trainees: a qualitative study. *Medical Education* 49: 783–795.

43 Owen, D. and Shohet, R. ed. (2012). *Clinical Supervision in the Medical Profession: Structured Reflective Practice*. Milton Keynes: Open University Press.

44 Sommers, L. and Launer, J. ed. (2013). *Clinical Uncertainty in Primary Care: The Challenge of Collaborative Engagement*. New York: Springer.

45 Cottrell, D., Kilminster, S., Jolly, B., and Grant, J. (2002). What is effective supervision and how does it happen? A critical incident study. *Medical Education* 36: 1042–1049.

46 Department of Health (2004). *Mentoring for Doctors: Signposts to Current Practice for Career Grade Doctors*. London: Department of Health, Doctors' Forum.

47 Viney, R. and Harris, D. (2011). Mentoring supervision for doctors and dentists in a National Health Service (NHS) deanery. In: *Coaching and Mentoring Supervision Theory and Practice* (ed. T. Bachkirova, P. Jackson and D. Clutterbuck), 251–257. Milton Keynes: Open University Press.

48 Sackin, P., Barnett, M., Eastaugh, A., and Paxton, P. (1997). Peer supported learning. *British Journal of General Practice* 47: 67–68.

49 Whitmore, J. (1996). *Coaching for Performance*. London: Nicholas Brierley.

50 Kersley, S. (2006). *Prescription for Change: For Doctors Who Want a Life*, 2e. Oxford: Radcliffe Publishing.

51 Proctor, B. (2000). Postscript. In: *Mentorship, Preceptorship and Clinical Supervision: A Guide to Professional Support Roles in Clinical Practice*, 2e (ed. A. Morton-Cooper and A. Palmer), 221. Oxford: Blackwell.

52 Kolb, D. (1984). *Experiential Learning*. Englewood Cliffs: Prentice Hall.

53 Honey, P. and Mumford, A. (1992). *Manual of Learning Styles*. Maidenhead: Honey.

54 Houghton, A. and Allen, J. (2005). Understanding personality type: doctor–patient communication. *BMJ Career Focus* 330: 36–37.

55 Wackman, D.B., Miller, S., and Nunnally, E.W. (1976). *Student Workbook: Increasing Awareness and Communication Skills*. Minneapolis: Interpersonal Communication Programmes.
56 Launer, J. (2013). Narrative-based supervision; training in narrative-based supervision. In: *Clinical Uncertainty in Primary Care: The Challenge of Collaborative Engagement* (ed. L. Sommers and J. Launer), 147–176. New York: Springer.
57 General Medical Council (2016). Recognition and approval of trainers. https://www.gmc-uk.org/education/how-we-quality-assure/medical-schools/recognition-and-approval-of-trainers (accessed 28 June 2018).
58 Royal College of General Practitioners. Curriculum Statement 3.7: Teaching, Mentoring and Clinical Supervision. https://www.gponline.com/rcgp-curriculum/teaching-mentoring-clinical-supervision (accessed 28 June 2017)
59 Pront, L., Gillham, D., and Schuwirth, L.W.T. (2016). Competencies to enable learner-focused supervision: a thematic analysis of the literature. *Medical Education* 50: 486–495.

14 Interprofessional Education

Della Freeth[1], *Maggi Savin-Baden*[2], *and Jill Thistlethwaite*[3]

[1]Royal College of Physicians, London, UK
[2]Institute of Education, University of Worcester, Worcester, UK
[3]School of Communication, University of Technology, Sydney, Australia

KEY MESSAGES

- Interprofessional education (IPE) is just a special case of professional education, so knowledge about good practice for learning and teaching in a wide range of contexts can be applied directly to IPE.
- The uniqueness of IPE lies in deliberately creating heterogeneous groups. By bringing together participants from different professions around a task, it is anticipated that the increased diversity of knowledge and perspectives will enhance the learning of all.
- The *raison d'être* for IPE is to enhance professional practice: uniprofessional as well as interprofessional.
- Poorly planned or delivered IPE may be damaging if it generates a reluctance to engage in subsequent interprofessional collaboration or reinforces negative stereotypes.
- IPE is not limited to formally planned and overtly labelled education. Whenever practitioners meet in multiprofessional groups to address complex needs or to improve clinical services, there is potential for interprofessional learning. This serendipitous IPE is a feature of daily practice.

Introduction

Interprofessional education (IPE) is simply a special case of professional education. Thus, everything we know about good practice in professional education, a broad sweep of which is set out in the other chapters of this book, also applies to IPE. IPE differs, however, in the conscious decision to amplify the heterogeneity of learners by including students or members of different professions, bringing their differing professional perspectives. We take a broad view of education and include formal and informal interprofessional learning (IPL), serendipitous learning (see Box 14.1) [1] and the effects of the hidden curriculum [2]. We view IPE and IPL as focused on improving professional practice, care, and services to support and enhance the lives of individuals, communities, and populations. We argue that IPE and IPL contribute to this through promoting high-quality interprofessional collaboration (IPC), founded on complementary professional contributions. We use the term 'profession' broadly, extending it to include occupations contributing to care and services that may lack regulated entry and licensure, for example managers, technicians, and health care assistants. For brevity and consistency, we will use the term 'patient' throughout the chapter, whilst acknowledging that client or service user would be more appropriate in some health contexts.

Bringing people from different groups together (physically or electronically) for IPE is logistically complex in terms of timetabling, space, facilitation, funding, and identifying or creating learning resources that suit all participants. Before investment in IPE, therefore, we need to examine its effectiveness in different contexts and with different approaches: we will summarise some key literature in Box 14.3, but first we will provide some definitions.

Defining Interprofessional Education: The Importance of Pronouns

The terminology relating to IPE is not standardised and can be confusing. Box 14.1 provides simple definitions of phrases used in this chapter. The most widely recognised definition of IPE is that of the UK Centre for the Advancement of Interprofessional Education (CAIPE): 'Interprofessional education occurs when two or more professions learn with, from and about each other to improve collaboration and the quality of care' [3], later extended to include members or students of professions, and both care and services [4]. These definitions position IPE as that which prompts IPL, while its purpose is to improve care and services through collaboration; or as the World Health Organization (WHO) publication succinctly framed it, 'Learning Together to Work Together' [5].

The WHO's *Framework for Action on Interprofessional Education and Collaborative Practice* reordered the pronouns to 'about, from, and with' [6]; both formulations appear in the literature. Bainbridge and Woods' study with students and faculty elicited associations with each pronoun

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

BOX 14.1 FOCUS ON: Terminology

Prefixes (e.g. inter-, multi-, trans-) and adjectives (e.g. shared, common, professional, disciplinary), combined with organisational and educational nouns (e.g. agency, sector, provider and learning, education, training), create a confusing array of competing and inconsistently used phrases to describe IPE. Simple descriptions of commonly used terms are sufficient for this chapter. For a more elaborated exploration of terminology see the 'Instructions to Authors' in the *Journal of Interprofessional Care* [7].

Interprofessional education (IPE)
Learning with, from, and about each other to improve collaboration and the quality of care and services [4]. In the North American literature, in particular, this is often termed: Interprofessional Education for Collaborative Practice (IPECP).

Interprofessional learning (IPL)
Learning arising from interaction between members (or students) of two or more professions. This may be a product of formal IPE or learning may happen serendipitously in the workplace or education settings [1]. (See the later section on 'The Diversity of Interprofessional Education' for more about serendipitous IPL.)

Multiprofessional education (MPE)
Members (or students) of two or more professions learn side by side for whatever reason, for example they may have a common need to master specific knowledge or skills.

Uniprofessional education
Uniprofessional education, with participants from a single profession, forms the bulk of each profession's individual pre-licensure education and some profession-specific post-licensure education; consequently it is also an important venue for the development of knowledge, skills, and attitudes needed to underpin effective (and usually interprofessional) teamwork.

Interprofessional collaboration (IPC)
The process of developing and maintaining effective interprofessional working relationships with learners, practitioners, patients, families, and communities to enable optimal health outcomes [8].

BOX 14.2 'With, from, and about' (Adapted from Bainbridge and Woods) [9]

Learning about: understanding other's roles and responsibilities, knowing people outside their professional role, overcomes stereotypes, may be more superficial than the next two categories.
Learning with: active engagement; interaction, co-location; equity; sharing values; non-judgemental; collaboration; teamwork; trust.
Learning from: trust, respect; open communication, dialogue, confidence in other's knowledge and skills.

(Box 14.2) [9]. For us, the pronouns 'with, from, and about' highlight that IPL necessitates *active* and *interactive* learning. IPE is not about a mixed group of people acquiring the *same* knowledge or developing the *same* clinical skill.

The Rationale for Interprofessional Education for Collaborative Practice (IPECP)

The drivers for IPECP have been well elucidated during the past half-century and include the increasing incidence of complex and chronic conditions globally, the patient safety agenda, recruitment and retention of staff, and global workforce inequities [10–14]. In the USA a resurgence of interest in IPECP has focused on the 'triple aim' of:
- improving the quality of patients' health care experiences and patient satisfaction
- improving the health of communities and populations
- reducing the cost of health care delivery [15].

In low- and middle-income countries, priorities focus more on building workforce capacity for primary health care because of mal-distribution of health professionals and a projected shortfall of 18 million health workers by 2030 [16].

The report, *Health Professionals for a New Century: Transforming Education to Strengthen Health Systems in an Interdependent World* [17], called for wide-ranging and fundamental changes in health professionals' education, and in the relationships between health care providers, educational institutions, and the populations they serve. It suggested (p. 8) there is a 'mismatch of professional competencies to patient and population priorities because of fragmentary, outdated and static curricula producing ill-equipped graduates' [17]. The need for team-based care, and therefore team-based learning, was stressed and IPE advocated as part of a continuum of training.

IPE should help to develop insights, shared knowledge, and teamwork skills that promote effective collaboration to deliver high quality care efficiently. It can examine how health care 'teams' vary and collaborative practices may range from well-rehearsed task-focused teamwork, to the more fluid concept of 'knotworking'. Building upon Engeström's work [18], Bleakley (p. 140) describes knotworking as 'expert work taking place in rapidly shifting contexts, where a number of 'loose ends" of activity are constantly being tied together or untied, to create the conditions for collaborative production of knowledge or new work practices' [19]. More recently, Paradis and colleagues have examined the variety of ways in which IPC has been conceptualised in a leading medical education journal [20].

Traditional hierarchies, roles, and boundaries have the potential to inhibit effectiveness: IPE can contribute to challenging and renegotiating established ways of thinking and being. For example, it may help participants to appreciate the contributions of different members of the team; it can promote shared examination of multifaceted problems and the formation of team plans, rather than a tangle of criss-crossing separate responses; it may promote healing and collaboration when external pressures have damaged collaborative practice.

Can IPE make a Difference? Does it Work?

The rationale for IPE is multifaceted and strongly argued. However, delivering IPE requires effort outside (and mostly in addition to) usual work and processes: leadership, collaborative effort, overcoming constraints and barriers, and nurturing enabling factors. Summaries in reviews by Lawlis and colleagues [21] and Reeves and colleagues [22] underline this.

Questions such as 'Can IPE make a difference?' and 'Does it work?' are often asked. The answer is the same as for any type of education: well-focused, well-designed, well-delivered, contextually appropriate IPE makes a range of positive differences (see Box 14.3). There are hundreds of peer-reviewed studies of IPE, and a growing body of evidence of effectiveness, although few studies relate to developing countries [23]. As with other topics in health professional education, the quality of studies varies. IPE, as

BOX 14.3 WHERE'S THE EVIDENCE: Interprofessional education

Systematic reviews

There have been many systematic and scoping reviews of IPE studies and, increasingly, syntheses of reviews [27, 35–42]. In one example, Reeves and colleagues appraised reviews reporting IPE study outcomes relating to learner gains, collaborative practice, and patient care [43]. They synthesised eight reviews (range 10–133 studies) and noted:

- the variable quality of reviews and the IPE studies they included, and signs of improvement as the IPE literature matures
- the diverse range of activities within IPE, involving various health professions over different time periods in different settings
- most studies were conducted at a single site, and examined only short-term impact
- most studies reported positive findings in relation to the learner-focused outcomes of satisfaction, changes in attitudes, and/or change in knowledge and skills; fewer reported changes in individual behaviour; a small number found positive changes in organisational practice and some found changes in clinical outcomes.

They concluded (p. 66): 'this updated review-of-reviews revealed that IPE can nurture collaborative knowledge, skills, and attitudes. It also found more limited, but growing, evidence that IPE can help enhance collaborative practice and improve patient care' [43].

It is more common for studies to report positive outcomes than mixed, neutral, or negative outcomes [22]. This may be a symptom of publication bias [44].

Post-licensure IPE

For post-licensure learners there is evidence that IPE (mostly embedded in quality improvement initiatives) can improve the quality of care and the quality of working lives. Here are just a few examples:

- improved preventative care, including increased screening and immunisation rates [45–47]
- improved teamwork, fewer errors, and swifter life-saving treatment observed in emergency departments [48, 49]
- more regular briefing and better teamwork in operating theatres [50]
- increased interprofessional participation in planning and reviewing care [51]
- more patient-centred communication [52]
- increased and self-sustaining networking among primary mental health care professionals, particularly in rural areas, which increased knowledge about other practitioners and confidence about referrals [53].

Pre-licensure IPE

Studies of pre-licensure IPE have provided a wide range of insights. The list below is illustrative and the examples are just a small selection from an ever-increasing pool:

- sustainable models of delivery that, over time, can accommodate large numbers of students [54–59]
- variable responses from students from different professions [56, 60, 61]
- IPE can develop more positive perceptions of members of other professions, constructive 'mutual inter-group differentiation', a more sophisticated understanding of roles within teams [62–64], and significantly higher scores on a scale measuring perceived need for professional cooperation [65]
- IPE can increase interest in working in places or specialties where there are recruitment shortfalls [66, 67]
- patients are pleased with care and advice provided by interprofessional student teams [68, 69], and improved patient outcomes have been recorded [70]
- interprofessional student teams can identify gaps or alternative approaches that enable qualified practitioners to improve care and patient outcomes [69].

a complex intervention in complex dynamic systems, presents research challenges. To support better studies and evidence syntheses, a conceptual model has been developed comprising an 'interprofessional learning continuum', 'enabling or interfering factors', and a range of outcome categories [24]. Theoretical contributions to the IPE research literature are increasing [25–27] and there are practical guides for those wishing to evaluate IPE [1, 28, 29].

The breadth of health professionals' education and continuing professional development aims to improve health outcomes and the quality of care. However, it is difficult to demonstrate that any specific facets of the pre-licensure curriculum do this (including IPE), because there are multiple confounding factors and varying amounts of time between learning activities and a professional subsequently being able to provide unsupervised care [30]. Correlation is easier to show than causation. Nevertheless, it is possible to evaluate pre- and post-licensure IPE to explore whether it helps learners achieve the learning outcomes (or competencies) relevant to interprofessional practice that have been defined by professional accreditation bodies, such as CanMEDS [31], and means of doing this are being developed and refined [32, 33]. At post-licensure level the causal chain from IPE to patient and service outcomes can be shorter, although the contexts for IPE and its intended outcomes remain complex and challenging, particularly when health systems are changing rapidly [34]. Nevertheless, in a range of contexts interprofessional continuing professional development or workplace learning has achieved contextually important outcomes for patients, professionals, or services (Box 14.3).

The Diversity of Interprofessional Education

IPE may be described with two dimensions: first, variation in emphasis; and second, variation in the degree of planning and formalisation [1]. The emphasis dimension runs from a primary focus on interprofessional collaboration as the subject matter for the IPE [71, 72] to a secondary focus on interprofessional collaboration and a primary focus elsewhere (e.g. specific patient or client group [73, 74], or professional skills [75] and policy innovation [76]). Many IPE initiatives seek to pay balanced attention to interprofessional collaboration and some other substantive content, for example, pain [77].

The planning and formalisation dimension recognises that serendipitous IPL is influential and should be acknowledged. Serendipitous IPL often happens in daily practice when members of different professions review their work together, or encounter something unusual, causing them to pause and more closely observe some aspect of overlapping concern or seek information from one another. It is axiomatic that we cannot plan serendipitous learning. However, we can pay more attention to creating the right conditions for positive unplanned IPL, such as promoting recognition of complementary expertise and willingness to share ideas [78].

More predictable than serendipitous IPL, *informal* IPL/IPE occurs because of work systems or the structure of educational programmes. Multidisciplinary team reviews of patients or processes can be good examples of informal IPL/IPE. These may be labelled as team meetings [79], action learning, audit, or external inspection, each with a different emphasis and potential for informal and formally recognised IPL. Nisbet and colleagues advocate valuing more highly workplace IPL arising informally from daily work and improvements, and making it more explicit [80]. Furthermore, students from different professions are often in the same clinical areas at the same time, providing opportunities for informal IPE.

Formal pre- and post-licensure IPE comprises planned activities to promote learning *with, from, and about* members of other professions, and is the main focus of the IPE literature. Formal IPE normally supports serendipitous and informal IPL during less structured periods, such as refreshment breaks; it is worth structuring formal IPE to leverage these by-products.

Designing Effective Interprofessional Education

Effective IPE is effective education with the added value of harnessing the knowledge, learning needs, and dynamics of an interprofessional group purposefully. This book, with adjustment for interprofessionality, provides a wealth of advice to underpin the design of IPE. It is also important to consider theories of learning and change that will inform the design of IPE (Box 14.4). Learning experiences evoke emotions such as excitement, satisfaction, empathy, anxiety, boredom, fatigue, and disaffection so IPE developers and facilitators need to plan and manage IPL in ways that create positive emotions and, if necessary, acknowledge and work with negative emotions. Poor-quality IPE may be particularly damaging if it creates increased reluctance to engage in subsequent interprofessional collaboration or reinforces negative stereotypes.

Activities that do not allow each participating group to contribute to more or less the same extent are unlikely to be a good foundation for sound IPE. The aim is for everyone to learn something productive through balanced exchanges, not for one group to plunder the expertise of another.

The Perceived Relevance of the Learning Opportunity

Learners are life-centred and problem-centred and are motivated to develop their knowledge and skills when they encounter an idea, a task, or a problem that matters to them in their current context. It is difficult to engage with things that are not interesting and seem to have little relevance.

During IPL participants are likely to want to develop their knowledge and practice from multiple perspectives (as individuals, as members of a particular profession, and as members of diverse teams and collaborations). Furthermore, their primary focus will vary over time and in response to external demands. IPE normally addresses this personal, professional, and team development by

BOX 14.4 Focus on theory

Theory informs and helps us better understand IPE. Previously, the interprofessional field was considered under-theorised in relation to curriculum design, evaluation, and research [36, 81] although many potentially relevant theories had been identified [82]. There is now an interprofessional community of practice dedicated to theory, scholarship, and collaboration [83]. A single theory is not sufficient due to the complexity of IPE 'where different groups of learners meet for a variety of purposes at different stages of their professional development' [84, p. 81]. It is also important to consider theory from outside health care, for example in the wider literatures about organisations and workplaces and formal and informal learning.

Chapter 4 of this book discusses key theories of learning that can inform the practice of pre- and post-licensure IPE in workplaces and elsewhere. We would also draw attention to:

- Illeris' work highlighting the interplay between the content of learning, emotions, and the context of learning [85]
- 'threshold concepts' (discipline-related concepts essential for understanding and creating knowledge in a discipline) and 'conceptual thresholds' (moments of enlightenment and leaps in a learning journey, often referred to as 'aha!' moments) [86]
- non-formal learning and tacit knowledge [87]
- professional identity formation and agency [88–90].

Chapters 4 and 12 in this book discuss important social learning theories for this context (social cognitive, social constructivism, and socio-cultural). We would also highlight the socio-material perspective in which workplace learning and workplace practices emerge from dynamic relationships between people and material things (such as technology, clinical notes, and space), through embodied practice (as enacted through the body and emotions) and influenced by culture [91–94]. Activity theory [95, 96] or cultural-historical activity theory (CHAT) [97], actor-network theory [98–100], and complexity theory [101–103] are the most popular examples of socio-material theories.

Social identity theory [104] based on the contact theory of Allport [105] includes the contact hypothesis. This suggests that hostility between social groups would be reduced if members experienced greater contact, with each group having equal status: interaction should be conducted in a cooperative atmosphere, and participants should be working towards a common goal. The contact hypothesis provides a rationale for IPE and guides its design and delivery [106, 107].

appealing to shared interest in delivering good and safe care to *patients* (the 'object' of 'activity' [108]). This is most obvious within post-licensure IPE, particularly in relation to quality improvement. People may not even notice they are engaged in IPE. Participants from diverse backgrounds are focused on and motivated by their shared practice-based problem.

At the pre-licensure level, where students are highly focused on their specific professional knowledge, establishing relevance and authenticity requires active attention from curriculum developers and facilitators. Learning opportunities should be aligned with the participants' concerns, interests, and levels of expertise. This can be particularly challenging for IPE because the diversity of concerns, interests, and expertise is normally greater within an interprofessional group. Students have been shown to engage more fully with IPE when they perceive it as supportive of their own, profession-specific, development [89].

The Perceived Demands of the Learning Context

Learners' *perceptions* of the learning environment and what is expected from them affect what and how they learn [109, 110]. Overloading learners is known to encourage a reproducing (surface) approach to learning, faulty learning, disengagement or a strategic approach to studying [111]. Thus, for example, it is not helpful to place an optional, experiential IPL opportunity shortly before a high-stakes summative assessment. While this may seem too obvious to mention, it is surprisingly easy to overlook important conflicting demands faced by one or more of the groups from whom participation is desired. Such oversights are more easily avoided when care is taken to include a member of each participating group in the planning process.

Learners' perceptions are shaped by explicit and implicit messages. Explicit messages include the following:

- the course description (as published and as spoken by facilitators)
- the intended learning outcomes (normally stated in a course handbook or webpage)
- learning materials and assessment requirements.

These all convey the 'target understanding' [112] that curriculum developers, tutors, and examiners have in mind (with any lack of alignment creating confusion).

Implicit messages, the 'hidden curriculum' [2], include the following:

- perceptions of the importance of a learning opportunity based on, for example, who chooses to attend or otherwise contribute and the attitudes they display
- attendance or assessment requirements
- the physical space (e.g. located in a bright and airy room; configured so participants can face one another rather than sitting in rows all facing a facilitator or presenter)
- access to adequate technology and refreshments
- timing of events (e.g. the event is pushed to the fringes of working time, such as late on Friday afternoon)
- absence of particular professional groups
- many other subtle ways of reinforcing or undermining the official explicit messages.

IPE can be enhanced or undermined by explicit and implicit messages delivered and *perceived* about its relationship to the dominant activities of uniprofessional education and practice. Interestingly, the connection between individual learning and IPE is complex. Work linking threshold concepts (Box 14.4) and disjunction (getting stuck in learning) has examined the relationship between autonomy in learning and the personal challenge of student experiences in clinical education. It found that autonomy should be regarded as something that develops in relation to others, not as a merely individual phenomenon, and that IPE enhances students' experiences of autonomy [113].

Learning Spaces

There is increasing interest in the notion of space. Interest in the use and design of physical space (and virtual worlds [VWs], discussed in the section Simulation: 'Real-world' and Virtual below) is consistent with increased attention to socio-material perspectives on learning (Box 14.4). For example, Temple [114, 115] reviewed research into the built environment of universities, linking this to the organisational nature of higher education in terms of how universities are governed and managed including: changing relations with their students; research relating to how students learn; and factors influencing the learning process. He developed a useful agenda for future research, much of which remains to be done. More recent contributions to this discourse have emphasised evolving space usage, the importance of spaces that can be used flexibly, and alignment between space and the curriculum [116–118]. As a counterbalance, Thomas challenges the dominance of physical spaces in the discourse about space and learning [119]. He argues that there is little recognition that our conceptions of learning are bounded by the 'physical situatedness' of learning itself, creating unhelpful and inadequately challenged distinctions between conceptions on learning in different types of space (e.g. classrooms, libraries, cafés, clinical areas, and online).

There has been relatively little consideration of space as a site of learning and more particularly as a site of power in IPE. Kitto and colleagues [120] argued that space and place are under-conceptualised in the health professions' literature suggesting that:

- There is a need to examine how the notion of space is utilised for learning, and the impact that place has on expectations, types, and evaluations of learning.
- Health sector research into space and place has predominantly focused on patient and family experiences of care; interest in space, place, and professional practice has grown slightly, but research is developing within, rather than across, professions. There is little research that explores space, place, and learning together.
- Research using a broader range of methodologies would enable greater understanding of how students and practitioners learn in clinical settings, and how both people and places affect one another.

Nordquist and colleagues [121], alongside providing a commentary on a collection of papers addressing the impact of space on learning in professional education and IPE, discourage designing physical learning spaces based on specific educational methods (potential 'living museums' p. 81), but instead encourage the design of flexible hybrid spaces for learning that can remain relevant as learning and teaching changes. They advocate strengthening alignment between the curriculum and space provision through sequential consideration of:

- the vision of the curriculum: an exploration of the need for and type of formal and informal spaces
- existing spaces and an analysis of how these might be classified
- the gap between the vision and what existing spaces provide, to guide redevelopment or creation of new learning spaces.

This is less radical than Thomas's vision of blurred distinction between physical and online spaces for learning, since these increasingly meld together in a learner's multifaceted learning experiences and it is difficult to articulate 'where' particular learning occurs [119]. Thomas also calls for learning spaces to be designed as adaptive, malleable, and enchanting spaces which provide opportunities for emergent types of learning. Moving beyond the discourses of physical and virtual learning spaces, Savin-Baden [122, 123] argues for the need to see spaces between people and places in terms of:

- *Territorial spaces* between the tribes of academia, whether disciplinary, professional, or departmental. These are places in which understandings about issues of power, status, and emphasis are important.
- *Space between learner and teacher*: the concerns and agendas of learners and teachers are different spaces with diverse emphases, and such spaces are often complex and difficult to manage. Often, these spaces are not just different in territory but also in language and social practices.

The notion of translation is useful in understanding the complexity of these forms of space. Translation is normally seen as finding parallels between two languages or as a means of mediation between professions' languages. Yet – in the process of translation – words, discourse, and practices change and their meanings are often mislaid and misunderstood. The difficulty with attempting to translate different professionals' ideas into something simplified and accessible to all IPE participants can make matters worse. Perhaps these spaces between learners, and between learners and teachers, should not be managed through translation, but by acknowledging the differences and the complexity of bridging these. IPE participants should be given opportunities and encouragement to ask for clarification when they are uncertain, or when ideas and language are unclear.

Interprofessional Education and the Curriculum

We argue that there is no ideal or essential location for IPE within a curriculum, rather that there are many opportunities for enhancing learning through IPE. One useful model to consider what, when, and how, is the taxonomy developed by Bainbridge and Wood [124] who argue for three levels of IPE based on stages of becoming a professional.

1. *Exposure*: junior students engage in learning experiences with peers from other professions. The intention here is mainly 'learning about', gaining an understanding

of other professions, whilst making sense of their own. Learning tends to be through simulation, case-based learning (including e-learning), group work, discussion, and reflection.

2 *Immersion:* senior students have practice-based collaborative learning experiences to examine the strengths and boundaries of their profession whilst beginning to develop an interprofessional world view. Challenging learning activities include appreciative inquiry, problem-based learning (PBL), team projects, and group activities in clinical settings.
3 *Mastery*: pre- and post-licensure learners are prompted to use critical thinking to develop a deep sense of their own context, profession, and values in relation to other professions. At this stage students have useful profession-specific knowledge and experiences that they can share, and they tend to be hungry for new experiences [66]. Learning may include involvement in cutting-edge practices, emerging technologies, activities such as student run (or guided) clinics, and workplace learning.

Curriculum Alignment in an Era of Competency-based Education

The argument for 'constructive alignment' [125] highlights the connection, and impacts of disconnection or misalignment, between intended learning outcomes, the learning opportunities provided, and assessment (both what and how). In the field of medical education we are in a time of competency-based education (CBE) and a number of interprofessional competency frameworks have been published [126] with competence being defined as what graduates should be able to do in practice [127]. Two influential examples are: the *National Interprofessional Competency Framework* published by the Canadian Interprofessional Health Collaborative (CIHC) in 2010 [128] and the 2011 Interprofessional Education Collaborative's (IPEC) list of core competencies for interprofessional collaborative practice from the USA [129], updated in 2016 [130].

Learning outcomes or competencies are defined within each health profession's accreditation standards. In many countries these now include interprofessional standards. However, within a given country professions may use different language to capture interprofessional competencies (or outcomes, or capabilities, etc.) – this opens up spaces for misalignment, gaps, or unresolvable profession-specific constraints. Until professions jointly agree and adopt national interprofessional competencies, the design and assessment of IPE will present difficulties related to differing uniprofessional requirements for accreditation.

Delivering Effective Interprofessional Education

Most modes of educational delivery potentially have some application in IPE, but some are more naturally suited to the task; we will discuss case-based and PBL, simulation, shadowing, and clinical work, before discussing online IPE (see also Chapters 10–12). Of course, these modes of delivery overlap.

Subsequently, we will consider the roles of facilitation and assessment (see also Chapters 9, 13, and 21–25).

Case-based Learning (CBL) and Problem-based Learning (PBL)

CBL and PBL trigger learning by adapting real cases and incidents. Patient cases and narratives help link theory to practice [131]. All the participating professions are likely to be familiar with learning in this way and scenarios can be tailored to be relevant to all learners. Well-selected or well-crafted triggers are vital to ensure that each profession can make a valued contribution. One study has shown that augmenting case triggers with an interprofessional team reasoning framework and video examples of interprofessional interactions improved students' perceptions of team skills and their case presentations [132].

It is important to clarify what CBL or PBL means to each professional group as there may be different conceptions, based on different prior experiences of CBL/PBL processes adapted to suit profession-specific needs or traditions. Different expectations may underpin unanticipated difficulties in PBL processes during IPE [133]. There is a broad range of PBL approaches and practices worldwide and the diversity is growing. Differences occur in respect of constituent dimensions, such as problem type, form of interaction, knowledge focus, form of facilitation, focus of assessment, and learning emphasis. Within a particular IPE experience, the conceptions of PBL brought by participants form a particular 'constellation' [134] of overlapping expectations, both harmonious and clashing. To date, there is relatively little understanding of the impact of different PBL constellations on IPE.

Simulation: 'Real-world' and Virtual

The term simulation covers everything from table-top exercises and simple role play (e.g. telephone call) to medium-fidelity simulation in clinical skills centres, and on to high-fidelity clinical simulations supported by sophisticated technology and/or highly skilled simulated patients (see also Chapter 11). Palaganas and colleagues [135] argue that simulation and IPE are natural partners, helping to overcome some of the logistical challenges of IPE in other contexts – providing a motivating, engaging learning environment and providing a safe environment (removing risks to patients and facilitating a psychologically safe environment to explore challenging issues such as social hierarchy, diversity, and divisions). Examples of simulation for IPL include extending earlier uniprofessional learning through the development of interprofessional scenarios for rehearsing aspects of communication with simulated patients [62]. This is a good example of a spiral curriculum in action (Chapter 5). Many other examples of simulation-based IPE centre on managing complex cases or deteriorating patients (represented by manikins or actors in physical simulations, and variously portrayed in online simulations) in order to rehearse and reflect upon clinical skills, interprofessional teamwork, leadership, and effective workload management [54, 136–138]. Simulation-based learning generally includes an element of role play – learning and teaching processes linked to an IPE example are explored by van Soeren and colleagues [139], whose findings resonate with the wider role play literature (see also Chapter 11).

Teaching and learning through the use of technologies such as virtual worlds has expanded rapidly in recent years [123, 140, 141]. VWs are virtual learning spaces such as Second Life, Blue Mars, and Kaneva, comprising open computer-based simulations populated and built by online communities in which people can create a personal avatar. For example, an IPE pilot [142] used an existing virtual hospital in Second Life, augmented with documents in Wikispaces, to enable health care students from four professional groups and two universities to make profession-specific clinical assessments of a simulated elderly patient whose care needs had increased. They then meet in the virtual environment to develop a collaborative care plan. The pilot was generally well received. It partly mitigated lack of opportunity for collaborative learning in real clinical placements, and raised students' awareness of the roles of other health care professionals; opportunities to improve the VW simulation and learning experience were identified.

Many designers and tutors, especially those working in distance-learning contexts, describe VWs as supporting social interaction and learners' motivation [143, 144], dialogic learning [145], action learning [146], communal constructivism [147], experiential learning [148], role-playing [149], and PBL [150]. Attempts have been made to map these pedagogical practices across VWs. Literature syntheses examining VW usage highlighted the prevalence of 'simulation of space' [151] and 'collaborative simulation activities' [152]. Thus, returning to our earlier discussion of spaces, VWs could be seen as offering IPE a means to develop and use teaching spaces in different ways.

Shadowing

Reciprocated shadowing, in which health care students or practitioners from different professions observe each other at work in clinical settings, when complemented by associated discussion and reflection, can make an excellent contribution to interprofessional understanding of roles, responsibilities, constraints, expertise, and models of practice. Concentration wanes during passive observation and therefore observations need to be actively processed to become integrated with wider professional learning. An interprofessional shadowing experience needs both structure and follow-up activities to promote IPL and increase the chance that that learning will be integrated in subsequent professional practice. Some wider IPE initiatives include shadowing elements [153–155].

Clinical Work in Interprofessional Student Teams

Several IPE initiatives, in a variety of clinical contexts, have involved interprofessional student teams providing care under the supervision of qualified practitioners [54, 56, 68]. Two main models are interprofessional training wards (IPTW) and student-run clinics (SRC). These models show how a rolling programme of IPE in a service delivery setting can, over time, allow large numbers of students to rehearse and reflect on interprofessional teamwork. Follow-up studies show that students retain strong and largely positive memories of these types of IPE [156, 157]. However, these models can be vulnerable to sudden changes in the clinical area, resulting in loss of staff to provide the level of supervision students require, or a change in the caseload rendering a clinical area too demanding for student teams.

Training wards originated in Sweden in 1996 and much has been written about them since their inception. Many studies have shown that students find IPTW experiences meaningful (see for example evaluations from Sweden [54], the UK [68], and Australia [158]). Patients tend to be highly satisfied with care provided by interprofessional student teams [68,159]. However, such wards are not easy to implement: careful planning and attention to legal and bureaucratic requirements within a country's health system are required, they are resource intensive, and with large student numbers across the health professions it is difficult to provide adequate time on the ward for each student and team. Studies have also noted that supervising student teams and facilitating effortful reflection can be a draining role for clinicians and faculty [156].

In the United States SRCs are becoming more widespread and were originally implemented to provide free access to health care for uninsured and socially deprived or underserved populations [159]. They are also referred to as student-led or student-assisted clinics, and involve students from a wide range of health professions under appropriate supervision. A systematic review suggested that SRCs give students 'the optimal and most realistic form of learning by doing' [160, p. 250]. However, interprofessional student clinics may cost more than conventional clinical placements for students [161] and, as with much health professional education, evaluation through longer-term follow up of impact is still required.

A related model of IPE includes direct contact with patients but stops short of providing care; this involves interprofessional student team members assessing patients with complex needs and conducting an interdisciplinary case conference to integrate their findings and develop a care plan [70]. This model of IPE forms the basis of the health care team challenge (HCTC) – a student team competition held annually at many universities worldwide [162].

Online IPE

With its emphasis on the contact hypothesis [105], IPE has tended to emphasise face-to-face synchronous learning and has encountered well-documented challenges [21–23]. Online IPE may help solve some of the logistical issues of face-to-face activities [163]. At pre-licensure level, online interprofessional modules may be woven into the wider profession-specific curriculum for several professions; this allows large numbers of students across different faculties, and even different institutions, to undertake IPL activities asynchronously and simultaneously [164]. At post-licensure level, examination of computer-mediated communication (CMC) among dispersed members of a rural interdisciplinary health care team illuminated group dynamics in virtual interprofessional teams, providing valuable insights for planners and facilitators of e-IPL [165].

Online IPE has been shown to be effective and a sustainable solution to help learn foundational teamwork knowledge [166] and promote self-directed IPL [167]. It is, however, important to remember that online IPE also requires interaction, which might be achieved through moderated or unmoderated electronic discussion forums within a virtual learning environment (VLE), or through social media [168], in addition to blended learning formats that combine, for example, online and face-to-face learning [169]. A study of online IPE groups of students from six professions working through a social networking site, compared three levels of facilitation and task-structuring. It found that a facilitated group with moderate task-structuring was more successful than both an unfacilitated minimally structured group and a highly facilitator-structured group [170]. This highlights the careful balancing act between too much and too little intervention by curriculum developers and learning facilitators. Blending e-learning with informal self-directed IPL activities, such that learners are encouraged to meet online team members face-to-face when possible, demonstrates that e-learning does not have to be an all or nothing approach [171].

The Role of Facilitation

While many principles of good small group facilitation apply to IPE, some additional dimensions warranting attention include the deliberate heterogeneity of groups, and the aim of harnessing diverse perspectives, skills, and insights of participants from different professions. This is challenging, even for those with considerable uniprofessional facilitation experience [172]. There is a considerable body of literature on facilitation and IPE [173–176] and there continues to be a wide range of staff development programmes in IPE [177]. Unique demands of facilitating online IPE have also been identified. For example, Dalley-Hewer and colleagues highlighted the phenomenon of 'polite agreement' in interprofessional online discussion forums and were concerned that exploring 'meaningful disagreement' respectfully should be a defining feature of IPL. Restructuring 'e-tivities' and scenarios resulted in a more critical discourse which can help learners to reach new and shared conclusions [178].

Recently, Evans and colleagues [179] examined the impact of facilitating IPE on facilitators themselves – currently a relatively under-explored area. The findings indicate undertaking interprofessional facilitation affected facilitators' own interprofessional workplace behaviour, such as working more collaboratively when planning care, treating patients, and supervising students. A recent qualitative synthesis of 12 articles [180] examined the nature of interprofessional facilitation and found that it was influenced by contextual factors, such as the need for good organisational and e-learning support. Facilitator experiences and the use of different facilitation strategies were also found to be important, as were initial preparation, ongoing support, the opportunity for co-facilitation, and the ability to be flexible in terms of adopting a variety of approaches to facilitation. This synthesis helpfully provides guidance on both staff development for interprofessional facilitation and curriculum design, as well as suggestions about how facilitators may best be supported in this challenging role.

Interprofessional co-facilitation is one way to mitigate facilitators' inadequate knowledge of all participating professions and their usual approaches to learning and teaching. The extra cost of additional facilitators may need to be offset by larger groups or reduced contact time. In addition, co-facilitators can feel that their own professional expertise is under the spotlight to a greater extent than in their routine work – an experience that may be enjoyable or nerve-wracking. Co-facilitators need to role model high-quality interprofessional collaboration, otherwise the credibility of the learning experience may be damaged.

IPE researchers have reported that gender balance and the balance of professional membership can affect group dynamics [181]. There may be opportunities for facilitators to allocate group membership in ways that maintain sufficient balance to safeguard productive interprofessional discussion.

Discomfort generated by lack of familiarity with IPE may cause some participants to try to change the nature of the learning experience so that it becomes more familiar. Conflict may be more likely in a group where firmly held professional positions are scrutinised. Facilitators need to develop skills to productively harness the energy of conflict, reflecting it back to participants, and set clear limits on acceptable behaviour [182]. The contact hypothesis suggests how prejudice and its associated conflict might be reduced [106]. It highlights the importance of facilitators drawing out both similarities and differences between participating groups. Most interprofessional facilitators find that conflict remains productive and manageable if there is a central focus on patients and improving the quality of services. Improving the quality of working lives by improving team communication and local processes is also an effective focus at the post-qualification level.

Assessment and Interprofessional Education

The importance of the assessment of interprofessional learning outcomes should be no different from that of uniprofessional or generic outcomes. However, a 2015 review suggested that few IPL activities are assessed and that if they are it is rarely through assessment of performance [183]. An Australian audit indicated that IPL is most frequently assessed through attendance followed by essays and group presentations [184]. Assessment is important to ensure that students do not consider IPE different from their uniprofessional learning – it is not optional or peripheral but important: 'an integral and necessary component in the education of health and human service professionals, regardless of discipline' [185, p. 101].

In 2016, in response to an invitation from the programme committee of the 17th International Ottawa conference on the Assessment of Competence in Medicine and the Health care Professions, an international working group organised several meetings and discussions with colleagues around the globe on the topic of the assessment of IPL. This process, which involved 75 contributors from 15 countries, resulted in the 'International consensus statement on the assessment of interprofessional learning outcomes' [33]. Box 14.5 summarises this statement.

The statement includes a rationale for assessment, similar to that of assessment in health professional education in general with the addition of showing that health professionals are able: 'to meet the needs and expectations of patients, clients and communities, as well as carers and families, for effective cooperation and interprofessional communication between health and social care workers' [33, p. 4].

For qualification and licensure purposes, students' and trainees' competence is assessed at the individual level, although teamwork and group activities may be assessed

BOX 14.5 FOCUS ON: Assessment: summary of the International Consensus Statement on the Assessment of Interprofessional Learning Outcomes [33]

What to assess:

The defined learning outcomes for the particular university and health profession – which will include the broad areas of role understanding, interprofessional communication, interprofessional values, coordination and collaborative decision-making, reflexivity, and teamwork.

How to assess:

Methods should draw on best practice in assessment, which may need to be 'situated and contextualised'. Students should be assessed both individually and within a group or team setting. Suggested approaches include scenario-based MCQs, team-based projects, simulation with observation, practice-based activities, reflective journaling, and oral or written critique of teams that students observe or 'join'. In addition there are numerous tools for the assessment of teamwork and collaborative practice (see for example the collection on the National Center for Interprofessional Practice and Education's website [189]). Assessments may form the basis of an interprofessional portfolio in which students provide evidence of meeting the defined learning outcomes over the course of their programme of study.

Who should assess?

Many jurisdictions require that students should be assessed summatively by a member of their own health profession. However, if interprofessional assessment is part of a programmatic assessment process, this becomes less of an issue as there is not one final examination but rather on-going assessment throughout the programme by different health care professionals, including from their own profession (see Chapter 25 for more on formative and programmatic assessment processes). Evidence of learning may also be shown by peer assessment and multisource feedback, including patient opinion.

BOX 14.6 FOCUS ON: Global developments in IPE

The World Health Organization

The WHO continues to support the need for IPE. In recent documents it has called for health professional education accreditation at a national level to include IPE for collaborative practice [16]. Its five-year action plan for 2017–2021 calls for 'provision of interprofessional education and organization of multidisciplinary care, including recommendations on skills mix and competencies to achieve integrated people-centred care' [191, p. 11].

Japan

The Japan Association for Interprofessional Education (JAIPE) was founded in 2008 by 46 representatives from health care, welfare, universities, hospitals, and other institutions. Health and social care in Japan recognises the need for interprofessional collaborative practice as the country becomes a super-ageing society. IPE is being implemented across academic disciplines and aims to help learners master team-working competencies [192]. It is still a developing field within higher education.

USA

- The IPEC, which includes representatives from professional associations of dentistry, pharmacy, nursing, public health, osteopathic medicine, and allopathic medicine, has published an influential report on core competencies for interprofessional collaborative practice [129]. The number of US medical schools mandating IPE increased from 56 in 2007–2008 to 130 in 2014–2015 [193].
- The National Center for Interprofessional Practice and Education in Minneapolis, Minnesota, was developed through cooperative agreement with the Health Resources and Services Administration (HRSA), the primary federal agency of the US Department of Health and Human Services, with the mandate of improving access to health care and also has funding through three private foundations. Part of its mission is to conduct rigorous evaluation of IPECP. Its work began in October 2012 [194].

Australia

The standards of all health professions accredited with AHPRA (the Australian Health Practitioner Regulation Agency) include items relating to interprofessional practice. However, there is wide diversity in how, when, and what is included in health professional curricula at the individual institutional level [184].

Africa and Asia

IPE has a new network in Africa (http://afriPEN.org) and is being implemented in developing countries with similar challenges to those in developed countries [23]. Other examples in Asia are Indonesia [195] and Malaysia [196].

as part of wider in-course assessments. Lingard has contrasted the individual emphasis to a collectivist approach to competence [186]. She suggests three key premises underpinning a collectivist approach (p. 55):

1 Competence is achieved through participation in authentic situations.
2 Competence is distributed across a network of persons and artefacts.
3 Competence is a constantly evolving set of multiple, interconnected behaviours enacted in time and space.

The collectivist discourse takes a systems approach and acknowledges the intricacies of collaborative practice. It will necessitate a rethink of the current assessment infrastructure in order to measure team performance and a move from thinking about '*competent practitioners* to talking also about *competent performances* of teams' [186, p. 67, italics in the original]. This will not be easy and is likely to be more feasible and acceptable post-licensure. Orchard, drawing on the work of Kvarnström [187], advocates post-licensure assessment to focus on team dynamics and the knowledge contribution from each team member, taking into account the organisational environment [188]. Of course, health professionals may work in different teams and wider collaborations depending on their role and location within a health service, adding to the complexity of deciding when, where, and what to assess. Some 'teams' come together for very specific tasks such as the management of a cardiac arrest – there the importance is the role and not the person. Other teams work together over much longer periods of time.

Conclusion

Whichever beach you are sitting on, it is likely to feel as if interest in IPE comes and goes like the tide. This applies to areas of practice as well as geographical locations, and beaches differ with respect to the amount of variation between high and low tide. At the time of writing there are exciting developments in IPE worldwide [190], a few of which are described in Box 14.6.

References

1 Freeth, D., Hammick, M., Reeves, S. et al. (2005). *Effective Interprofessional Education: Development, Delivery and Evaluation*. Oxford: Blackwell Publishing.
2 Hafferty, F.W. and O'Donnell, J.F. ed. (2015). *The Hidden Curriculum in Health Professional Education*. Lebanon, NH: Dartmouth College Press.
3 CAIPE (2006). CAIPE re-issues its statement on the definition and principles of interprofessional education. *CAIPE Bulletin* 26: 3.
4 CAIPE (2016). Home page. http://caipe.org (accessed 12 March 2017).
5 World Health Organization (1988). *Learning Together to Work Together*. Geneva: WHO.
6 World Health Organization (2010). *Framework for Action on Interprofessional Education and Collaborative Practice*. Geneva: WHO.
7 Bainbridge, L. and Wood, V.I. (2012). The power of prepositions: learning with from and about others in the context of interprofessional education. *Journal of Interprofessional Care* 26 (6): 452–458.
8 http://www.tandfonline.com/action/authorSubmission?show=instructions&journalCode=ijic20
9 CIHC – Canadian Interprofessional Health Collaborative (2010). http://www.cihc.ca/files/CIHC_IPCompetencies_Feb1210r.pdf (accessed 12 March 2017).
10 Barr, H. (2005). *Interprofessional Education: Today, Yesterday and Tomorrow* (Revised). London: LTSN for Health Sciences and Practice https://www.unmc.edu/bhecn/_documents/ipe-today-yesterday-tmmw-barr.pdf (accessed 12 March 2017).
11 Meads, G. and Ashcroft, J. (2005). *The Case for Interprofessional Collaboration*. Oxford: Blackwell Publishing.
12 Thistlethwaite, J.E. (2012). Interprofessional education: a review of context, learning and the research agenda. *Medical Education* 46: 58–70.
13 Salas, E. and Rosen, M.A. (2013). Building high reliability teams: progress and some reflections on teamwork training. *BMJ Quality and Safety* 22 (5): 369–373.
14 Epstein, N.E. (2014). Multidisciplinary in-hospital teams improve patient outcomes: a review. *Surgical Neurology International* 5 (Suppl 7): S295–S303.
15 Berwick, D.M., Nolan, T.,.W., and Whittington, J. (2008). The triple aim: care, health and cost. *Health Affairs* 27: 759–769.
16 World Health Organization (2017). *National Health Workforce Accounts: A Handbook*. Draft for Consultation. Geneva: WHO.
17 Frenk, J., Chen, L., Bhutta, Z.A. et al. (2010). Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. *The Lancet* 376 (9756): 1923–1958.
18 Engeström, Y. (2008). *From Teams to Knots: Activity-Theoretical Studies of Collaboration and Learning at Work*. New York: Cambridge University Press.
19 Bleakley, A. (2014). *Patient-Centred Medicine in Transition. The Heart of the Matter*, *Advances in Medical Education 3*. Basel: Springer International Publishing.
20 Paradis, E., Pipher, M., Cartmills, C. et al. (2017). Articulating the idea: 50 years of interprofessional collaboration. *Medical Education* 51: 861-872.
21 Lawlis, T.R., Anson, J., and Greenfield, D. (2014). Barriers and enablers that influence sustainable interprofessional education: a literature review. *Journal of Interprofessional Care* 28 (4): 305–310.
22 Reeves, S., Fletcher, S., Barr, H. et al. (2016). A BEME systematic review of the effects of interprofessional education: BEME guide no 39. *Medical Teacher* 38 (7): 656–668.
23 Sunguya, B.F., Hinthong, W., Jimba, M., and Yasuoka, J. (2014). Interprofessional education for whom? – challenges and lessons learned from its implementation in developed countries and their application to developing countries: a systematic review. *PLoS ONE* 9 (5): e96724.
24 Institute of Medicine (2015). *Measuring the Impact of Interprofessional Education on Collaborative Practice and Patient Outcomes*. Washington, DC: The National Academies Press.
25 Billet, S.R. (2014). Securing intersubjectivity through interprofessional workplace learning experiences. *Journal of Interprofessional Care* 28 (3): 206–211.
26 Gregory, L.R., Hopwood, N., and Boud, D. (2014). Interprofessional learning at work: what spatial theory can tell us about workplace learning in an acute care ward. *Journal of Interprofessional Care* 28 (3): 200–205.
27 Reeves, S., Goldman, J., Gilbert, J. et al. (2011). A scoping review to improve conceptual clarity of interprofessional interventions. *Journal of Interprofessional Care* 25 (3): 167–174.
28 Reeves, S., Boet, S., Zierler, B., and Kitto, S. (2015). Interprofessional education and practice guide no. 3: evaluating interprofessional education. *Journal of Interprofessional Care* 29 (4): 305–312.
29 Teodorczuk, A., Khoo, T.K., Morrissey, S., and Rogers, G. (2016). Developing interprofessional education: putting theory into practice. *The Clinical Teacher* 13 (1): 7–12.

30 Thistlethwaite, J.E. (2016). Research and evaluation: the present and the future. In: *Leading Research and Evaluation in Interprofessional Education and Practice* (ed. D. Forman, M. Jones and J.E. Thistlethwaite), 355–368. London: Palgrave-Macmillan.
31 Frank, J.R., Snell, L., and Sherbino, J. ed. (2015). *CanMEDS 2015 Physician Competency Framework*. Ottawa: Royal College of Physicians and Surgeons of Canada.
32 Lockeman, K.S., Dow, A.W., DiazGranados, D. et al. (2016). Refinement of the IPEC competency self-assessment survey: results from a multi-institutional study. *Journal of Interprofessional Care* 30 (6): 726–731.
33 Rogers, G., Thistlethwaite, J.E., Anderson, E.S. et al. (2016). International consensus statement on the assessment of interprofessional learning outcomes. *Medical Teacher* 39 (4): 347–359.
34 Gilman, S.C., Chokshi, D.A., Bowen, J.L. et al. (2014). Connecting the dots: Interprofessional health education and delivery system redesign at the Veterans Health Administration. *Academic Medicine* 89 (8): 1113–1116.
35 Brashers, V., Phillips, E., Malpass, J., and Owen, J. (2015). *Measuring the Impact of Interprofessional Education on Collaborative Practice and Patient Outcomes*, 67–133. Washington, DC: National Academies Press.
36 Pauze, E. and Reeves, S. (2010). Examining the effects of interprofessional education on mental health providers: findings from an updated systematic review. *Journal of Mental Health* 19 (3): 258–271.
37 Abu-Rish, E., Kim, S., Choe, L. et al. (2012). Current trends in interprofessional education of health sciences students: a literature review. *Journal of Interprofessional Care* 26 (6): 444–451.
38 Brody, A. and Galvin, J. (2013). A review of interprofessional dissemination and education interventions for recognizing and managing dementia. *Gerontology & Geriatrics Education* 34 (3): 225–256.
39 Broyles, L., Conley, J., Harding, J., and Gordon, A.J. (2013). A scoping review of interdisciplinary collaboration in addictions education and training. *Journal of Addictions Nursing* 24 (1): 29–36.
40 Reeves, S., Perrier, L., Goldman, J., et al. (2013). Interprofessional education: Effects on professional practice and healthcare outcomes (update). *Cochrane Database of Systematic Reviews 3* (Art. No.: CD002213). doi: 10.1002/14651858.CD002213.pub3.
41 Brandt, B., Lutfiyya, N., King, J., and Chioreso, C. (2014). A scoping review of interprofessional collaborative practice and education using the lens of the triple aim. *Journal of Interprofessional Care* 28 (5): 393–399.
42 Sockalingam, S., Tan, A., Hawa, R. et al. (2014). Interprofessional education for delirium care: a systematic review. *Journal of Interprofessional Care* 28 (4): 345–351.
43 Reeves, S., Palaganas, J., and Zierler, B. (2017). An updated synthesis of review evidence of interprofessional education. *Journal of Allied Health* 46 (1): 56–61.
44 Song, F., Hooper, L., and Loke, Y.K. (2013). Publication bias: what is it? How do we measure it? How do we avoid it? *Open Access Journal of Clinical Trials* (5): 71–78.
45 Shafer, M.B., Tebb, K.B., Pantell, R.H. et al. (2002). Effect of a clinical practice improvement intervention on chlamydial screening among adolescent girls. *JAMA: The Journal of the American Medical Association* 288 (22): 2846–2852.
46 Janson, S.L., Cooke, M., McGrath, K.W. et al. (2009). Improving chronic care of type 2 diabetes using teams of interprofessional learners. *Academic Medicine* 84 (11): 1540–1548.
47 Taylor, C.R., Hepworth, J.T., Buerhaus, P.I. et al. (2007). Effect of crew resource management on diabetes care and patient outcomes in an inner-city primary care clinic. *Quality and Safety in Health Care* 16 (4): 244–247.
48 Capella, J., Smith, S., Philp, A. et al. (2010). Teamwork training improves the clinical care of trauma patients. *Journal of Surgical Education* 67 (6): 439–443.
49 O'Leary, K.J., Sehgal, N.L., Terrell, G., and Williams, M.V. (2012). Interdisciplinary teamwork in hospitals: a review and practical recommendations for improvement. *Journal of Hospital Medicine* 7 (1): 48–54.
50 Weaver, S.J., Rosen, M.A., DiazGranados, D. et al. (2010). Does teamwork improve performance in the operating room? A multilevel evaluation. *Joint Commission Journal on Quality and Patient Safety* 36 (3): 133–142.
51 Cameron, K.A., Engel, K.G., McCarthy, D.M. et al. (2010). Examining emergency department communication through a staff-based participatory research method: identifying barriers and solutions to meaningful change. *Annals of Emergency Medicine* 56 (6): 614–622.
52 Helitzer, D.L., Lanoue, M., Wilson, B. et al. (2011). A randomized controlled trial of communication training with primary care providers to improve patient-centeredness and health risk communication. *Patient Education and Counseling* 82 (1): 21–29.
53 Fletcher, J., King, K., Christo, J. et al. (2014). An evaluation of an Australian initiative designed to improve interdisciplinary collaboration in primary mental health care. *Evaluation and Program Planning* 45: 29–41.
54 Bridges, D.R., Davidson, R.A., Soule Odegard, P. et al. (2011). Interprofessional collaboration: three best practice models of interprofessional education. *Medical Education Online* 16: 6035.
55 Pelling, S., Kalen, A., Hammar, M., and Wahlström, O. (2011). Preparation for becoming members of health care teams: findings from a 5-year evaluation of a student interprofessional training ward. *Journal of Interprofessional Care* 25 (5): 328–332.
56 Ponzer, S., Hylin, U., Kusoffsky, A. et al. (2004). Interprofessional training in the context of clinical practice: goals and students' perceptions on clinical education wards. *Medical Education* 38 (7): 727–736.
57 Lennox, A. and Anderson, E.D. (2009). The Leicester model of interprofessional education: developing, delivering and learning from student voices for 10 years. *Journal of Interprofessional Care* 23 (6): 557–573.
58 Brewer, M.L. and Barr, H. (2016). Interprofessional education and practice guide no. 8: team-based interprofessional practice placements. *Journal of Interprofessional Care* 30 (6): 747–753.
59 Shrader, S., Hodgkins, R., Laverentz, D. et al. (2016). Interprofessional education and practice guide no. 7: development, implementation, and evaluation of a large-scale required interprofessional education foundational programme. *Journal of Interprofessional Care* 30 (5): 615–619.
60 McFadyen, A.K., Webster, V.S., Maclaren, W.M., and O'Neill, M.A. (2010). Interprofessional attitudes and perceptions: results from a longitudinal controlled trial of pre-registration health and social care students in Scotland. *Journal of Interprofessional Care* 24 (5): 549–564.
61 Buckley, S., Hensman, M., Thomas, S. et al. (2012). Developing interprofessional simulation in the undergraduate setting: experience with five different professional groups. *Journal of Interprofessional Care* 26 (5): 362–369.
62 Cooke, S., Chew, G., Boggis, C., and Wakefield, A. (2003). 'I never realised that doctors were into feelings too': changing student perceptions though interprofessional education. *Learning in Health and Social Care* 2 (2): 137–146.
63 Mires, G., Williams, F., Harden, R., and Howie, P. (2001). The benefits of a multiprofessional education programme can be sustained. *Medical Teacher* 23 (3): 300–304.
64 Hawkes, G., Nunney, I., and Lindqvist, S. (2013). Caring for attitudes as a means of caring for patients–improving medical, pharmacy and nursing students' attitudes to each other's professions by engaging them in interprofessional learning. *Medical Teacher* 35 (7): e1302–e1308.
65 Lehrer, M.D., Murray, S., Benzar, R. et al. (2015). Peer-led problem-based learning in interprofessional education of health professions students. *Medical Education Online* 20: doi: 10.3402/meo.v20.28851.
66 Mu, K., Chao, C.C., Jensen, G.M., and Royeen, C.B. (2004). Effects of interprofessional rural training on students' perceptions of interprofessional health care services. *Journal of Allied Health* 33 (2): 125–131.

67 McNair, R., Stone, N., Sims, J., and Curtis, C. (2005). Australian evidence for interprofessional education contributing to effective teamwork preparation and interest in rural practice. *Journal of Interprofessional Care* 19 (6): 579–594.

68 Reeves, S., Freeth, D., McCrorie, P., and Perry, D. (2002). 'It teaches you what to expect in future …': interprofessional learning on a training ward for medical, nursing, occupational therapy and physiotherapy students. *Medical Education* 36 (4): 337–344.

69 Anderson, E.S. and Thorpe, L.N. (2014). Students improve patient care and prepare for professional practice: an interprofessional community-based study. *Medical Teacher* 36: 495–504.

70 Shiyanbola, O., Randall, B., Lammers, C. et al. (2014). Impact of an Interprofessional diabetes education model on patient health outcomes: a longitudinal study. *Journal of Research in Interprofessional Practice and Education* 4 (2): 2–21.

71 Gilbert, J., Camp, R., Cole, C. et al. (2003). Preparing students for interprofessional teamwork in health care. *Journal of Interprofessional Care* 14 (3): 223–235.

72 Byrnes, V., O'Riordan, A., Schroder, C. et al. (2012). South Eastern Interprofessional Collaborative Learning Environment (SEIPCLE): nurturing collaborative practice. *Journal of Research in Interprofessional Practice & Education* 2 (2): 169–186.

73 Horbar, J.D., Rogowski, J., Plsek, P.E. et al. (2001). Collaborative quality improvement for neonatal intensive care. *Pediatrics* 107 (1): 14–22.

74 Quinn, K., Hudson, P., Ashby, M., and Thomas, K. (2008). Palliative care: the essentials: evaluation of a multidisciplinary education program. *Journal of Palliative Medicine* 11 (8): 1122–1129.

75 Treadwell, M., Frank, L., and Vichinsky, E. (2002). Using quality improvement strategies to enhance pediatric pain assessment. *International Journal for Quality in Health Care* 14 (1): 39–47.

76 Morey, J., Simon, R., Gregory, D. et al. (2002). Error reduction and performance improvement in the emergency department through formal teamwork training: evaluation results of the MedTeams project. *Health Services Research* 37 (6): 1553–1561.

77 Hunter, J., Watt-Watson, J., McGillion, M. et al. (2008). An interfaculty pain curriculum: lessons learned from six years' experience. *Pain* 140: 74–86.

78 Engeström, Y., Engeström, R., and Kärkkäinen, M. (1995). Polycontextuality and boundary crossing in expert cognition: learning and problem solving in complex work activities. *Learning and Instruction* 5 (4): 319–336.

79 Nisbet, G., Dunn, S., and Lincoln, M. (2015). Interprofessional team meetings: opportunities for informal interprofessional learning. *Journal of Interprofessional Care* 29: 426–432.

80 Nisbet, G., Lincoln, M., and Dunn, S. (2013). Informal interprofessional learning: an untapped opportunity for learning and change within the workplace. *Journal of Interprofessional Care* 27 (6): 469–475.

81 Reeves, S. and Hean, S. (2013). Why we need theory to help us better understand the nature of interprofessional education, practice and care. *Journal of Interprofessional Care* 27: 1–3.

82 Reeves, S., Goldman, J., Martinianakis, T. et al. (2007). *A Scoping Review to Identity Organizational and Education Theories Relevant for Interprofessional Practice and Education*. Canadian Interprofessional Health Collaborative http://www.cihc.ca/files/publications/ScopingReview_IP_Theories_Dec07.pdf (accessed 9 April 2017).

83 Hean, S., Anderson, E., Bainbridge, L. et al. (2013). IN_2_THEORY – Interprofessional theory, scholarship and collaboration: a community of practice. *Journal of Interprofessional Care* 27: 88–90.

84 Hean, S., Craddock, D., and Hammick, M. (2012). Theoretical insights into interprofessional education: AMEE guide no. 62. *Medical Teacher* 34 (2): e78–e101.

85 Illeris, K. (2007). *How We Learn: Learning & Non-Learning in School and Beyond*. London: Routledge.

86 Meyer, J.H.F. and Land, R. (2005). Threshold concepts and troublesome knowledge (2): epistemological considerations and a conceptual framework for teaching and learning. *Higher Education* 49 (3): 373–389.

87 Eraut, M. (2000). Non-formal learning, implicit learning and tacit knowledge in professional work. In: *The Necessity of Informal Learning* (ed. F. Coffield), 12–31. Bristol: Policy Press.

88 Thistlethwaite, J.E., Kumar, K., and Roberts, C. (2016). Becoming interprofessional: professional identity formation in the health professions. In: *Teaching Medical Professionalism*, 2e (ed. R. Cruess, S. Cruess and Y. Steinert), 140–154. Cambridge Medicine: Cambridge.

89 Green, C. (2012). Relative distancing: a grounded theory of how learners negotiate the interprofessional. *Journal of Interprofessional Care* 27 (1): 34–42.

90 Varpio, L., Aschenbauser, C., and Bates, J. (2017). Tackling wicked problems: how theories of agency can provide new insights. *Medical Education* 51: 353–365.

91 Fenwick, T., Nerland, M., and Jensen, K. (2012). Sociomaterial approaches to conceptualising professional learning and practice. *Journal of Education and Work* 25: 1–13.

92 Gherardi, S. (2012). Why do practices change and why do they persist? Models and explanations. In: *Practice, Learning and Change: Practice-Theory Perspectives on Professional Learning* (ed. P. Hager, A. Lee and A. Reich), 217–232. Dordrecht: Springer.

93 Jones, M. (2014). A matter of life and death: exploring conceptualizations of sociomateriality in the context of critical care. *MIS Quarterly* 38 (3): 895–925.

94 McMurtry, A., Rohse, S., and Kilgour, K.N. (2016). Socio-material perspectives on interprofessional team and collaborative learning. *Medical Education* 50: 169–180.

95 Engeström, Y. (2015). *Learning by Expanding: An Activity-Theoretical Approach to Developmental Research*, 2e. New York: Cambridge University Press.

96 Engeström, Y. (2001). Expansive learning at work: toward an activity theoretical reconceptualization. *Journal of Education and Work.* 14 (1): 133–156.

97 Foot, K.A. (2014). Cultural historical activity theory: exploring a theory to inform practice and research. *Journal of Human Behavior in the Social Environment* 24 (3): 329–347.

98 Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.

99 Bleakley, A. (2012). The proof is in the pudding: putting actor-network-theory to work in medical education. *Medical Teacher* 34: 462–467.

100 Fenwick, T.J. and Edwards, R. (2019). *Actor-Network Theory in Educational Research*. London: Routledge.

101 Bogg, J. and Geyer, R. ed. (2007). *Complexity, Science and Society*. Oxford: Radcliffe Publishing.

102 Bleakley, A. (2010). Blunting Occam's razor: aligning medical education with studies of complexity. *Journal of Evaluation in Clinical Practice* 16: 849–855.

103 Fenwick, T. (2012). Complexity science and professional learning for collaboration: a critical reconsideration of possibilities and limitations. *Journal of Education and Work* 25: 141–162.

104 Tajfel, H. and Turner, J.C. (1986). The social identity theory of intergroup behavior. In: *Psychology of Intergroup Relations* (ed. S. Worchel and L.W. Austin). Chicago: Nelson-Hall.

105 Allport, G.W. (1954). *The Nature of Prejudice*. Reading: Addison-Wesley.

106 Hewstone, M. and Brown, R. (1986). Contact is not enough: an intergroup perspective on the 'contact hypothesis'. In: *Contact and Conflict in Intergroup Encounters* (ed. M. Hewstone and R. Brown), 1–44. Oxford: Blackwell.

107 Furness, P.J., Armitage, H.R., and Pitt, R. (2012). Qualitative evaluation of interprofessional learning initiatives in practice: application of the contact hypothesis. *International Journal of Medical Education* 3: 83–91.

108 Engeström, Y. (1999). Expansive visibilization of work: an activity-theoretical perspective. *Computer Supported Cooperative Work* 8 (1, 2): 63–93.

109 Marton, F. and Saljo, R. (1976). On qualitative differences in learning: II outcomes as a function of the learner's conception of the task. *The British Journal of Educational Psychology* 46 (2): 115–127.

110 Prosser, M. and Trigwell, K. (1999). *Understanding Learning and Teaching: The Experience in Higher Education*. Buckingham: Society for Research into Higher Education and Open University Press.

111 Entwistle, N.J. (2009). *Teaching for Understanding at University: Deep Approaches and Distinctive Ways of Thinking*. Basingstoke: Palgrave Macmillan.

112 Entwistle, N. and Smith, C. (2002). Personal understanding and target understanding: mapping influences on the outcomes of learning. *The British Journal of Educational Psychology* 72 (3): 321–342.

113 Fredholm, A., Savin-Baden, M., Henningsohn, L., and Silén, C. (2015). Autonomy as both challenge and development in clinical education. *Learning, Culture and Social Interaction* 5: 20–27.

114 Temple, P. (2007). *Learning Spaces for the 21st Century: A Review of the Literature*. York: The Higher Education Academy www.heacademy.ac.uk/knowledge-hub/learning-spaces-21st-century (accessed 11 April 2017).

115 Temple, P. (2008). Learning spaces in higher education: an under-researched topic. *London Review of Education* 6 (3): 229–241.

116 Beckers, R., van der Voordt, T., and Dewulf, G. (2015). A conceptual framework to identify spatial implications of new ways of learning in higher education. *Facilities* 33 (1/2): 2–19.

117 SCHOMAS, AUDE, UCISA 2016. The UK Higher Education Learning Space Toolkit. Oxford: Universities & Colleges Information System Association (UCIAS). www.ucisa.ac.uk/learningspace (accessed 11 April 2017).

118 Nordquist, J., Sundberg, K., and Laing, A. (2016). Aligning physical learning spaces with the curriculum: AMEE guide no. 107. *Medical Teacher* 38 (8): 755–768.

119 Thomas, H. (2010). Learning spaces, learning environments and the dis'placement' of learning. *British Journal of Educational Technology* 41 (3): 502–511.

120 Kitto, S., Nordquist, J., Peller, J. et al. (2013). The disconnections between space, place and learning in interprofessional education: an overview of key issues. *Journal of Interprofessional Care* 27 (S2): 5–8.

121 Nordquist, J., Sundberg, K., Kitto, S. et al. (2013). Future learning environments: the advent of a "spatial turn"? *Journal of Interprofessional Care* 27 (sup2): 77–81.

122 Savin-Baden, M. (2008). *Learning Spaces. Creating Opportunities for Knowledge Creation in Academic Life*. Maidenhead: McGraw Hill.

123 Savin-Baden, M. (2015). *Rethinking Learning in an Age of Digital Fluency: Is Being Digitally Tethered a New Learning Nexus?* London: Routledge.

124 Bainbridge, L. and Wood, V.I. (2013). The power of prepositions: a taxonomy for interprofessional education. *Journal of Interprofessional Care* 27 (2): 131–136.

125 Biggs, J.B. and Tang, C. (2011). *Teaching for Quality Learning at University: What the Student Does*, 4e. Maidenhead: McGraw-Hill, Open University Press.

126 Thistlethwaite, J.E., Forman, D., Rogers, G. et al. (2014). Competencies and frameworks in interprofessional education: a comparative analysis. *Academic Medicine* 89: 869–874.

127 Gruppen, L.D., Mangrulkar, R.S., and Kolars, J.C. (2012). The promise of competency-based education in the health professions for improving global health. *Human Resources for Health* 10: 43.

128 CIHC – Canadian Interprofessional Health Collaborative (2010). http://www.cihc.ca/files/CIHC_IPCompetencies_Feb1210r.pdf (accessed 12 April 2017).

129 IPEC – Interprofessional Education Collaborative Expert Panel. (2011). Core Competencies for Interprofessional Collaborative Practice: Report of an Expert Panel. Washington, DC: Interprofessional Education Collaborative. http://www.aacn.nche.edu/education-resources/ipecreport.pdf (accessed 12 April 2017).

130 IPEC - Interprofessional Education Collaborative Expert Panel. (2016). Core Competencies for Interprofessional Collaborative Practice: 2016 Update. Washington, DC: Interprofessional Education Collaborative. https://ipecollaborative.org/uploads/IPEC-2016-Updated-Core-Competencies-Report__final_release_.PDF (accessed 23 April 2017).

131 Thistlethwaite, J.E., Davies, D., Ekeocha, S. et al. (2012). The effectiveness of case-based learning in health professional education: a BEME systematic review. *Medical Teacher* 34: e421–e444.

132 Packard, K., Chelal, H., Maio, A. et al. (2012). Interprofessional team reasoning framework as a tool for case study analysis with health professions students: a randomized study. *Journal of Research in Interprofessional Practice and Education* 2 (3): 250–263.

133 Reeves, S. and Freeth, D. (2000). An innovative interprofessional experience for pre-qualifying medical, nursing, occupational therapy and physiotherapy students. *Journal of Interprofessional Care* 14 (2): 202–204.

134 Savin-Baden, M. (2014). Problem-based learning: new constellations for the 21st century. *Journal of Excellence in College Teaching* 25 (3/4): 197–219.

135 Palaganas, J.C., Epps, C., and Raemer, D.B. (2014). A history of simulation-enhanced interprofessional education. *Journal of Interprofessional Care* 28 (2): 110–115.

136 Freeth, D. and Chaput de Saintonge, M. (2000). Helping medical students become good house officers: interprofessional learning in a skills Centre. *Medical Teacher* 22 (4): 392–398.

137 Ker, J., Mole, L., and Bradley, P. (2003). Early introduction to interprofessional learning: a simulated ward environment. *Medical Education* 37 (3): 248–255.

138 Baker, V., Cuzzola, R., Knox, C. et al. (2015). Teamwork education improves trauma team performance in undergraduate health professional students. *Journal of Educational Evaluation for Health Professions* 12: 36.

139 van Soeren, M., Devlin-Cop, S., MacMillan, K. et al. (2011). Simulated interprofessional education: an analysis of teaching and learning processes. *Journal of Interprofessional Care* 25 (6): 434–440.

140 Hew, K.F. and Cheung, W.S. (2010). Use of three-dimensional (3-D) immersive virtual worlds in K-12 and higher education settings: a review of the research. *British Journal of Educational Technology* 41 (1): 33–55.

141 Wang, F. and Burton, J.K. (2013). Second life in education: a review of publications from its launch to 2011. *British Journal of Educational Technology* 44 (3): 357–371.

142 Davis, D.L., Hercelinskyj, G., and Jackson, L.M. (2016). Promoting Interprofessional collaboration: a pilot project using simulation in the virtual world of second life. *Journal of Research in Interprofessional Practice and Education* 6 (2).

143 Pellas, N. and Kazanidis, I. (2014). Engaging students in blended and online collaborative courses at university level through second life: comparative perspectives and instructional affordances. *New Review of Hypermedia and Multimedia* 20 (2): 123–144.

144 Minocha, S. and Hardy, C. (2016). Navigation and wayfinding in learning spaces in 3D virtual worlds. In: *Learning in Virtual Worlds: Research and Applications*. Part of the "Issues in Distance Education" series (Series Editor: Terry Anderson, Athabasca University, Canada) (ed. S. Gregory, M.J.W. Lee, B. Dalgarno and B. Tynan), 3–41. Athabasca University Press.

145 Petrakou, A. (2010). Interacting through avatars: virtual worlds as a context for online education. *Computers & Education* 54 (4): 1020–1027.

146 Wagner, C. and Ip, R.K.F. (2009). Action learning with second life: a pilot study. *Journal of Information Systems Education* 20 (2): 249–258.
147 Girvan, C. and Savage, T. (2010). Identifying an appropriate pedagogy for virtual worlds: a communal constructivism case study. *Computers & Education* 55 (1): 342–349.
148 Jarmon, L., Traphagan, T., Mayrath, M., and Trivedi, A. (2009). Virtual world teaching, experiential learning, and assessment: an interdisciplinary communication course in second life. *Computers & Education* 53 (1): 169–182.
149 Jamaludin, A., Chee, Y.S., and Ho, C.M.L. (2009). Fostering argumentative knowledge construction through enactive role play in second life. *Computers & Education* 53 (2): 317–329.
150 De Jong, N., Savin-Baden, M., Cunningham, A. M. and Verstegen, D.M.L. (2014). Blended learning in health education: three case studies, *Perspectives in Medical Education* 3(4):278–88.
151 Kim, S.H., Lee, J., and Thomas, M.K. (2012). Between purpose and method: a review of educational research on 3D virtual worlds. *Journal of Virtual Worlds Research* 5 (1): 1–18.
152 Duncan, I., Miller, A., and Jiang, S. (2012). A taxonomy of virtual worlds usage in education. *British Journal of Educational Technology* 43 (6): 949–964.
153 Steven, A., Dickinson, C., and Pearson, P. (2007). Practice-based interprofessional education: looking into the black box. *Journal of Interprofessional Care* 21 (3): 251–264.
154 Price, D., Howard, M., Hilts, L. et al. (2009). Interprofessional education in academic family medicine teaching units. *Canadian Family Physician* 55: 901–901e.5.
155 Wright, A., Hawkes, G., Baker, B., and Lindqvist, S.M. (2012). Reflections and unprompted observations by healthcare students of an interprofessional shadowing visit. *Journal of Interprofessional Care* 26 (4): 305–311.
156 Reeves, S. and Freeth, D. (2002). The London training ward: an innovative interprofessional learning initiative. *Journal of Interprofessional Care* 16 (1): 41–52.
157 Hylin, U., Nyholm, H., Mattiasson, A.-C., and Ponzer, S. (2007). Interprofessional training in clinical practice on a training ward for healthcare students: a two-year follow-up. *Journal of Interprofessional Care* 21 (3): 277–288.
158 Brewer, M.L. and Stewart-Wynne, E.G. (2013). An Australian hospital-based student training ward delivering safe, client-centred care while developing students' interprofessional practice capabilities. *Journal of Interprofessional Care* 27 (6): 482–488.
159 Thistlethwaite, J.E. (2016). Interprofessional education: 50 years and counting. *Medical Education* 50: 1082–1086.
160 Schutte, T., Tichelaar, J., Dekker, R.S. et al. (2015). Learning in student-run clinics: a systematic review. *Medical Education* 49: 249–263.
161 Haines, T.P., Kent, F., and Keating, J.L. (2014). Interprofessional student clinics: an economic evaluation of collaborative clinical placement education. *Journal of Interprofessional Care* 28 (4): 292–298.
162 Moran, M., Boyce, A.R., and Nissen, L. (2011). Competition the new collaboration? Employing inter-team competitiveness to motivate health students to learn together. In: *Sociology of Interprofessional Health Care Practice* (ed. S. Kitto, J. Chesters, J. Thistlethwaite and S. Reeves), 155–168. New York: Nova Science Publishers.
163 Luke, R., Solomon, P., Baptiste, S. et al. (2009). Online interprofessional health sciences education: from theory to practice. *Journal of Continuing Education in the Health Professions* 29 (3): 161–167.
164 Bluteau, P. and Jackson, A. (2009). An eLearning model of interprofessional education. In: *Interprofessional Education: Making It Happen* (ed. P. Bluteau and A. Jackson), 107–121. Basingstoke: Palgrave Macmillan.
165 Vroman, K. and Kovacich, J. (2002). Computer-mediated interdisciplinary teams: theory and reality. *Journal of Interprofessional Care* 16 (2): 159–170.
166 Djukic, M., Adams, J., Fulmer, T. et al. (2015). E-learning with virtual teammates: a novel approach to interprofessional education. *Journal of Interprofessional Education* 29 (5): 476–482.
167 Shoemaker, M.J., Platko, C.M., Cleghorn, S.M., and Booth, A. (2014). Virtual patient care: an interprofessional education approach for physician assistant, physical therapy and occupational therapy students. *Journal of Interprofessional Care* 28 (4): 365–367.
168 Cain, J. and Chretien, K. (2013). Exploring social Media's potential in Interprofessional education. *Journal of Research in Interprofessional Practice and Education* 3 (2).
169 Carbonaro, M., King, S., Taylor, E. et al. (2008). Integration of e-learning technologies in an interprofessional health science course. *Medical Teacher* 30 (1): 25–33.
170 Pittenger, A.L. (2013). The use of social networking to improve the quality of interprofessional education. *American Journal of Pharmaceutical Education* 77 (8): 174.
171 Casimiro, L., MacDonald, C.J., Thompson, T.L., and Stodel, E.J. (2009). Grounding theories of W(e)learn: a framework for online interprofessional education. *Journal of Interprofessional Education* 23 (4): 390–400.
172 Egan-Lee, E., Baker, L., Tobin, S. et al. (2011). Neophyte facilitator experiences of interprofessional education: implications for faculty development. *Journal of Interprofessional Care* 25: 333–338.
173 Silver, I.L. and Leslie, K. (2009). Faculty development for continuing interprofessional education and collaborative practice. *Journal of Continuing Education in the Health Professions* 29 (3): 172–177.
174 Cavanaugh, J.T. and Konrad, S. (2012). Fostering the development of effective person centered healthcare communication skills: an interprofessional shared learning model. *Work* 41 (3): 293–301.
175 Sargent, J., Hill, T., and Breau, L. (2010). Development and testing of a scale to assess Interprofessional education (IPE) facilitation skills. *Journal of Continuing Education in the Health Professions* 30 (2): 126–131.
176 Evans, S., Knight, T., Sønderlund, A., and Tooley, G. (2014). Facilitators' experience of delivering asynchronous and synchronous online interprofessional education. *Medical Teacher* 36 (12): 1051–1056.
177 West, C., Graham, L., Palmer, R.T. et al. (2016). Implementation of interprofessional education (IPE) in 16 US medical schools: common practices, barriers and facilitators. *Journal of Interprofessional Education & Practice* 4: 41–49.
178 Dalley-Hewer, J., Clouder, D.L., Jackson, A. et al. (2012). Facilitating critical discourse through "meaningful disagreement" online. *Journal of Interprofessional Care* 26 (6): 472–478.
179 Evans, S., Shaw, N., Ward, C., and Hayley, A. (2016). "Refreshed… reinforced…reflective": a qualitative exploration of interprofessional education facilitators' own interprofessional learning and collaborative practice. *Journal of Interprofessional Care* 30 (6): 702–709.
180 Reeves, S., Pelone, F., Hendry, J. et al. (2016). Using a meta-ethnographic approach to explore the nature of facilitation and teaching approaches employed in interprofessional education. *Medical Teacher* 38 (12): 1221–1228.
181 Kilminster, S., Hale, C., Lascelles, M. et al. (2004). Learning for real life: patient-focused interprofessional workshops offer added value. *Medical Education* 38 (7): 717–726.
182 Thistlethwaite, J.E. and Jackson, A. (2014). Conflict in practice-based settings: nature, conflict and resolution. *International Journal of Practice-based Learning in Health and Social Care* 2 (2): 2–13.
183 Thistlethwaite, J.E., Moran, M., Kumar, K. et al. (2015). An exploratory review of pre-qualification interprofessional education evaluations. *Journal of Interprofessional Care* 29: 292–297.
184 Interprofessional Curriculum Renewal Consortium (2013). *Curriculum Renewal for Interprofessional Education in Health*. Sydney:

Centre for Research in Learning and Change, University of Technology, Sydney.
185 Gilbert, J.H.V. (2005). Interprofessional learning and higher education structural barriers. *Journal of Interprofessional Care* 19 (Supl1): 87–106.
186 Lingard, L. (2012). Rethinking competence in the context of teamwork. In: *The Question of Competence* (ed. B.D. Hodges and L. Lingard), 42–69. Ithaca: ILR Press.
187 Kvarnström, S. (2008). Difficulties in collaboration: a critical incident study of interprofessional healthcare teamwork. *Journal of Interprofessional Care* 22: 191–203.
188 Orchard, C. (2015). Assessment of learning within interprofessional client-centered collaborative practice – challenges and solutions. In: *Interprofessional Client-Centered Collaborative Practice* (ed. C. Orchard and L. Bainbridge), 171–188. New York: Nova.
189 https://nexusipe.org/informing/resource-center?f[0]=im_field_subject%3A1290&f[1]=im_field_resource_type%3A1360 (accessed 23 April 2017).
190 Barr, H. (2015). Interprofessional Education. The Genesis of a Global Movement. Fareham, UK: CAIPE. https://www.caipe.org/resources/publications/barr-h-2015-interprofessional-education-genesis-global-movement (accessed 23 April 2017).
191 World Health Organization. High-Level Commission on Health Employment and Economic Growth. (2016). Health Employment and Economic Growth: A five-year action plan (2017–2021). Version for consultation. Geneva: WHO.
192 Takahashi, H.E., Watanabe, H., and Kinoshita, M. (2015). The Japan Association for Interprofessional Education. In: *Leadership and Collaboration: Further Developments in Interprofessional Education* (ed. D. Forman, M. Jones and J.E. Thistlethwaite), 47–67. Basingstoke: Palgrave.
193 Association of American Medical Colleges (2016). Curriculum inventory and reports. www.aamc.org/initiatives/cir/ (accessed 21 June 2018).
194 Thistlethwaite, J.E. and Brandt, B. (2016). The formation and development of the National Center for Interprofessional practice and education. In: *Leading Research and Evaluation in Interprofessional Education* (ed. D. Forman, M. Jones and J.E. Thistlethwaite), 23–39. Basingstoke: Palgrave.
195 Kusumawati, K. and Orbayinah, S. (2015). Interprofessional education from pilot to formal curriculum. In: *Leadership and Collaboration* (ed. D. Forman, M. Jones and J.E. Thistlethwaite), 26–46. Basingstoke: Palgrave.
196 Tong, S.-F., Mohamed, N., Tan, C.-E. et al. (2016). Transition from uniprofessional towards interprofessional education: the Malaysian experience of a pragmatic model. In: *Leading Research and Evaluation in Interprofessional Education* (ed. D. Forman, M. Jones and J.E. Thistlethwaite), 169–192. Basingstoke: Palgrave Macmillan.

Further Reading

Many of the other chapters in this book are valuable sources of additional reading to inform the development and delivery of IPE.

Good studies of IPE are dispersed across many journals, their location often reflecting the clinical setting for the IPE or the professional backgrounds of the authors. However, the most extensive single collection of papers about IPE – descriptions, evaluations, and theoretical debate – can be found in the *Journal of Interprofessional Care* (http://www.informa http://healthcare.com/jic).

15 Patient Involvement in Medical Education

John Spencer[1], *Judy McKimm*[2], *and Jools Symons*[3]
[1]Newcastle University, Newcastle, UK
[2]College of Medicine, Swansea University Medical School, Swansea, UK
[3]Faculty of Medicine and Health, Leeds University, Leeds, UK

KEY MESSAGES

- Patient involvement in medical education is embedded within a broader context of both policy and practice.
- There is a growing literature on patient involvement but much of it remains descriptive and uninformed by theory.
- A wide variety of initiatives and settings has been described.
- A wide range of benefits has been shown for all stakeholders.
- There is a need for appropriate support, training, and remuneration for all involved.
- 'Patients' include patients themselves, carers, patient groups, and patient representatives and advocates.
- There is a need for further research, including assessment of the strengths and weaknesses of different approaches, evaluation of long-term impact, and factors influencing sustainability.

Introduction

> My method (is to) lead my students by hand to the practice of medicine, taking them every day to see patients in the public hospital, that they may hear the patient's symptoms and see their physical findings. Then I question the students as to what they have noted in their patients and about their thoughts and perceptions regarding the causes of the illness and the principles of treatment.
>
> Sylvius (1614–1672) [1]

In the early seventeenth century Sylvius' teaching methods would have been unusual, indeed it would have been considered distinctly eccentric to involve patients in medical education to such a degree. Whilst the traditional physician apprenticeship – dating back (at least) to Hippocrates – relied on contact with sick people, by the time medical education was formally established in the universities of Europe in the thirteenth century, the patient had all but 'disappeared'. It was not until after the Renaissance that universities began to introduce bedside experience. And by the eighteenth century it was, in the words of one author, 'axiomatic' that students should supplement their book learning by spending time 'walking the wards' [2]. Clinical experience through patient contact gradually assumed its place at the centre of medical education such that by the turn of the twentieth century, Sir William Osler's assertion that 'it is a safe rule to have no teaching without a patient for a text, and the best teaching is that taught by the patient himself' [3] had become part of the rhetoric of a modern medical education.

Despite this evolving centrality, the patient's role in clinical education has historically been largely passive. At worst, a hapless hospital inmate unable to say 'No' to a gaggle of students at the foot of the bed; 'imposition' more than 'involvement'. Even at best, the patient, though treated with courtesy, was often no more than a medium, 'an interesting case', through which clinical teaching took place.

This chapter considers the importance of actively involving patients in the education of doctors and other health professionals, describes models of involvement, explores aspects of the growing literature on the subject, including evidence of benefits and disadvantages, discusses problems and challenges, and identifies areas for further inquiry. Before doing that, however, it is important that we first consider terminology.

Patients, Users, and Consumers

Nomenclature in this area is a potential source of controversy and confusion, and the issues are complex. The language used reflects values and power relations, generates strong emotions, and may impair both scholarly activity, such as searching the literature, and collaboration between interested groups [4, 5]. Views vary greatly about how people prefer to be described, and preferences and language change over time [5]. For simplicity's sake we will mainly use the term 'patient' (and, where relevant, 'carer') throughout this chapter, because, for all its limitations, it is probably

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

BOX 15.1 FOCUS ON: Terminology

'User' or 'service user' is commonly used in the UK, but in North America may be more associated with illicit drug use. Furthermore, the term implies that medical care is simply a technical service rather than a more holistic, relationship-based interaction. However, 'user' does imply a degree of active participation.

'Lay' is also used frequently; it acknowledges that the person may not necessarily be either sick or under active care. However, 'lay' defines someone essentially not by any positive attributes, but by what they are not (i.e. not 'professional') and what they do not have (i.e. they don't have medical expertise).

'Consumer', 'client', or 'customer' connote health as a commodity and health care as a market, and thus suggest a commercial relationship.

'Survivor' and 'person in recovery' are terms mostly restricted to use in relation to cancer and mental health.

'Patient' is probably the most unambiguous term, although it implies that the person is sick and under active care; the term connotes passivity; and the care relationship is medicalised.

'Standardised patients' were originally real patients trained to present standardised representations of their own problems; the term often overlaps with 'simulated patient'.

the most widely recognised term in *medical* education – although terminology differs amongst user groups as well as different professionals and countries (see Box 15.1). We use 'patient' to mean both people with health problems, whether or not they are currently receiving care, and healthy people, although we recognise that this may be contentious and will not acknowledge everyone's preference. We also recognise that 'on the ground' it is important to be mindful of the power of language [4]. The problem of terminology also extends to describing the active roles increasingly played by patients. A wide range of terms are used, including instructor, educator, expert, associate, and mentor.

Context of Patient Involvement

Engaging patients and the public in health care can cover many aspects. For example, Carman et al. [6] described a model of 'patient and family engagement' with three critical dimensions: the continuum of forms of engagement (from consultation to partnership and shared leadership), the different levels at which engagement may occur (ranging from direct care to policy making), and the factors influencing whether and to what extent engagement occurs.

Patient involvement in the development, delivery, and management of health care has been enshrined in health policy worldwide for several decades. Taking the UK as an example, successive waves of health reform have aimed to ensure that patient and public involvement should be part of everyday practice in the NHS [7, 8], reflected in the catchphrase 'no decision about me, without me'. This principle is now enshrined in the NHS Constitution which is absolutely clear that the NHS (in England) 'should support individuals to promote and manage their own health but also that NHS services must reflect, and should be coordinated around and tailored to, the needs and preferences of patients, their families and their carers' [9, p. 3]. This development has inevitably encompassed the education and training of health professionals [10] and the latest standards for education and training from the General Medical Council (GMC) includes the requirement that 'the development of medical school curricula must be informed by … patients, families and carers' [11]. The GMC has also produced 'supplementary guidance' on the involvement of patients and the public [12]. These trends are not confined to medical education nor to the UK [13].

Aside from national policy agendas, there is an obligation for medical schools to demonstrate 'social accountability' which involves schools directing 'their education, research and service activities towards addressing the priority health concerns of the community, region and/or nation they have the mandate to serve' [14]. Social accountability has evolved as a major contemporary discourse, embracing concepts such as 'the social contract' (between professions and society), 'social responsibility' (of doctors and other health professionals), and 'social responsiveness' (of institutions). Woollard and Boelen [15] highlight the challenge for medical schools 'to strive for and demonstrate greater impact on health through their bonds with society', which, they contend, is the very purpose of social accountability. They argue that medical schools must demonstrate a commitment to social accountability in both formal programmes and the 'hidden curriculum'. At the same time, and influenced by similar social and political forces, the concept of 'professionalism' has been revisited and redefined, with obvious implications for education [16–18]. While debate continues about the nature of professionalism, its theoretical basis, and how best to teach, assess, and research it, there is general agreement that professionalism is underpinned by a commitment to patients' interests and must be 'based on virtue, deeper attitudes rather than mere behaviour, and requiring of practical wisdom' [19].

Aside from changes in expectations, there is now a greater appreciation of the psychological and social consequences of ill health and health care treatments, and the need for new models to guide practice, such as the 'biopsychosocial model' and 'patient-centredness', which put the patient's perspective and priorities at the centre [20, 21]. Increased knowledge about variation between patients in values, preferences, and responses to illness and their effect on treatment outcomes has brought the health care professional's interaction with the patient to centre stage. And whilst the technical complexity of medicine offers greater choice of diagnostic tests and treatments and more complex interventions, financial constraints often lead to doctors having to navigate their way through the tensions between expectations and feasible options.

These developments have occurred in parallel with changes in public expectations, changes often described as part of the move from paternalism to partnership. With the decline in uncritical deference to the professions, the rise of consumerism, and a greater understanding of what health care can and cannot achieve, many patients expect to have their concerns addressed and their requests heard, and to be fully informed about their condition, briefed about risks of treatment, involved in decisions about their care, and educated and supported to manage their own conditions [22].

Co-production is a model of service delivery developed over the past 30–40 years which emphasises the central importance of the contribution of service users [23]. Collaborative co-production involves a relocation of power whereby the user is seen as an expert in their own circumstances, and professionals move from 'being fixers to facilitators' [23]. It requires a new relationship with professionals, and all parties need training to take on new roles. Over the same period, shared decision-making has evolved as a new model for the clinical encounter. The underlying ethical principles require acceptance that self-determination is desirable and that the clinician's role is to support this. Shared decision-making – for which there is a growing evidence base to support effectiveness [24] – depends on building an effective relationship in the clinical encounter using a range of specific communication skills.

Such changes make working with patients ever more demanding – for example, supporting patients to exercise choice in situations of uncertainty, enabling them to understand the options available and the risks and dangers involved, and helping them appreciate restrictions on choice. Learning how to do all these things needs input from patients and challenges educators to seek the most appropriate ways of enabling students and trainees to learn, whilst respecting the rights and needs of patients [25].

Active involvement of patients and carers in health care professional education was a logical development from involvement in other areas, including policy and research. It started in medicine, nursing, social care, and education of (non-physician) mental health professionals but has spread to other disciplines such as occupational therapy, pharmacy, and physical therapy. Nonetheless at the time of writing, to quote a recent position paper, it is 'still not well established in the mainstream of educational practice', and is 'often limited to a specific population of patients … is fragmented and not embedded in the educational institution, and lacks appropriate infrastructure and sustained leadership and resources' [26, p. 19].

Scope of Patient Involvement

Many people might assume that 'patient involvement in the curriculum' was limited to direct involvement in teaching, learning, and assessment; indeed, these areas are the main focus of this chapter. Nonetheless, there is potential for people to make a contribution to *all* aspects of the educational process. These include the following:

- student selection and admission
- curriculum development
- course management
- faculty development
- practice placements
- programme evaluation.

However, several recent major literature reviews highlight that patient involvement in most of the above areas is still relatively unusual. A number of frameworks help us explore the potential scope of patient involvement, and here we describe three: Tew et al.'s [27] 'ladder of patient involvement', the 'Cambridge framework' [28], and a taxonomy of active involvement described by Towle et al. [5].

Ladder of Patient Involvement

Tew et al. [27] describe a 'ladder of involvement', which they propose can be used to establish and monitor patients' involvement within individual programmes and institutions. The tool was developed in the context of non-physician mental health education and training, but could be applied both across the educational spectrum and across disciplines (see Box 15.2).

The Cambridge Framework

Spencer et al. [28] reviewed the patient's role in medical education and suggested a framework (the 'Cambridge framework') to facilitate discussion about patient involvement. It is based on four sets of attributes of contexts in which patients, students, and teachers interact, under the headings 'Who?', 'How?', 'Where', and 'What?', providing a template against which patient involvement can be planned or evaluated

Who?

This reflects the individual background, culture, experience, and expectations of each patient, their family, and carers. Patients vary immensely in terms of the clinical problems with which they present, as well as their age, gender, ethnicity, sexual orientation, emotional and intellectual capacity, and socio-economic status.

How?

Students and trainees work in a wide range of settings (such as hospital wards, hospital and community ambulatory clinics, emergency departments), which present different educational opportunities. These depend on factors such as whether encounters are planned or opportunistic, pressures of time, available supervision, and so on. Considering these issues may help teachers plan how patients may be involved based on the intended learning outcomes and the setting.

Where?

Recognising that health care takes place in a wide range of locations and settings, and that context will inevitably influence the nature and quality of learning, questions addressed under this heading explore issues to do with place, safety, identity, and power relationships. The 'Where?' also includes whether it is a 'real' or 'simulated' environment, such as a training ward, and the contrast between 'uniprofessional' or 'multiprofessional' settings to distinguish between situations in which doctors alone are learning with

BOX 15.2 Ladder of patient involvement [27]

Level	Description of involvement
1 No involvement	The curriculum is planned, delivered, and managed with no consultation or involvement of service users or carers.
2 Limited involvement	Outreach with local service user or carer groups. Service users/carers invited to 'tell their story' in a designated slot and/or be consulted about course planning or management, student selection, student assessment, or programme evaluation. Payment offered but no opportunity to participate in shaping the course as a whole.
3 Growing involvement	Service users/carers contribute regularly to at least two of the following: planning, delivery, student selection, assessment, management, or evaluation. Payment at normal visiting lecturer rates. However, key decisions on matters such as curriculum content, learning outcomes, or student selection made in forums in which service users/carers are not represented. Some support before and after sessions, but no consistent programme of training and supervision. No discrimination against service users and carers accessing programmes as students.
4 Collaboration	Service users/carers involved as full team members in at least three of the following: planning, delivery, student selection, assessment, management, or evaluation. Underpinned by a statement of values. Service users/carers contribute to key decisions on matters such as curriculum content. Facility for contributors to the programme to meet and regular provision of training, supervision, and support. Positive steps to encourage service users and carers to access programmes as students.
5 Partnership	Service users, carers, and staff work together systematically and strategically across all areas, underpinned by an explicit statement of partnership values. All key decisions made jointly. Service users and carers involved in the assessment of practice learning. Adequately funded infrastructure to provide induction, support, and training. Service users and carers employed as lecturers on secure contracts and/or contracts established between programmes and independent groups. Positive steps made to encourage service users and carers to join learning sessions, even if not (yet) in a position to achieve qualifications.

patients and those in which a range of health and social care professionals are learning and working.

What?

This set of attributes deals with 'content': the clinical problems presented, the specific skills and knowledge that may be learned, and underlying attitudes and values. Consideration of these should help teachers realise the maximum potential of particular situations and assess the likely impact on both patients and learners.

Taxonomy of Active Patient Involvement

Towle et al. [5] propose a taxonomy combining elements of both the 'Cambridge Framework' and 'ladder of involvement', which they argue helps both clarify the patient's role and makes communication of research findings easier to articulate, synthesise, and compare. Their classification considers the wide variety of ways other than real patients in the workplace in which patients may be encountered, describing a continuum of involvement grounded in five attributes at six levels (see Box 15.3).

Levels of Patient Involvement

In this section, we describe some of the issues that may arise for medical educators in each of the categories of involvement described in Box 15.3.

Cases and Scenarios

Paper-based cases have long been used to supplement real patient contact, with video-based and electronic cases increasing in use as the technology has developed. The use of virtual patients (VPs) was first described in the early 1970s. A VP has been defined as a: 'specific type of computer program that simulates real-life clinical scenarios: learners emulate the roles of health care providers to obtain a history, conduct a physical examination, and make diagnostic and therapeutic decisions' [29]. Virtual patients can be static or dynamic, used passively or interactively, and may be linear or branching [30].

Whilst some reviews concluded that the evidence base to inform the very wide variety of uses of VPs was very weak, a more recent discussion paper contended that much progress had been made in the thoughtful application of VPs, to the point where their use has become embedded in curricula, including areas such as PBL cases, interactive lectures and seminars, and both formative and summative assessments [31]. Drivers of these developments include international collaboration, decreasing costs, better authoring systems, and greater ease of dissemination. Although not intended to replace authentic patient contact, the use of VPs appears to be able to usefully complement clinical experience – in particular, from both theoretical and empirical perspectives, in the development of clinical reasoning, requiring, as it does, exposure to multiple cases and variations [31, 32]. However, there has been relatively little

BOX 15.3 Taxonomy of patient involvement [5]

Degree to which the patient is actively involved in the learning encounter. *N.B. In all instances patients are 'real' and assumed to be representing themselves.*	Duration of contact with learner	Patient autonomy during the encounter	Training for the patient	Patient involvement in planning the encounter and curriculum	Institutional commitment to patient involvement in education
1 Paper-based or electronic case or scenario: Patient is focus of a paper-based, electronic, or web-based case or scenario.	None	Not applicable	Not applicable	None	Low
2 Standardised/volunteer patient in clinical setting: Patient encounter with student is scripted and serves as an example to illustrate or reinforce learning, e.g. teacher asks patient to provide student with history or student practises a clinical exam.	Encounter-based	None	None	None	Low
3 Patients share their experience with students within faculty-directed curriculum: Patient is invited to share experience; faculty plan the encounter but patient determines personal comfort and level of participation.	Encounter-based	None to low	Brief, simple	None	Low
4 Patient teacher(s) involved in teaching and/or evaluating students: Patient is given preparation for specific teaching role, may actively question students, and may be involved in giving feedback and evaluating their performance.	Variable	Moderate	Structured, extensive	Low to moderate	Low to moderate
5 Patient teacher(s) as equal partners in student education, evaluation, and curriculum development: Patients are involved in many aspects of educational delivery, development, and evaluation, beyond specific courses to the curriculum as a whole – a true partnership in which patients make meaningful and valued contributions to decision-making.	Moderate to extensive	High	Extensive	Moderate to extensive	Moderate
6 Patient(s) involved at the institutional level in addition to sustained involvement as patient teacher(s) in education, evaluation, and curriculum development for students: As [5] above but there also are institutional policies that ensure involvement in decision-making bodies within undergraduate, graduate, and continuing health professional education.	Extensive	High	Extensive	High	High

high-quality research into, for example, the most effective design, sequencing, or balance within the curriculum. Consideration of the evidence-based principles that underpin effective instructional design in the related field of simulation-based education may be helpful. These include range of difficulty, repetitive and distributed practice, cognitive interactivity, multiple learning strategies, individualised learning, mastery, and feedback [33, 34]. See Chapter 11.

However, fast-evolving technology has the potential to build VPs with far greater realism. As Poulton and Balasubramaniam note: 'It is now possible to consider the extension of the current relatively lightweight VP into a truly interactive patient simulation, an "e-human" or "digital avatar" … offering authentic patient management, clinical and communication skills training, and the potential capability to mimic the health or disease of any citizen' [31]. This development will need to draw on new collaborations between medical educators, learning technologists and content experts, and, crucially, patients and carers. However, a note of caution was raised in a systematic review of the literature about the utility of VPs: they are expensive to develop, which begs the need for a virtual commons or online community where resources and ideas can be shared [30].

On a more holistic level, there have been increasing calls to adopt a narrative approach in all aspects of health care, clinical practice, research, *and* education, with potential benefits claimed for all parties [35, 36]. As the widely quoted Barbara Hardy noted in 1968 'we dream in narrative, daydream in narrative, remember, anticipate, hope, despair, believe, doubt, plan, revise, criticise, construct, gossip, learn, hate and love by narrative' [37, p. 5].

One kind of narrative account is what Aronson [38] called the 'autopathography', more simply referred to as a 'patient's tale', or perhaps more cynically, the 'medical confessional'. He analysed and classified characteristics of nearly 300 such book-length tales. Reading and reflecting on such stories may help health professionals, both in training and in practice, better understand and empathise with their patients, and 'teach them things they won't learn from textbooks', indeed 'that cannot be arrived at by any other means' [35].

There are many ways in which such resources may be used educationally. Powley and Higson [39] suggest the following simple process for using written narratives in teaching: read, discuss, facilitate, analyse responses, and discuss applications. Questions such as 'What is the story about?' and 'What effect did it have on me?' promote reflection and help learners focus on key messages and apply in new contexts. Grounding the exercise in reality reinforces relevance, which in turn helps motivate learners. Sufficient time must be allowed for reflection and discussion.

In terms of choice of text, Aronson's own criteria for recommending a book are 'that it should provide a judicious balance between emotional expression and analytical discourse, and that it should have informed, and above all entertained me' [38]. He suggested a 'top ten' books in his article, but many more have been and continue to be published since he wrote.

Perhaps a more commonly used source of such stories nowadays is the Internet. 'Google' any disease, common or rare, and a significant proportion of 'hits' will be personal blogs about the problem, as either patient or carer. Online resources such as 'Patient Voices' or 'Healthtalk' (see Box 15.4) are readily available. Others can be purchased or developed to suit specific purposes, although the cost, time, and expertise required to do this should not be underestimated. The literature is sparse in this area, but involving patients and carers in designing e-learning materials is self-evidently important to ensure that patient experiences are more comprehensively and accurately represented [40].

Standardised Patients

Learning from real patients in clinical settings is central to medical education, enabling learners to consolidate and synthesise learning from a range of sources. Indeed, 'bedside teaching' is the only setting in which all of the technical

BOX 15.4 Case study: Online resources for patient narratives

- *Healthtalk* (http://www.healthtalk.org) was established in the late 1990s as DIPEX, the 'Database of Individual Patient Experiences', later becoming Healthalkonline, now Healthtalk. The material is based on qualitative research into patient experiences of a wide (and expanding) range of illnesses and conditions, led by researchers in the 'Health Experiences Research Group' at the University of Oxford. At the time of writing there are over 2500 video and audio clips covering over 100 health conditions with a section on how to use the material in education. There is also a section dedicated to young persons' experiences at http://www.healthtalk.org/young-peoples-experiences
- *Patient Voices* (www.patientvoices.org.uk) was founded in 2003 and 'aims to facilitate the telling and the hearing of some of the unwritten and unspoken stories of ordinary people so that those who devise and implement strategy in health and social care, as well as the professionals and clinicians directly involved in care, may carry out their duties in a more informed and compassionate manner.' The stories 'of health and illness, tragic loss and miraculous recoveries' are usually gathered during workshops, and use video, audio, still images and music to convey each unique account. It provides more than 1000 digital stories from patients, carers, and health workers on a range of topics.
- Another category of website with potential for use in education and training has burgeoned recently in the UK, namely sites soliciting patient feedback, one of the best known being *Patient Opinion* (www.patientopinion.org.uk).

Such resources can be included in face-to-face or e-learning tutorials, or lectures or workshops (e.g. embedded in PowerPoint presentations), providing illustrative or trigger scenarios about different clinical conditions or critical incidents including promoting discussion about responding to complaints and preventing error. These are particularly helpful when it is inappropriate or difficult for learners to work with real patients – for example, those who have rare conditions, are terminally ill, or have mental health problems.

and non-technical skills, behaviours, and applied knowledge that constitute 'doctoring' are modelled by clinical teachers and can be learnt as an integrated whole.

Ramani reminds us that 'the bedside is the perfect venue for unrehearsed and unexpected triangular interactions between teacher, trainees, and patient … physician teachers should be vigilant about grabbing teachable moments' [41] during these encounters.

'Standardised' (or 'programmed') patients were first introduced by Barrows and Abrahamson [42] in the 1960s. Although originally real patients trained to represent their problem(s) in a consistent manner for the purposes of teaching and assessment, nowadays, in the words of Barrow himself, the 'patient' will generally be 'a well person, with or without a thespian background, trained to simulate a patient's illness' [43] – see the section later in this chapter.

Whenever clinical teaching occurs, patients are usually the most passive and vulnerable of the parties involved. Notwithstanding this, most patients find involvement rewarding and are willing participants, often commenting that they recognise that students 'have to learn'. However, it should not be tacitly assumed that patients will engage in teaching; their wishes and feelings should always be respected, and they should know that, whatever their decision, their treatment and care will not be affected. Patients must always be informed that learners may be present and may be providing care, whatever the setting. This allows them to prepare for the initial encounter and to raise anxieties (see the section on 'Ethical Issues' later in this chapter). There are also benefits from briefing patients explicitly about a session's aims, what teacher and learners hope to get out of it, and expectations of the patient, for example, whether they will be asked to give feedback. Patients need to be aware of the number and level of the learners who may be present, each person's role should be clarified, and verbal or written agreement obtained and recorded as appropriate. McKimm's [44] clinical 'trialogue' – a 'discussion or conversation in which three persons or groups participate' [45] that attends to the developing relationships between all three 'players', rather than consciously trying to think about teaching and clinical practice – is a useful model that may help teachers plan and work actively with patients and learners.

Patients Share their Experience *and* Patient Teachers Involved in Teaching and/or Evaluation

Selecting real patients for teaching is often opportunistic, but a more structured approach employing trained patients is being increasingly used within undergraduate and postgraduate training. Such patient educators can be drawn from many settings, including areas where concerns might be expressed about potential harm to patients from ad-hoc teaching encounters, such as those who are terminally ill or have mental health problems.

The concept of the expert patient is enshrined in the wider patient involvement agenda. In the UK, for example, the 'expert patient initiative was … part of the government's commitment to place patients at the heart of health care which is … part of the transformational focus of the clinical governance agenda' [46]. This was primarily targeted at people with long-term conditions to help them 'become key decision makers in their own care' [46]. It was supported by an educational programme for the patients themselves, and there is some evidence that it resulted in changes in confidence and self-efficacy in relation to self-management of symptoms (such as pain, tiredness, and depression) [47]. However, results of several randomised controlled trials looking at outcomes of such programmes failed to show any effect on use of health care [48]. Expert patients may need support and encouragement, ideally from other patients through sharing individual patient journeys, to prepare them for a more active role in the education of health professionals [49].

There are advantages in working with patients who are not under current active care, particularly when teaching inexperienced learners when more time needs to be taken than clinical demands allow. Patient educators have the benefits of being:

- motivated individuals with an interest in medical training
- 'real' (not simulated) with authentic clinical histories and possibly clinical signs
- able to give structured feedback from a patient's perspective, such as the pressure of the hands or the way in which a history is taken.

Patient instructors (PIs) with rheumatological conditions have been well described, with most studies demonstrating high levels of satisfaction from both learners and PIs, and that PIs can be as effective as clinicians in enhancing learners' knowledge and skills, especially in talking about the impact of living with the condition [50]. Another distinctive category of patient teacher is the 'gynaecology teaching associate', women trained to teach pelvic and breast examination through examination of themselves. They are widely used in North America and Scandinavia, and increasingly in other Western countries, although there is still much debate about students performing intimate examinations [51]. Such teaching can help free up clinical tutors as, once trained, the associates need little assistance in running sessions. They have been found to be acceptable and effective – in one comparative study students taught by associates had better skills in an end-of-attachment assessment than students who were not [52].

Certain groups have been under-represented in professional education programmes, including people with learning disabilities, those whose first language is not the majority language, and the terminally ill, as well as traditionally 'hard-to-reach' communities such as the homeless, asylum seekers and refugees, and people with substance misuse problems. Actively involving such patients presents particular challenges but successful interventions have been described, for example in one of the authors' institutions refugees came into the medical school to tell their stories and answer questions about their experience; the sessions were invariably positively evaluated by all parties.

Carers represent another large and important population (one in eight people are said to be carers in the UK, for example, see: https://www.carersuk.org/news-and-campaigns/press-releases/facts-and-figures), whose own health and emotional needs are often overlooked, and who potentially have a great deal to offer. Innovative sessions have been described which put the carer's perspective at the centre, for example in learning about dementia [53].

Well-managed patient involvement can benefit both patients and learners. The challenge is to be sensitive to both patients' and learners' needs and to identify suitable patients who feel equipped to participate. See Box 15.5 and Box 15.6.

Patient Teachers as Equal Partners

It is unusual in medical education to find this level of involvement, with some notable exceptions, but models have been developed in other disciplines, particularly (non-psychiatric) mental health, nursing, and social care education. Guidelines and recommendations derived through a variety of processes have been published [55–57]. and UK guidance about involving patients in research highlights issues that may be relevant to professional education [58]. Both the GMC and British Medical Association have published advice about active patient involvement in medical education and training [12, 59].

Patients Involved at the Institutional Level

Involvement at this level is even rarer than at the previous one. One of the best examples in the UK is the Universities/Users Teaching and Research Action Partnership (UNTRAP), based at Warwick University [60], and which is a partnership between Warwick and Coventry universities, users of health and social care services and their carers, and the NHS. Patients are involved at different levels, some in one-off events, and others more heavily. The central philosophy of UNTRAP is that everyone will benefit if service users, carers, academics, and professionals share their experience. Patients and carers helped strategically develop Warwick Medical School's case-based curriculum launched in 2013 and are active throughout the curriculum. UNTRAP have also developed the first accredited training for partnership working [61]. Other institutional approaches have also been described [27].

BOX 15.5 Case study: Interprofessional health mentors programme UBC [54]

In a programme at the University of British Columbia (UBC), students from different disciplines were attached to a person in the community with a long-term condition for a period of three semesters.

During six themed meetings, the health mentors ('experts-by-experience' recruited through community groups) worked with students to help them learn about living with a long-term condition from the patient's perspective, and to develop interprofessional competencies. Participants in these 'self-managed learning communities' were encouraged to explore their own questions and to be creative in their approach to working towards the goals. A symposium two-thirds of the way through the programme enabled sharing of ideas and reflection on progress.

The programme was rated very highly and as beneficial by mentors and students alike. A wide range of learning outcomes were documented by students, for example recognising the benefits of collaboration and the expertise and resourcefulness of patients. Mentors benefited from being able to describe their 'complete journey', and the opportunity to 'give back'. All parties acknowledged the importance of long-term relationships.

Four factors were identified as key to success:

- the uniqueness of patient-centred learning, with faculty acting as facilitators
- sustained partnerships with community organisations
- keeping things simple
- encouraging diversity, creativity and flexibility.

BOX 15.6 FOCUS ON: Body donation

Probably the ultimate contribution a person can make to the education of doctors and other health professionals is body donation. Cadaveric dissection was central to anatomy education for several centuries, but whether students need to have hands-on experience of dissection or even be exposed to human cadavers at all has been subject to recent debate. Proponents argue that it provides a three-dimensional perspective, insights into variation, develops manual dexterity, basic surgical skills, and aspects of professionalism such as team working, and promotes humanistic values and respect for the dead. Critics highlight the potentially de-humanising effect of exposure to cadavers, arguing that it is an outdated 'rite of passage', and that alternative teaching methods using modern technology, imaging, interactive multi-media, plastinated models, body painting, and life drawing, have the potential to promote learning in a more engaging way [62, 63]. Use of cadavers is subject to various legal and ethical regulations and considerable variation between countries exists reflecting cultural and religious differences (e.g. concerning the acceptability of using unclaimed, unconsented bodies). This has resulted in calls for the adoption of a common framework which includes informed consent, liaison with families, services of commemoration, and discouraging commercialisation [64].

Relatively little is known about the motivation of people to donate their bodies, other than it is a complex issue. A survey of donors in the Netherlands identified three main dimensions: a desire to be useful after death, negative attitudes to funerals (e.g. expense or burden placed on families), and an expression of gratitude to 'the system' [65]. In a similar survey in New Zealand, South Africa, and Eire, 80% of respondents expressed a desire to help medical science [66]. A contemporaneous community survey in India, however, identified lack of awareness of and considerable negativity towards body donation, partly related to religious beliefs and customs, along with concerns about whether bodies would be treated with respect [67].

Benefits and Disadvantages

The thought of training doctors *without* direct patient contact would nowadays be considered absurd. Medical schools aspire to maximise it, teachers and administrators strive to deliver it, students demand as much as possible, and patients seem only too willing to help. At the postgraduate level, with increasing emphasis on in-service training grounded in clinical practice, patient contact is obviously crucial. A significant literature has accumulated – one of the most comprehensive bibliographies identified over 400 relevant papers [68] – providing corroboration for some of the theoretical benefits. Contact between patients and learners is generally very well received, with relatively few apparent adverse effects or disadvantages for the former, and even fewer for the latter. Patients recognise their contribution – for example, by acting as 'experts' in and/or exemplars of their condition, showing and telling, aiding the development of professional skills and attitudes, and boosting learners' confidence [69]. Whilst most studies report distinctive, largely positive, outcomes for patients, concerns exist about the possible emotional and psychological impact on both patients *and* students of recounting painful and traumatic experiences, and the 'professionalisation' of some patients through repeated telling of their stories [70]. Another study using a phenomenological approach, showed that involvement in clinical teaching for most patients was, in fact, often characterised by its ordinariness [71].

Most of the research informing this literature has limitations in that it is descriptive, is not informed by theory (see below), is often based on self-report, and provides insufficient information about educational interventions or research design. There have been few attempts at evaluating long-term impacts and publications are found in disparate outlets, which, along with problems posed by use of different terms, has bedevilled searching and synthesis of the literature [5]. Nevertheless, several comprehensive reviews have been published over the past few years showing, on the whole, consistent findings about benefits and disadvantages [4, 5, 10, 27, 72–76]. See Box 15.7.

BOX 15.7 WHERE'S THE EVIDENCE: Patient involvement

Benefits to students

- motivation through relevance
- increased empathy
- development of professional skills and attitudes
- increased confidence
- social responsibility
- development of clinical reasoning
- new insights and understanding
- recognition of cultural diversity and lifestyle factors
- improved performance in examinations

Benefits for patients

- satisfaction at contributing to student learning
- improved relationships with professionals
- altruistic feelings, for example, giving something back to 'the system'
- being valued and increased self-esteem
- development of own skills
- catharsis
- increased knowledge about their own condition
- getting a better service from their clinicians, for example, 'a good going over'
- companionship and relief from social isolation
- the long-term relationship and continuity established in longitudinal programmes

Disadvantages to students

Few disadvantages are reported by students, but they include:

- embarrassment
- emotionally challenged in certain situations
- being a burden to patients
- concern about representativeness of some patients

Disadvantages for patients

These are mainly in relation to:

- mental health problems and potentially embarrassing situations (e.g. intimate examination)
- concerns about confidentiality and choice, previous poor experiences, large numbers of learners

There is little evidence of significant adverse effects on the health and well-being of either students or patients.

Theoretical Considerations

To date, change and innovation has been pragmatic, driven largely by social developments and/or in response to policy. Neither practice nor research appears to have been much informed by theory, with notable exceptions. Rees et al.'s [4] study used the lens of 'situated learning' to explore how medical students learn 'with' rather than just 'about' service users. They posited that both patients and students were 'legitimate peripheral participants' struggling in parallel with the challenges of power imbalance, identity, and roles, as they moved towards greater participation. The authors offered a set of recommendations to encourage more active participation. Monrouxe and colleagues adopted Goffmann's dramaturgy theory to explore the many roles played by participants in hospital bedside teaching, including 'actor', 'director', 'audience', 'non-person', and 'prop' [77]. Bleakley and Bligh [78] used contemporary post-structuralist theory to explore the concept of patients and learners engaging in collaborative knowledge production. They argued for a 'radical overhaul of conventional doctor-led education … that also challenges the orthodoxies of individualistic student-centred approaches' which could lead to development of an educational model whereby the locus of learning shifts from the relationship between doctor as educator and student to the relationship between patient and student, with the doctor as a resource and facilitator. Regan de Bere and Nunn [79] demonstrated how 'activity theory' provides a framework for understanding the complexities of patient and public involvement, allowing, as it does, consideration of context, change, challenge, and conflict and providing potentially rich descriptions of the activity under review. Both practice and scholarship would benefit from more theoretical explorations of this nature.

Principles and Practice

In this section we explore general principles for active patient involvement, focus on three important ethical issues, and consider the challenge of representativeness.

General Principles

As described above, a number of reports supporting and guiding patient involvement have been published, and these invariably also highlight barriers and challenges. From this growing literature [4, 12, 55, 58, 59, 74, 76, 80] we have listed practical issues to consider when establishing a new initiative (see Box 15.8). We have not explored barriers and challenges in any detail, but these are argued to include the following:

- different, sometimes conflicting, values and expectations
- power imbalances
- perceptions of intimidation (such as unfamiliar even hostile environments)
- lack of consent, choice, and confidentiality
- time constraints
- institutional inertia
- inadequate resourcing
- lack of training, support, and debriefing
- problems with language and communication.

The practical design considerations in Box 15.8 should be underpinned by a set of principles that include ensuring institutional support, providing adequate resources, and providing safe and comfortable environments. Patients should be treated as equals and attention paid to language (e.g. 'working with' rather than 'using'), and avoiding technical jargon and terminology. Educators should be prepared to be flexible, should embrace change, and be prepared to be challenged and think in new ways. Involvement should be seen as a process, not a one-off, and opportunities sought for further development and capacity building.

BOX 15.8 HOW TO: Develop a new educational initiative involving patients

1. Assemble a team, ensuring patients are involved from the start (not as an afterthought, or worse still, imposition); learners can also make an important contribution; a named, dedicated lead is crucial (ideally not an add-on to someone's already busy role).
2. Invest time and effort in building relationships and developing ways in which patients and/or the community may benefit from involvement.
3. Set up a steering committee with clear terms of reference and some authority and broad membership to encourage ownership.
4. Involve patients in designing and developing as well as delivering the educational programme, and encourage collaborative learning.
5. Develop a recruitment strategy, including a selection process involving, as appropriate, patient support groups and other local networks.
6. Provide orientation and training appropriate to specific roles and ensure an on-going support system is in place, including mechanisms for responding to feedback.
7. Consider practicalities such as timing (including meetings as well as educational sessions) and location (including accessibility), as well as intellectual property rights.
8. Establish a clear policy on remuneration being mindful of its impact, for example on welfare benefits.
9. Explore ways of explicitly recognising involvement (other than remuneration), such as job titles, certificates of participation, recognition events, access to academic facilities (e.g. library).
10. Build evaluation into the system from the beginning, incorporating the perspectives of all stakeholders.
11. Work to ensure diversity is reflected and minority views are represented.
12. Underpin any strategy with a statement of values (this might include issues such as gender, ethnicity, religion, and sexuality, as well as expectations).

Ideally, all the above should be developed with patients and carers from the start, not through post-hoc consultation or, worse still, through imposition.

Ethical Issues

Ethical issues to be considered when involving patients can be summarised as the 'three Cs': consent, choice, and confidentiality. The main message emerging from policy documents, good practice, and the literature is that simply assuming that patients will be involved in teaching and learning without making this explicit through formal systems, professional conversations, and ethical practice is no longer acceptable.

Consent

Medical law and ethics enshrine the principle of informed consent, which should routinely guide patient involvement, not just those encounters involving intimate examinations or invasive procedures [59]. 'Arguments for not informing patients in advance seem to be based more on prejudice than on empirical evidence' [81], and providing information about learner involvement before the clinical encounter does not appear to adversely influence patients' decisions about participation. Perhaps unsurprisingly, there is evidence that patients are more positive about involvement when consent *is* obtained [82]. Obtaining consent should be 'a continuous process that begins with the first contact the service has with the patient' [59], and all patients should be informed that students may be present and, as appropriate, involved in care. It is important to recognise that the presence of a learner will inevitably change the dynamics of a consultation, although there is little, if any, evidence of significant negative effects on quality of care [83]. An interesting argument has been made that there should be an expectation that all patients should be willing to contribute to medical education as a default stance, the quid pro quo being that all learners should be required to demonstrate competence in the relevant task in simulation before being 'let loose' [84].

Choice

Facilitating patient choice is challenging when students and trainees need to learn within the 'turbulent here and now of care delivery' [84], with little time to ensure that each encounter is set up optimally for all. Promoting active choice shows basic respect but also acknowledges that the patient is an expert about the way their own condition affects them. It moves clinicians away from 'operating from within the safety of a powerful expert role and performing habitual and ritualized tasks that depersonalize the transaction of caring' [85].

Seeking informed consent about teaching should ideally be done without the learner in attendance, then confirmed in their presence [59]. Building in 'moment-to-moment' opportunities for patients to say 'No' is another way of empowering them and acknowledging their needs [28]. Patients should be informed about the level of experience and identity of any learner intending to carry out a procedure on them. Lack of personal power and space, and the more urgent need for treatment, mean that a different approach may need to be taken in hospital settings compared with primary care or in the community, where there is usually a more intimate relationship, more privacy, and patients have greater autonomy [86].

Confidentiality

Confidentiality in relation to patients involved in education must be maintained. Some patients express concerns 'about students' access to their case notes and whether discussions about patients occurred after they had left the consulting room' [87], which raises questions about how aspects of choice, consent, and confidentiality should be raised with patients.

Practical steps include:

- providing sufficient information so people can understand the boundaries of confidentiality
- reassuring patients that learners are bound by the same duty to respect confidentiality as are 'fully fledged' health professionals
- involving patients in discussions
- finding private spaces to discuss intimate or distressing issues, remembering that curtains around a bed or cubicle are not soundproof (!)
- raising issues of confidentiality routinely with learners as part of preparation and debriefing.

When patient information is being used in teaching, permission must be obtained for the use of images, sound recordings, and extracts from case notes, particularly identifiable information. Increasing use of electronic records and mobile communications and technologies is creating new challenges.

Medical ethics and law is complex and ever-changing, and all clinicians have a responsibility to keep up-to-date and informed. Clinical teachers are key role models; keeping the 'three Cs' – consent, choice, and confidentiality – in mind ensures these are seen as fundamental pillars of good practice, not as options. Embedding these principles in institutional practices and policies is an important step.

Patient Representation

It is easy to forget that the 'patients' are not a homogeneous group; indeed, diversity is the norm. 'Patients' do not think alike any more than professionals do, yet much of the literature on 'involvement' seems to treat all users, carers, survivors, clients, patients, etc. as the same (even if only implicitly). Charlotte Williamson [88] of Picker Institute, Europe, proposes the following three broad categories of 'patient' who might get involved.

- *Individual patients* who can describe their own experience but cannot necessarily speak for others, with implications for the generalisability of their experience.
- *Patient group members* who usually *do* know about the experiences of others like themselves, but may still have a narrow perspective. Consulting all relevant groups in a locality is important.
- *Patient representatives or advocates* who generally have broader experience, perhaps of working with several groups, wider knowledge about strategic and policy issues, and understand 'the bigger picture'.

Ideally, consultation with 'patients' should involve all three categories. In Williamson's words: 'The patient side of health care is complex but not mysterious. Consulting the "right" patients can be feasible and rewarding' [88]. However, a note of caution must be struck regarding working with patient groups. Many, probably the majority, have links with the pharmaceutical industry, and whilst this is not in itself a problem, educators engaging with such groups should be aware of the potential for conflicts of interest [89].

Simulated Patients

We end our discussion with a brief word on simulated patients, reflecting the increased use of simulation to complement both classroom and clinical learning. Simulation cannot replace authentic experiential learning but can potentially prepare learners for the real world of clinical practice. The first simulated patients were real patients presenting 'standardised' problems [42], but the more common contemporary simulated patient portrays a range of scenarios *outside* their own experience (Silverman J and Britten N, personal communication). The variability in the use of the terms 'standardised' and 'simulated' (often used interchangeably) can cause confusion when trying to identify good practice or interpret research findings. A useful way of thinking about the difference is that situations involving a *simulated* patient (or 'role player') focus predominantly on authenticity, whereas with a *standardised* patient the emphasis is on consistency (e.g. of clinical signs).

Advantages of working with simulated, rather than real, patients include authenticity, consistency, predictability, convenience, and efficiency; challenging situations such as breaking bad news or communicating about sensitive issues can be explored and rehearsed; and scenarios can be customised.

Research has shown that people generally cannot easily distinguish between real patients and well-trained simulators [90, 91]. They are acceptable to learners and faculty, and are effective, reliable, and valid in both instruction and assessment. There is now wide international experience, although some mental health and paediatric problems, as well as those of the frail elderly, may be difficult to simulate and thus may be under-represented, as may certain patient groups, for example, ethnic minorities or people with learning disabilities. The choice of whether to use real or simulated patients in a particular setting will be determined by a range of factors, including the nature of the phenomena to be simulated, intended learning outcomes, local circumstances, and available resources [92]. It is important to remember though that simulation's main purpose is to enable skills development and rehearsal, *not* primarily to ensure incorporation of the patient's voice.

The literature on the effects of simulation on simulated patients themselves is limited. There is potential for harm in some situations – for example, when portraying emotionally intense scenarios – indeed, it has been argued that only professional actors should undertake such demanding roles [93], but the general consensus is that benefits outweigh any disadvantages so long as people are appropriately selected and supported in the role(s) [94]. It is important to pay attention to recruitment, including exploring the person's reasons for wanting to get involved, training and support, and debriefing and de-roling [91, 94]. The use of simulation and simulated patients is discussed further in Chapter 11.

Areas for Further Research

In light of the variable quality of much of the research in this area, a large research agenda has been identified. Typical questions include the following:

- What are the drivers of patient involvement?
- What are the strengths and weaknesses of different approaches, and how do these vary between professions and disciplines, and between countries and across cultures?
- What factors influence what works, and why?
- How do structural and organisational factors such as location, access, and safety influence development of programmes?
- What factors influence patients' experiences of involvement?
- What are the key outcomes, short *and* long term, for all parties?
- What factors influence sustainability of programmes?

Conclusion

From an early stage of training, students and trainees need to actively engage with patients, carers, and families so they can learn to consolidate their learning and put learning from other contexts into practice in the real clinical environment. Appropriate involvement, carried out professionally and sensitively, provides immense benefits not only for the learners, but also for patients. Many patients want to 'give something back' to those who care for them, and engaging in medical education at all levels is one way of so doing.

Teachers and learners need to be aware when learning on 'real' patients is inappropriate. However, many alternatives are available, ranging from paper case scenarios to high-

BOX 15.9 Priorities for action [26]

The 'Vancouver Statement' was developed at an international conference in 2015 [26]. It broadly summarises the current state of patient and public involvement across the continuum of education and training, including benefits and barriers. The statement lays out nine priorities for action in the areas of policy, recognition and support, innovation, research and evaluation, and dissemination which the authors believe 'are necessary in order to embed patient involvement'.

1. Promote patient involvement through directives such as accreditation standards, external and internal policies, pronouncements from professional bodies, and best practice statements.
2. Foster institutional, local, national, and global recognition of patient expertise that grounds and values it; recognise achievement and celebrate success.
3. Increase the diversity of people involved by harnessing the motivation and enthusiasm of patients, community agencies, patient advocacy organisations, and community members.
4. Introduce initiatives to learners as early as possible, and sustain them throughout the educational continuum.
5. Target patient involvement in new and emerging learning activities in order to facilitate a more holistic approach to partnerships and teamwork.
6. Explore and create models to promote collaboration between educational institutions and community organisations to promote patient involvement.
7. Conduct high quality research in partnership with patients to provide further evidence of short-term and long-term impact of patient involvement.
8. Lobby committees to involve patients in planning, delivery, and evaluation of conferences and educational events; lobby community organisations, colleges, and universities and funding bodies to provide grants for people to attend and present at conferences.
9. Create regional networks of people and champions to collaborate, disseminate information, share promising practices, and plan further meetings.

fidelity simulations. The greater emphasis on professionalism, including attending to legal and ethical issues, and the changing agendas relating to patient empowerment and social accountability, shared decision-making, and co-production mean that, for a host of reasons, educators need to pay close attention to seeking active, informed involvement in educational activities from patients and carers. This approach will help put the rhetoric of 'patients as partners' at the centre of the teaching and learning environment, and ultimately at the heart of clinical practice. See Box 15.9.

Acknowledgements

Our thanks to Professor Stephen McHanwell, Newcastle University, for advice about body donation and to Dr Judy Purkis, University of Warwick, for information about UNTRAP.

References

1 Jolly, B. (1998). Historical and theoretical background. In: *Medical Education in the Millennium* (ed. B. Jolly and L. Rees), 171–187. Oxford: Oxford University Press.
2 Rosner, L. (1997). The growth of medical education and the medical profession. In: *Western Medicine* (ed. I. Loudon), 147–159. Oxford: Oxford University Press.
3 Osler, W. (1905). *Aequanimitas* (ed. W. Osler), 332. Philadelphia: PK Blakiston's Son & Co.
4 Rees, C.E., Knight, L.V., and Wilkinson, C.E. (2007). User involvement is a sine qua non, almost, in medical education: learning with rather than just about health and social care service users. *Advances in Health Sciences Education* 12 (3): 359–390.
5. Towle, A., Bainbridge, L., Godolphin, W. et al. (2010). Active patient involvement in the education of health professionals. *Medical Education* 44: 64.
6 Carman, K.L., Dardess, P., Maurer, M. et al. (2013). Patient and family engagement. A framework for understanding the elements and developing interventions and policies. *Health Affairs* 32: 223–231.
7 Department of Health (2005). Creating a patient-centred NHS. http://webarchive.nationalarchives.gov.uk/+/www.dh.gov.uk/en/publicationsandstatistics/publications/publicationspolicyandguidance/dh_4106506 (accessed 1 February 2017).
8 UK Government (2014). The Health and Social Care Act 2008 (Regulated Activities) Regulations 2014. www.legislation.gov.uk/uksi/2014/2936/contents/made (accessed 24 April 2017).
9 Department of Health (2012). The NHS Constitution for England. https://www.gov.uk/government/publications/the-nhs-constitution-for-england (accessed 27 April 2017).
10 Morgan, A. and Jones, D. (2009). Perceptions of service user and carer involvement in healthcare education and impact on students' knowledge and practice: a literature review. *Medical Teacher* 31: 82–95.
11 General Medical Council (2016). Promoting excellence: standards for medical education and training. http://www.gmc-uk.org/education/27394.asp (accessed 1 February 2017).
12 General Medical Council (2010). Patient and Public Involvement in Undergraduate Medical Education. Advice supplementary to Tomorrow's Doctors. GMC, London. http://www.gmc-uk.org/Patient_and_public_involvement_in_undergraduate_medical_education___guidance_under_review_0815.pdf_56438926.pdf (accessed 1 February 2017).
13 Crawford, M.J., Rutter, D., Manley, C. et al. (2002). Systematic review of involving patients in the planning and development of health care. *British Medical Journal* 325: 1263.
14 Woollard, R.F. (2006). Caring for a common future: medical schools' social accountability. *Medical Education* 40: 301–313.
15 Woollard, B. and Boelen, C. (2012). Seeking impact of medical schools on health: meeting the challenges of social accountability. *Medical Education* 46: 21–27.
16 Royal College of Physicians (2005). *Doctors in Society – Medical Professionalism in a Changing World*. London: RCP.
17 Medical Professionalism Project (2002). Medical professionalism in the new millennium: a physicians' charter. *Lancet* 359: 520–521.
18 Hilton, S. and Southgate, L. (2007). Professionalism in medical education. *Teacher and Teacher Education* 23: 265–279.
19 Brody, H. and Doukas, D. (2014). Professionalism: a framework to guide medical education. *Medical Education* 48: 980–987.
20 Engel, G.L. (1989). The need for a new medical model: a challenge for biomedicine. *Journal of Interprofessional Care* 4: 37–53.
21 Stewart, M., Brown, J.B., Weston, W.W. et al. (2003). *Patient-Centred Medicine. Transforming the Clinical Method*, 2e. Oxford: Radcliffe Medical Press.
22 Coulter, A. and Ellins, J. (2007). Effectiveness of strategies for informing, educating and involving patients. *British Medical Journal* 335: 24–27.
23 Realpe, A. and Wallace, L.M. (2010). *What Is Co-production?* London: Health Foundation http://personcentredcare.health.org.uk/sites/default/files/resources/what_is_co-production.pdf (accessed 1 March 2017).
24 Elwyn, G., Frosch, D., Thomson, R. et al. (2012). Shared decision making: a model for clinical practice. *Journal of General Internal Medicine* 27: 1361–1367.
25 Hasman, A., Coulter, A., and Askham, J. (2006). *Education for Partnership*. Oxford: Picker Institute.
26 Towle, A., Farrell, C., Gaines, M.E. et al. (2016). The patient's voice in health and social care professional education. The Vancouver statement. *International Journal of Health Governance* 21: 18–25.
27 Tew, J., Gell, C., and Foster, S. (2004). *Learning from Experience: Involving Service Users and Carers in Mental Health Education and Training*. York: Mental Health in Higher Education/NIMHE West Midlands/Trent WDC.
28 Spencer, J., Blackmore, D., Heard, S. et al. (2000). Patient-oriented learning: a review of the role of the patient in the education of medical students. *Medical Education* 34: 851–857.
29 Cook, D.A. and Triola, M.M. (2009). Virtual patients: a critical literature review and proposed next steps. *Medical Education* 43: 303–311.
30 Saleh, N. (2010). The value of virtual patients in medical education. *Annals of Behavioral Science and Medical Education* 16: 29–31.
31 Poulton, T. and Balasubramaniam, C. (2011). Virtual patients: a year of change. *Medical Teacher* 33: 933–937.
32 Eva, K. (2005). What every teacher needs to know about clinical reasoning. *Medical Education* 39: 98–106.
33 Issenberg, S.B., McGaghie, W.C., Petrusa, E.R. et al. (2005). Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Medical Teacher* 27: 10–28.
34 Cook, D.A., Hamstra, S.J., Brydges, R. et al. (2013). Comparative effectiveness of instructional design features in simulation-based education: systematic review and meta-analysis. *Medical Teacher* 35 (1): e867–e898.
35 Greenhalgh, T. and Hurwitz, B. (1999). Why study narrative? *British Medical Journal* 318: 48–50.
36 Engel, J.D., Zarconi, J., Pethtel, L.L., and Missimi, S.A. (2008). *Narrative in Health Care. Healing Patients, Practitioners, Profession, and Community*. Oxford: Radcliffe Publishing.
37 Hardy, B. (1968). Towards a poetics of fiction: an approach through narrative. *Novel* 2: 5–14.
38 Aronson, J.K. (2000). Autopathography: the patient's tale. *British Medical Journal* 321: 1599–1602.

39 Powley, E. and Higson, R. (2005). *The Arts in Medical Education. A Practical Guide*. Oxford: Radcliffe Publishing.

40 Soar, S., Ryan, S., and Salisbury, H. (2014). Using patients' experiences in e-learning design. *The Clinical Teacher* 11: 80–83.

41 Ramani, S. (2003). Twelve tips to improve bedside teaching. *Medical Teacher* 25: 112–115.

42 Barrows, H.S. and Abrahamson, S. (1964). The programmed patient: a technique for appraising clinical performance in clinical neurology. *Journal of Medical Education* 39: 802–805.

43 Barrows, H.S. (1993). An overview of the uses of standardized patients for teaching and evaluating clinical skills. *Academic Medicine* 68: 443–451.

44 McKimm, J. (2016). Involving patients in clinical teaching. Health Education England Multiprofessional Faculty Development. https://faculty.londondeanery.ac.uk/e-learning/involving-patients-in-clinical-teaching (accessed 1 March 2017).

45 http://Dictionary.com 2013. [online] http://dictionary.reference.com/browse/trialogue?s=t (accessed 1 March 2017).

46 Hardy, P. (2009). The Expert Patient Programme: a critical review. MSc Lifelong learning, policy and research. (www.pilgrimprojects.co.uk/papers/epp_msc.pdf; accessed 24 April 2017).

47 Squire, S. and Hill, P. (2006). The expert patients programme. *Clinical Governance* 11: 17–21.

48 Griffiths, C., Foster, G., Ramsay, J. et al. (2007). How effective are expert patient (lay led) education programmes for chronic disease? *British Medical Journal* 334: 1254–1256.

49 Henderson, K. (2017). The Patient as Teacher. Leeds Institute of Medical Education https://medhealth.leeds.ac.uk/info/842/medical_education_for_patients/811/the_patient_as_teacher (accessed 1 March 2017).

50 Hassell, A. (2102). Patient instructors in rheumatology. *Medical Teacher* 34: 539–542.

51 Rees, C.E. and Monrouxe, L.V. (2011). Medical students learning intimate examinations without valid consent: a multi-centre study. *Medical Education* 45: 261–272.

52 Pickard, S., Baraitser, P., Rymer, J., and Piper, J. (2006). Can gynaecology teaching associates provide high quality effective training for medical students in the UK? Comparative study. *British Medical Journal* 327: 1389–1392.

53 Teodorczuk, A. (2013). Learning from carers: evaluating the effectiveness of a carer-led intervention. Paper presented at Annual Scientific Meeting of the Association for the Study of Medical Education.

54 Towle, A., Brown, H., Hofley, C. et al. (2014). The expert patient as teacher: an interprofessional Health Mentors programme. *The Clinical Teacher* 11: 301–306.

55 Levin, E. (2004). *Involving Service Users and Carers in Social Work Education*. London: Social Care Institute for Excellence www.scie.org.uk/publications/guides/guide04/files/guide04.pdf (accessed 1 March 2017).

56 Howe, A. and Anderson, J. (2003). Involving patients in medical education. *British Medical Journal* 327: 326–328.

57 O'Keefe, M. and Jones, A. (2007). Promoting lay participation in medical school curriculum development: lay and faculty perceptions. *Medical Education* 41: 130–137.

58 Involve. Developing training and support (2013). National Institute for Health Research www.invo.org.uk/resource-centre/training-resource (accessed 24 April 2017).

59 British Medical Association (2008). Role of the Patient in Medical Education. https://www.yumpu.com/en/document/view/18917736/role-of-the-patient-in-medical-education-british-medical-bma (accessed 1 February 2017).

60 University of Warwick (2013). UNTRAP [online]. www2.warwick.ac.uk/fac/cross_fac/healthatwarwick/untrap (accessed 1 February 2017).

61 Jørgensen, C.R., Purkis, J., Blaxter, L., and Tulip, S. (2016). Accredited training for user involvement in higher education teaching – exploring an innovative training programme in public involvement and partnership working. *International Journal of Practice-based Learning in Health and Social Care* 4: 49–62.

62 McLachlan, J.C., Bligh, J., Bradley, P. et al. (2004). Teaching without a cadaver. *Medical Education* 38: 418–424.

63 Bergman, E.M., Verheijen, I.W.H., and Scherpbier, A.J.J.A. (2014). Influences on anatomical knowledge: the complete arguments. *Clinical Anatomy* 27: 296–303.

64 Reiderer, B. (2016). Body donations today and tomorrow. *Clinical Anatomy* 29: 11–16.

65 Bolt, S., Venbrux, E., Eisinga, R. et al. (2010). Motivation for body donation to science: more than an altruistic act. *Annals of Anatomy* 192: 70–74.

66 Cornwall, J., Perry, G.F., Louw, G. et al. (2012). Who donates their body to science? An international multicenter prospective study. *Anatomical Sciences Education* 5: 208–216.

67 Rokade, S.A. and Gaikawad, A.P. (2012). Body donation in India: social awareness, willingness and associated factors. *Anatomical Sciences Education* 5: 83–89.

68 Towle, A. and Godolphin, W. (2016). Patient involvement in health professional education: a bibliography 1975 November 2016 https://pcpe.health.ubc.ca/node/207 (accessed 1 August 2018).

69 Stacy, R. and Spencer, J. (1999). Patients as teachers: a qualitative study of patients' views on their role in a community-based undergraduate project. *Medical Education* 33: 688–694.

70 Jha, V., Quinton, N.D., Bekker, H.L., and Roberts, T.E. (2009). What educators and students really think about using patients as teachers in medical education: a qualitative study. *Medical Education* 43: 449–456.

71 McLachlan, E., King, N., Wenger, E., and Dornan, T. (2012). Phenomenological analysis of patient experiences of medical student teaching encounters. *Medical Education* 46: 963–973.

72 Wykurz, G. and Kelly, D. (2002). Developing the role of patients as teachers: literature review. *British Medical Journal* 325: 818–821.

73 Repper, J. and Breeze, J. (2007). User and carer involvement in the training and education of health professionals: A review of the literature. *International Journal of Nursing Studies* 44 (3): 511–519.

74 Jha, V., Quinton, N.D., Bekker, H.L., and Roberts, T.E. (2009). Strategies and interventions for the involvement of real patients in medical education: a systematic review. *Medical Education* 43: 10–20.

75 Moss, B., Boath, L., Buckley, S., and Colgan, A. (2009). The fount of all knowledge: training required to involve service users and carers in health and social care education and training. *Social Work Education* 28: 562–572.

76 Spencer, J., Godolphin, W., Towle, A., and Karpenko, N. (2011). *Can Patients Be Teachers? Involving Patients and Service Users in Health Professionals' Education*. London: Health Foundation www.health.org.uk/publication/can-patients-be-teachers (accessed 1 February 2017).

77 Monrouxe, L.V., Rees, C.E., and Bradley, P. (2009). The construction of patients' involvement in hospital bedside teaching encounters. *Qualitative Health Research* 19: 918–930.

78 Bleakley, A. and Bligh, J. (2008). Students learning from patients – let's get real in medical education. *Advances in Health Sciences Education* 13: 89–107.

79 Regan de Bere, S. and Nunn, S. (2016). Towards a pedagogy for patient and public involvement in medical education. *Medical Education* 50: 79–92.

80 Towle, A. and Godolphin, W. (2015). Patients as teachers: promoting their authentic and autonomous voices. *The Clinical Teacher* 12: 149–154.

81 Westberg, K., Lynøe, N., Löfgren, M., and Sandlund, M. (2001). Getting informed consent from patients to take part in the clinical training of students: randomized control trial of two strategies. *British Medical Journal* 323: 488.

82 Haffling, A. and Håkansson, A. (2008). Patients consulting with students in general practice: survey of patient's satisfaction and their role in teaching. *Medical Teacher* 30: 622–629.
83 Price, R., Spencer, J., and Walker, J. (2008). Does the presence of medical students affect quality in general practice consultations? *Medical Education* 42: 374–381.
84 Draper, H., Parle, J., and Ross, N. (2008). Medical education and patients' responsibility. Back to the future? *Journal of Medical Ethics* 34: 116–119.
85 Hardy, P. and Stanton, P. (2007). Cultivating compassion: seeing patient voices. *British Medical Journal* 335: 184–187.
86 Benson, J., Quince, T., Hibble, A. et al. (2005). Impact on patients of expanded, general practice based, student teaching: observational and qualitative study. *British Medical Journal* 331: 89.
87 O'Flynn, N., Spencer, J., and Jones, R. (1997). Consent and confidentiality in teaching general practice: survey of patients' views on presence of students. *British Medical Journal* 315: 1142.
88 Williamson, C. (2007). How do we find the right patients to consult? *Quality in Primary Care* 15: 195–199.
89 Taylor, J. and Denegri, S. (2017). Industry links with patient organizations. *British Medical Journal* 356: j1251.
90 Kurtz, S., Silverman, J., and Draper, J. (2005). *Teaching and Learning Communication Skills in Medicine*, 2e. Oxford: Radcliffe Medical Press.
91 Thistlethwaite, J. and Ridgway, G. (2006). *Making It Real. A Practical Guide to Experiential Learning*. Oxford: Radcliffe Publishing.
92 Cleland, J.A., Abe, K., and Rethans, J.J. (2009). The use of simulated patients in medical education. AMEE guide no.42. *Medical Teacher* 31: 477–486.
93 Eagles, J.M., Calder, S.A., Wilson, S. et al. (2007). Simulated patients in undergraduate education in psychiatry. *The Psychiatrist* 31: 187–190.
94 Spencer, J. and Dales, J. (2006). Meeting the needs of simulated patients and caring for the person behind them. *Medical Education* 40: 3–5.

Further Resources

Coulter, A. (2011). *Engaging Patients in Healthcare*. Maidenhead: Open University Press. A definitive text by one of the most prolific champions and scholars of patient engagement in health care.

Association for Standardized Patient Educators (ASPE) www.aspeducators.org (accessed 1 March 2017). ASPE is the international organisation for professionals in the field of standardised patient methodology. It is based in the USA and the website provides good ideas around the use of simulated patients.

16 Learning Medicine With, From, and Through the Humanities

Neville Chiavaroli[1], *Chien-Da Huang*[2,3,4], *and Lynn Monrouxe*[2]
[1]Department of Medical Education, Melbourne Medical School, University of Melbourne, Melbourne, Australia
[2]Chang Gung Medical Education Research Centre (CG-MERC), Chang Gung Memorial Hospital, Linkou, Taiwan (ROC)
[3]Department of Internal Medicine, Chang Gung Memorial Hospital, Chang Gung University College of Medicine, Taipei, Taiwan (ROC)
[4]Department of Thoracic Medicine and Medical Education, Chang Gung Memorial Hospital, Chang Gung University College of Medicine, Taipei, Taiwan (ROC)

KEY MESSAGES

- Medical practice is both *art* and *science*, so a balanced education is required to develop physicians who can combine scientific information, technical skills, and illness schemas to understand, alleviate, and/or cure a particular patient's illness.
- There is no consensus as to what comprises the *medical humanities;* it variously includes an interdisciplinary effort that draws on both creative and intellectual methodological aspects of disciplines such as anthropology, art, bioethics, drama and film, history, literature, music, philosophy, psychology, and sociology.
- Western models of the medical humanities have been privileged whereby the diversity and complexity of local cultural traditions across the world are at risk of becoming marginalised through the use of classic Western texts.
- The humanities in medical education have three major foci: expertise for mastering clinical skills (an *instrumental* rationale), dialogue for understanding patients' experiences and relational practices (an *epistemological* rationale), and expression/transformation of professional identity (*critical/intrinsic* rationale).
- The humanities have been incorporated into medical curricula both informally (by individual, passionate educators) and formally (as a core aspect or as an optional module).
- The humanities can be seen as providing a means through which specific outcomes are achieved, such as the development of communication skills, empathy, ethical reasoning, narrative medicine, reflective practice, and professionalism/professional identities.
- Evidence for the efficacy of the humanities in medical curricula suggests a positive impact on aspects such as empathy, observation skills, emotion, and anxiety.
- As in other areas of the medical curriculum, definitive proof of efficacy of the humanities in medical education is difficult to find; advocates of the humanities have responded to calls for more evidence by rallying to the challenge, resisting such calls, or reframing the nature of the issue.
- Being a good doctor may not require specialist knowledge of the humanities, but the focus of the humanities on meaning, judgement, and human experience make the humanities well suited to support the development of interpersonal competencies, clinical reasoning, and professional identity formation.

Part 1: Why the Humanities in Medical Education?

Introduction

One of the most enduring and widespread ideas about the practice of medicine is that it is both an art and a science [1]. Even today, when scientific and technical advances appear to have made medicine into a much more objective and rational profession, there remains an element of medical practice which is more akin to *art* than *science* and cannot be learnt directly via textbook or teacher [2, 3]. Thus the concept behind the art of medicine relates to the intuitive elements of professional practice, not necessarily acquired through successful completion of a course. As such, this art complements the scientific foundation and practically oriented skills associated with a medical degree. But the art of medicine can be interpreted in a more direct sense: the benefits that the medical profession realises as it draws on elements of knowledge, methods, and practice that are more strongly associated with the academic disciplines of the arts or, more broadly, the humanities. Accordingly, from the latter part of the twentieth century (and arguably earlier),

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

BOX 16.1 'What's in a preposition?'

Prepositions are small words with potentially big effects. In discussing the title of this chapter for the new edition of *Understanding Medical Education*, we wanted to very much keep the phrasing of the original chapter by Jill Gordon and Martyn Evans [4]. We felt their construction was a useful way of highlighting our educational focus ('learning medicine'), while preserving a broad view of the content, methods, and perspectives that are embraced by the field. However the preposition *from* suggests a unidimensional nature of the relationship: the humanities remain outside medicine, to be dipped into as necessary, but not integral to medicine. Our view is very different, hence we add *with* and *through*. The former suggests the humanities are a legitimate counterpart and the latter portrays them as a lens for seeing and understanding medicine: both useful additional metaphors inspired from a recent study [5]. Thus, medicine can, and does, draw on the humanities in each of the ways suggested by these three prepositions, and we feel it is useful to highlight this from the outset – with thanks to the above-mentioned authors.

clinical educators have deliberately drawn on the humanities when teaching medical students the necessary knowledge, skills, and/or attitudes for their future practice. This perspective takes the claim of the art of medicine more seriously, with the field known as the *medical humanities* being seen as a deliberate attempt to make this philosophy explicit both through the phrase itself and through the design of the curriculum. The pedagogical application of the humanities to medical and health professional curricula is the focus of this chapter as we consider the issues around medical professionals learning *with*, *from*, and *through* the medical humanities (see Box 16.1).

Many educators and scholars nowadays prefer the term *health humanities* because it affords inclusivity and broader applicability [6, 7]. Although we are sympathetic with the philosophy underlying this change in terminology, such a change has broader implications for practice and scholarship with the field. Given the focus on medical education throughout this volume, we situate our chapter within the tradition, practice, and scholarship of the field traditionally referred to as the *medical humanities*, but with particular attention to the way the humanities have been applied to educational purposes. Furthermore, we recognise that while early codification of the field has emanated from Western institutions, the medical humanities has been embraced worldwide with a recent rapid global expansion across non-Western cultures [8, 9]. We therefore draw on this wider literature too in the chapter, briefly addressing medical humanities in non-Western cultures and how this has been received and critiqued.

Defining the Humanities in Medical Education

For many years the term 'medical humanities' has been widely used and understood in everyday health care practice, despite the term being occasionally derided and an exact definition remaining elusive [10, 11]. Attempts to define the medical humanities often focus on capturing the range of disciplines that might legitimately lay claim to the field, such as:

> an interdisciplinary, and increasingly international endeavour that draws on the creative and intellectual strengths of diverse disciplines, including literature, art, creative writing, drama, film, music, philosophy, ethical decision making, anthropology, and history, in pursuit of medical educational goals [12, p. 1050].

Yet it is deceptively difficult to specify which disciplines belong to the medical humanities – particularly as even defining the humanities themselves presents a considerable challenge due to the inherent ambiguity of the term, their different approaches, their inherent interdisciplinarity, and their changing conceptions over time [13]. As such, any definition runs the risk of excluding non-mainstream or emerging disciplines (or methodologies) that could potentially be useful in bringing the humanities into medical education.

And what of the social sciences? Do disciplines such as anthropology, economics, psychology, politics, and sociology belong to the (medical) humanities? Indeed, such disciplines are variously located across educational institutions: sometimes within faculties of arts or humanities, sometimes together as social sciences, and sometimes placed within medical and/or health sciences faculties. Any definition adopted needs to consider the issue of applicability across medical humanities programmes and contexts. Thus we favour a definition that includes *any* discipline focusing on human experience and expression such as that promoted by the UK Association for Medical Humanities [14], or the definition proposed by Cole and colleagues: 'an inter- and multidisciplinary field that explores contexts, experiences and critical and conceptual issues in medicine and health care' [15, p. 12]. Such definitions that point to the interdisciplinary nature of the field, without naming any specifically, better acknowledge the broad and eclectic disciplinary base of the medical humanities, and embrace a variety of methodologies. Accepting the scope of the humanities in medicine as being simply concerned with the 'human experience of medicine' [16] is arguably sufficient to convey both the focus of their incorporation into medical education and their constituent sources. Thus a conception such as 'complex interdisciplinary developments within medical pedagogy that combine traditional humanities study with qualitative social sciences and fine arts' [17, p. 935], can be considered as a legitimate perspective alongside more traditional classifications of humanities disciplines.

The Value of the Humanities in Medical Education

The question of why should we bring humanities content, methodologies, and perspectives into medical education is of crucial importance: indeed, educational conservatism, positivistic leanings, and the issue of finding space in an already overcrowded medical curriculum [18] mean that the application of the humanities to clinical training needs strong supporting arguments. Many such arguments for the potential benefit of the humanities in clinical training have been made and categorised, based on one of the following four rationales [15, 19]: intrinsic or non-instrumental; instrumental or practical; critical or intellectual; and epistemological (see Box 16.2). These categories are

BOX 16.2 FOCUS ON: The rationales for and against the inclusion of the humanities in medicine

Intrinsic (non-instrumental) rationale

This emphasises the potential counterbalancing effect of the humanities in the medical curriculum by attempting to introduce an explicit humanistic perspective [20].

Supporting arguments

- Imparts a humanistic perspective to medical students [19–23].
- Focuses on a sensitive, open, and patient-focused approach to learning [24].
- Balances the empirical and objective perspective of biomedicine and technological advancement [16, 25].
- Fosters an 'educated mind' [19] and contribution to a broader academic citizenship [26].

Counter arguments

- Creates a 'false dichotomy' with instrumental approaches [27].
- The humanities have considerable utility in medical education beyond any intrinsic value [27, 28].
- Even intrinsic rationales reflect a fundamental instrumental orientation to medical knowledge and education [29].
- The 'bottom line' argument: 'Would you rather have a physician who is skilled or one who will hold your hand?' [30].

Instrumental (practical) rationale

Focuses on knowledge, skills, and attitudes directly related to clinical practice, such as communication, empathy, reflective practice, narrative competence (further explored in Part 2 of this chapter).

Supporting arguments

- Provides a source of relevant material and perspectives through which students develop insight and understanding of patient experiences and professional skills [15].
- The humanities in medical education will only be acceptable if they can demonstrate instrumental value through measurable outcomes [31, 32].

Counter arguments

- The humanities are devalued when characterised as being merely *in service* to medical education [29, 33], as usefully exemplified in the following quote:

… scholars have begun to worry that the success of the medical humanities is tied up with being *useful* to biomedicine … appearing as the domain of pleasant (but more or less inconsequential) helpmeets – lurking hopefully, poetry books in hand, at the edges of the clinical encounter's 'primal scene' [34, p. 35].

- The preparedness of the humanities faculty to collaborate with the (bio)medical faculty on the education of their students can be seen as a form of *complicity* rather than *collaboration*:

In bringing in the arts and taming them, or in bringing in tame arts (decorative rather than critical, apolitical, and aesthetically unchallenging) as a 'welcome relief' from the supposedly hard grind of science studies, medical education increase[s] the insensibility of medical students and in turn [does] the arts a disservice [35, p. 26].

Critical (intellectual) rationale

This brings a critical, independent, and polemical lens to medical education and health practices [34–38].

Supporting arguments

Facilitates an alternative perspective on the medical world through concepts such as: 'entanglement' [34, 36]; 'making strange' [38]; and 'radical hermeneutics' [37].

Questions the orthodoxy of medicine, including representations of caregivers and patients, the abuse of power and authority, and sciences' attempts to separate biology and culture [39].

Values and applies the humanities' methods of critical analysis and interdisciplinarity rather than simply drawing on narrative texts as sources of patient or practitioner perspectives [34].

Counter arguments

Its rejection of a 'serving' role [36] to medicine may limit its relevance to the primary aims and concerns of medical education [40].

The association with and the vocabulary of critical theory may appear esoteric to many clinical educators and medical students.

(*Continued*)

BOX 16.2 (*Continued*)

Epistemological rationale

This aims to identify and explain how the humanities disciplines, and their methods of inquiry, are fundamental to medical pedagogy and clinical practice [41–45].

Supporting arguments

- The humanities represent characteristic ways of understanding and reasoning which are highly relevant to medical practice, such as a focus on the particular [46], tolerance of ambiguity [35], and access to others' perspectives [47].
- Clinical judgement is made up of *technical* and *humane* components, unified by interpretation and insight, which are respectively underpinned by science on the one hand and arts and humanities on the other [19].
- This is often portrayed by the Aristotelian concept of *phronesis,* which represents the kind of thinking which the humanities deal with naturally:

  No matter how solid the science or how precise the technology that physicians use, clinical medicine remains an interpretive practice. Medicine's success relies on the physicians' capacity for clinical judgement. It is neither a science nor a technical skill (although it puts both to use) but the ability to work out how general rules – scientific principles, clinical guidelines – apply to one particular patient. This is – to use Aristotle's word – *phronesis*, or practical reasoning. It enables physicians to combine scientific information, clinical skill, and collective experience with similar patients to make sense of the particulars of one patient's illness and to determine the best action to take to cure or alleviate it [30, p. 5].

Counter arguments

- Clinical educators might argue that, while there may be commonalities, the epistemology of clinical practice falls within their discipline rather than within humanistic reasoning.
- As a relatively recent perspective, its validity or particular value has not yet been systematically explored in the medical humanities literature.

derived from the literature published under the banner of medical humanities over the past 20 years or so, but a recent scoping review provides further support for the categorisation of these different rationales [5]. After several analyses of the quantitative outcomes of medical humanities programmes, Dennhardt et al. [5] presented a conceptual framework characterising the medical humanities (abbreviated as *art* in their results) as having three major foci: expertise for mastering specific clinical skills (an instrumental rationale); dialogue for understanding patients' experiences and relational practices (an epistemological rationale); and expression/transformation for the professional identity process (a critical/intrinsic rationale).

Part 2: How the Humanities are Incorporated in Medical Curricula

The humanities may be incorporated into medical curricula in various ways, including through informal and formal approaches, as optional or core content, with a disciplinary orientation or an interdisciplinary approach, for the learning of clinical competencies or as a broader input in support of professional formation. In this part of the chapter we draw on a wide range of studies to illustrate the various ways the humanities have been incorporated into medical curricula. We also consider two relatively underappreciated aspects of the place of the humanities in medical curricula: the perspective of students, and the implications of adopting the humanities in cross-cultural contexts.

Informal Approaches

On one level, the humanities can be seen simply as providing sources of anecdote, examples, inspiration, or even a mode of reasoning intended to complement the established medical curriculum. Often clinical educators with a natural inclination for the humanities and humanistic ways of knowing instinctively implement such an approach. Undoubtedly, there have been many examples of inspiring educators who draw, in an apparently seamlessly way, on the humanities to challenge and educate students to think more broadly, deeply, and sensitively about their developing clinical skills and future practice. Sometimes these individuals leave a lasting legacy well beyond their own schools: the physician William Osler, neurologist and popular writer Oliver Sacks, medical educator and ethicist Edmund Pellegrino, and surgeon Atul Gawande are eminent examples of how a humanities perspective on clinical practice can make medicine more compassionate, inspiring, and even more appealing to the public. On the ground in the medical schools, the efforts and impact of dedicated 'champions' of the humanities usually remain local. While they undoubtedly have a significant influence on the clinical practice of their students, unfortunately, when such champions move on, the principles and commitment to the place of the humanities in the curriculum may not find a replacement, the traditional scientific basis of medical curricula resumes its dominance, and students subsequently have little exposure to such alternative perspectives [48, 49]. An example of this may be seen in the medical school of one of the authors of this chapter. Due to the teaching practices and strong advocacy of the humanities in medical education by a clinical educator in the 1970s [50, 51], the medical school at the University of Melbourne has been credited [4] as being one of the early exemplars of the application of the humanities to the medical curriculum. However, this attribution would surprise many educators in the current course in the Melbourne Medical School, where his legacy is little known except by those with a historical interest in the field.

Formal Approaches

In terms of formal implementation of medical humanities in a medical curriculum, a common distinction has been between *additive* and *integrated* approaches [52, 53]: the former applies the humanities to an essentially biomedically focused curriculum, while the latter results in a more fundamental transformation involving the 'nature, goals, and knowledge base of clinical medicine itself' [52, p. 1216]. A further distinction is the extent to which the humanities are optional or core elements of the medical programme. Macnaughton [19] describes three variations on this theme: purely voluntary (and highly popular) units, compulsory units (resisted and disparaged by many students in the example described), and mandatory elective units (commonly known as special study modules or SSMs, which allow students choice in the topic of the elective) [20, 25]. For Macnaughton, the third variation provides the more successful format for inclusion of the humanities: drawing students who are genuinely open to learning with the humanities, yet constituting an integral (and assessed) part of the medical curriculum. While their non-core status may still suggest to students and educators a relatively marginal value in clinical education [54] – medical humanities as merely 'decorative' [55] or 'ornamental' [56] – others might respond, quite reasonably, that the very presence of humanities content within a medical curriculum as a formal, identifiable credit-based subject is already a major advance. Furthermore, forcing medical students to take humanities subjects is no guarantee that they will learn to utilise the lessons clinically; the opposite effect is possible, as many educators have discovered when student evaluations come in [57].

This debate goes to the heart of the value of the humanities in medical education [35, 58] as discussed in Part 1 of this chapter. Rationales that view the humanities as essentially *external* to the actual practice of medicine, however desirable or beneficial ideally, will most likely lead to discretionary inclusion in the curricula. This can certainly be a useful strategy when curricular space is highly contested and there is resistance to any encroachment on time allocated to scientific and clinical disciplines. However, the espoused goal of most humanities educators in medical education is for their content and perspectives to be regarded as an essential and integral part of the curriculum, or as Peterkin puts it, for humanities-based content to ' "infiltrate" the standard curriculum as mandatory, fully intrinsic and highly valued' [54, p. 148]. Indeed, from an epistemological perspective, to *not* include humanistic content or reasoning can only be regarded as providing suboptimal training [25]. The programmes of medical schools that have managed this extent of integration have been periodically described in the medical education and medical humanities literature [58–65], and in particular in special issues; see, for example, the *Journal of Medical Humanities* (December 2013), *Academic Medicine* (October 2003), *Medical Education* (June 2003), and the *Journal for Continuing Education Professionals in Health Sciences* (1995, Volume 2 Issue 3), as well as the regular 'educational case studies' featured in *Medical Humanities* journal. However, for a more recent and comprehensive overview of medical humanities programmes around the world, Alan Bleakley's chapter entitled 'Where do the medical humanities come from and where are they going?' [35] is particularly recommended.

As we shall see later in this chapter, the outcomes of such programmes remain uncertain. Nonetheless, these reports may provide useful exemplars of what core curricula might look like, as opportunities for refinement and improvement, and as a source of insights into the experiences of the educators involved. Early curricular designs tended to treat the humanities as independent disciplines or methodologies, as is reflected in previous overviews of the medical humanities [4, 62]. In recent curriculum developments, however, educators have sought to focus more on specific clinical skills facilitated by study of the humanities, in particular: communication skills [66, 67], empathy [68–70], ethical reasoning [71–73], narrative medicine [74–76], reflective practice [77–79], and professionalism [80–86]. The last approach has provided a particularly fruitful point of entry into the curriculum for the humanities, so we focus here on recent work in this area as an exemplar of how the humanities can be implemented in a curriculum in an interdisciplinary, integrated manner for the development of clinically relevant skills.

Professionalism and Professional Identity

Educators have drawn upon concepts derived from the medical humanities in order to facilitate professionalism learning and, more recently, to support the development of professional identity. Early proponents of this include Coulehan [80, 81], who draws on narrative theory to distinguish between rule-based and narrative-based professionalism [80]. Rule-based professionalism prevails through competency-based notions of educating doctors. From this perspective, different accreditation bodies have drawn up lists of ideal attributes of doctors alongside recommended ways of behaving such that professionalism has become 'a list of required practices' [80, p. 893]. By distinguishing the term narrative-based professionalism, Coulehan asserts that professionalism cannot be learned unless it is embedded deeply within historical and contemporary lived experiences of actual physicians' stories across different cultures – echoing the importance of transformative aspects of medical education noted by other writers [5, 42]. This kind of learning opens up the questions of what it is *to know* and what it is *to be*. Others have followed in Coulehan's footsteps by proposing the study of literature as an antidote to the *professionalism as a competency* movement [82]. Thus it is argued that stories can be learned directly, through role models, or indirectly, through fictional or nonfictional novels, plays, and films [80–83].

Shapiro et al. highlight how the study of literature has an important role in the development of professionalism through the process of *close reading* [82]. They argue that the central tenet of close reading requires that the reader revisit the text to consider alternative or complementary ways of understanding what is meant, while recognising there are no right or wrong answers [82, 84]. In a close reading, the reader tries to understand why stories are told the way they are, who the narrator is, how others are portrayed, who could narrate the same events, and how these might change the nature of the story [82, 85]. In addition, considerations around why certain words are employed and others omitted, the use of metaphors, the specific tone of the story, and how it might shift across the narrating, are all important [82, 86, 87].

Furthermore, some suggest that this narrative world comes to life across hospital hallways, conference rooms, and unit stations [80]. Here, obstacles to the development of professionalism abound [88]: where doctors can be villains, patients can be the butt of gallows humour or used as plot devices against students, and students can be heroes [85, 89]. Thus, educational strategies through which students might make sense of these events, sharing their experiences through oral narratives, have been advocated [80, 88]. Through this sharing, students may come to understand their personal reactions to common professionalism-inhibiting situations, and (re)commit themselves to their future professional selves [80, 88]. In addition to narratives, a much wider range of arts-based tools – such as acting, drawing, poetry, and music – have been identified as being a positive force for professional development, including the development of reflection, communication, critical thinking, leadership, empathy, and complexity amongst a variety of health care groups including medical students [90].

Student Perspectives

An important perspective on the humanities often overlooked is that of the students themselves, for whose benefit the humanities in medical education are frequently invoked. Local evaluations of medical humanities programmes are bound to contain mixed views on the part of students, with interesting and perhaps somewhat contradictory views. For example, a survey of medical student attitudes towards the humanities in a British medical school showed that approximately 90% of students agreed that 'it is important for medical professionals to be broadly educated' and that 'medical professionals need a blend of scientific and humanitarian approaches' [91, p. 627]. However, when asked whether the medical humanities should be offered to students at their medical school, equal proportions (15%) expressed a definitive view for or against, with most opting for the noncommittal response of 'possibly' (69%). More tellingly, of the respondents who answered 'definitely' or 'possibly', three-quarters believed such offerings should be optional, and almost all thought it should not be examined. Predictably, most respondents (57%) felt that there was no room in the curriculum for any extra content. Such surveys remind us that while students' views are valuable in these discussions, and are worth considering, they need to be understood in the context of how the curriculum looks from their perspective – most likely very full [18], regardless of the theoretical benefits of a broader education.

Nevertheless, students frequently manage to articulate the importance of the humanities in ways that are clear and speak to their concerns as they look ahead to their eventual clinical practice. For example, one student wrote the following about the place of narrative medicine in her education:

> Competent medical practice necessarily requires compassion and imagination, and cannot avoid 'big' questions such as the nature and meaning of pain, suffering, and death. However, a doctor who is able to respond usefully to these fundamental questions requires training and skills beyond the merely technical and scientific [92, p. 65].

Reflecting back on his training, one medical resident wrote:

> In medical school, we are ostensibly taught to treat the person, not the disease, inspired by the so-called 'biopsychosocial' model. However, come exam time, the patient devolves back into a pattern of symptoms, signs, and abnormal test results matching a particular disease that we must identify out of a list of 4–5 choices, like a suspect in a police lineup. Little wonder then that the biomedical is still what prevails in medical training and propagates into clinical practice [93].

As illuminating (and ambivalent) as such perspectives may be, clinical educators who use the humanities in their teaching can attest to the issue that, for many medical students, the practices and perspectives of the humanities prove significantly more challenging and esoteric than the quantitative methods and positivist paradigms they experience prior to medical school. This is often voiced as dissatisfaction or rejection of the very premise of the humanities' place in medical education.

An example of the potential opposition of students to the humanities has been recounted by Wear and Aultman [94]. Far from providing the rich textual material and opportunity for students to empathise with the main characters in their chosen novel, they found disengagement, resistance, and outright dismissal of the intended messages. The authors argue that providing students with experiential material is not sufficient to build empathy or understanding: it still allows room for *spectating* as opposed to *witnessing*, particularly if the material proves too confrontational or overtly challenging. As highlighted previously, humanities teaching does not automatically make for humanitarian behaviour. In retrospect, the authors conclude that narrative approaches must move 'beyond a focus on the self and the patient in that individualised, circumscribed relationship and into a collective process involving the social, political, cultural and economic conditions that affect health and well-being' [94, p. 1057]. Context, it seems, really is everything.

A further example of resistance is explored by Birden and Usherwood in a study of Australian medical students' perspectives on how professionalism is taught [95]. A key finding was the way students gamed the system of professionalism assessment by inserting the kind of phrases and sentiments they felt the assessors wanted to hear into their reflective written assignments. Again, enforcing participation or reflection in ways perceived to be inauthentic or token is unlikely to be a productive way of achieving the kind of outcomes the humanities in medicine strive for. This is perhaps more a comment about curriculum theory and practice rather than the value of the humanities per se, but it is an important reminder that merely inserting humanities content or perspectives into medical curricula is not enough; it needs to be incorporated in a manifestly authentic, engaging, and motivating way.

Nevertheless, we should not view such resistance as wholly or even necessarily negative. The humanities can also provide a positive form of resistance, for students as well as teachers and practitioners, as advocates of critical perspectives remind us [36, 56]. A readily available format for such purposes is the cartoon. While a relatively under-utilised medium for teaching critical and interpretive reasoning and exploration of significant professional themes, the cartoon format can be a powerful tool for empowering students to critique poor models of clinical education and practice which can create the dissonance and dilemmas which many medical students experience [96, 97]. Two poignant examples of such cartoons are provided in Figures 16.1 and 16.2 where the objectification of patients and student abuse are portrayed [98]. These comprise the most common types of professionalism dilemmas experienced by UK medical and health care students in a recent large-scale multi-centre study (45% and over 50% respectively in the medical student cohort), resulting in students' experiencing moral distress [99].

But do activities and products of this kind actually count as humanities in medical education? From our perspective, they manifest several aspects of the humanities and what makes them unique in the context of a medical education is that they present a personal narrative (however brief), helping us to understand a particular perspective; they focus on and represent a value system (patient-centred medicine and a supportive education system); and, in this case, they critique a broader reality which is not consistent with those values. We contend that, taken seriously, they can teach as much, if not more, about professionalism as any didactic text or clinical practice lecture. This is precisely the critical function of the humanities in medical education at work.

Cross-cultural Contexts

Having outlined how the humanities have been used primarily in Western medical curricula, we now turn to the issue of cross-cultural transferability of the medical humanities. Recently, a rapid global expansion of the medical humanities is noticeable across non-Western regions; a welcome development, although we share the concerns articulated by several authors about privileging Western models of the medical humanities [100–102]. A particular concern is that non-Western versions of the humanities in medicine absorb *quasi-Western* versions in which the diversity and complexity of local cultural traditions can be marginalised when classic Western texts are adopted along with a medical humanities curriculum. Local circumstances demand local solutions; ideally medical schools would draw on resources relevant to their own cultures when drawing on humanities perspectives for medical education, an approach which is consistent with a fundamental humanities focus which is to acknowledge the importance of contextual and cultural factors – or 'situatedness' – in meaning and experience [103, p. 11]. Recent examples from Taiwan and Saudi Arabia

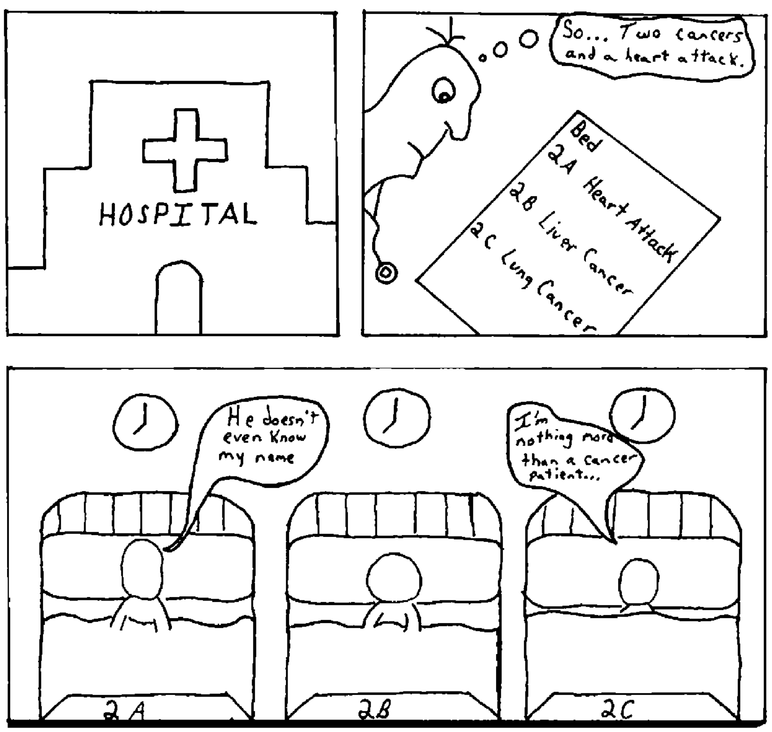


Figure 16.1 Objectification of patients; cartoon by Jordan C. Humphrey [98].

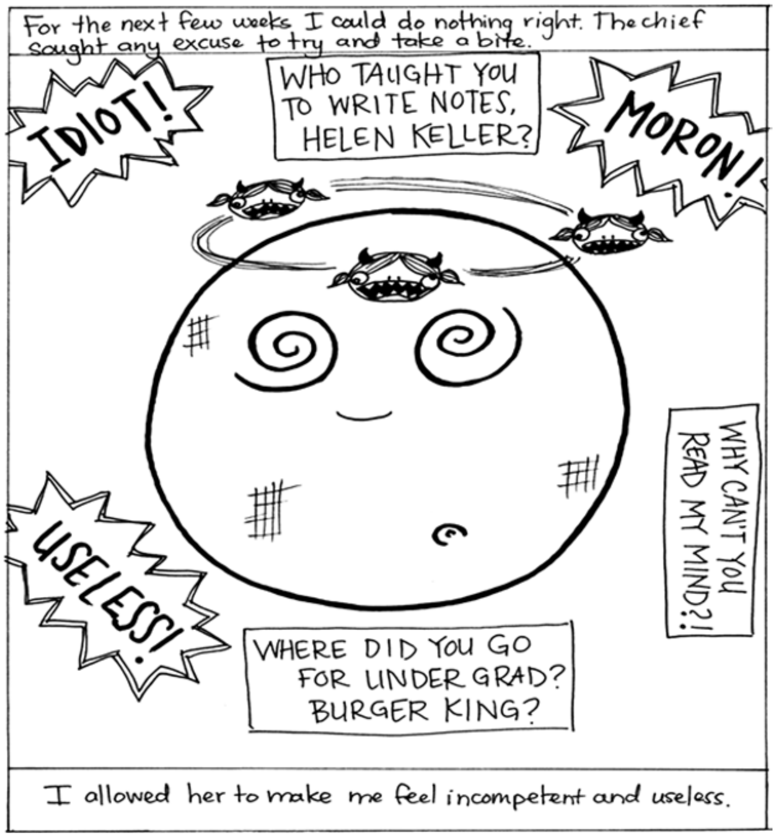


Figure 16.2 Student abuse; cartoon by Trey Banbury [98].

illustrate the incorporation of local cultural perspectives into the design of the medical humanities curriculum very well (see Box 16.3) [104, 105].

Another key example of local refinement of the Western medical humanities curriculum can be seen in the influence of Confucianism within East Asian medical schools. Confucian culture has undoubtedly been a significant influence, unmatched by any other school of thought [106], and is one of the most frequently cited social factors in health care research of older adults in such countries (including Japan, China, Taiwan, and Korea) [107]. It has also significantly influenced medical education in these countries, and presents important synergies with the potential role of the humanities in medicine, including critical approaches which impinge on issues such as patient autonomy, doctor–patient communication, end-of-life decisions, and other fundamental cultural and ethical issues that need to be questioned and understood within the complex mix of local cultural influences [108]. (See Box 16.3 for a brief outline of the recent developments of the humanities in medical schools across four Eastern cultural spaces.)

Part 3: Searching for Impact: Where is the Evidence for the Humanities?

Educators involved with proposing or implementing the inclusion of the humanities in a medical programme are likely to face the question at some point of the process: where is the evidence for the effectiveness of the humanities in medicine? That this question is seldom asked of traditional curricular elements provides no escape; the humanities are viewed as encroaching upon an already established curriculum, and therefore the onus of proof lies with the newcomer. Furthermore, the nature of such evidence, and its acceptability, is frequently determined by those working *outside* of humanities paradigms [35]. Of course, this is not an issue which is limited to medical education; the humanities have for some time faced the challenge of proving themselves in the face of changing educational and societal values [115]. Recent arguments on the value of the humanities represent an expanding genre of passionate defences aimed at academia, governing institutions, and general society [116–118].

Nevertheless, many stakeholders, both supporters and critics alike, have highlighted the apparent lack of evidence for the efficacy of the humanities in medicine. In a landmark article published in 2010, Ousager and Johannessen argued that the medical humanities literature seemed more concerned with 'pleading its case' and justifying ideologically their inclusion in medical curricula than seeking evidence of their impact – for which, the authors claimed, there was only 'sparse' evidence [31]. In an era of outcomes-based education, they argued further, such a lack of empirical evidence might cause the humanities in medical education to 'flounder' if the medical humanities community did not address the need for empirical evidence of its effectiveness. They were not alone in

BOX 16.3 FOCUS ON: The humanities in medicine across cultures: the Cases of Japan, China, Taiwan, and Saudi Arabia

Japan

A recent report on Japanese medical schools reveals that the medical humanities are widely implemented in Japan, with 92% of schools providing some degree of humanities education. The vast majority offer humanities content in the first year of study, with a focus on ethics, patient's rights (including informed consent) and doctor's responsibilities, safety and risk management, problem solving, and the doctor–patient relationship [9], skills reflecting a broader shift in national medical education towards more integrated core curricula [109].

China

China has seen similar reform of medical curricula over the past decade, with an increasing awareness of the possibilities and value of patient-centred approaches, medical humanities, and lifelong learning for professional development of physicians [110]. This new orientation builds upon the Confucian base of much of Chinese medical education discussed above [106, 108]. Nevertheless, challenges for the integration of the humanities into medical curricula remain, such as a lack of organisational independence, lack of appropriate teachers, and, in particular, a strong focus on 'technology-oriented medicine' [110]. A recent issue (#2, 2017) of the Japanese-based journal *Bioscience Trends* features several articles reporting on different aspects of the developments in the medical humanities in China.

Taiwan

Recent medical education reform in Taiwan has facilitated incorporation of the humanities into medical curricula, a development evidenced by a flourishing of recent literature on the topic [8, 9, 102, 104, 111–114]. The focus of such courses has included the development of communication skills, medical ethics, and the physician–patient relationship, with a particular emphasis on communication skills, ethics, and the development of reflective capabilities for professional formation and the enhancement of student empathy and critical thinking [111, 113]. Teaching by educators from nonmedical backgrounds is relatively commonplace in Taiwan [104], while the inclusion of the humanities in medical curricula has been largely endorsed by student groups at the various medical schools [111].

Saudi Arabia

A recent article outlines the relatively novel introduction of medical humanities into the medical curriculum in Saudi Arabia [105]. Two core and local curriculum components – Islamic studies and Arabic studies – were used as the medium for teaching in a medical humanities course. The course, which also draws on student-centred learning approaches, draws on the history of Islamic medicine, Islamic medical ethics, and Arabic medical poetry as the foundation of humanities content aimed at promoting a wide range of course objectives aimed at producing more compassionate, more rounded, and more reflective practitioners.

arguing that the humanities in medicine needed 'harder' evidence: other researchers had made similar claims [32]. The challenges posed by such calls for empirical evidence for the humanities appear to have prompted three main responses from scholars in the field. We consider each in turn, before reflecting on what these responses might mean for the humanities in medical education looking ahead.

Rallying to the Cause

For many researchers in the medical humanities, the necessary response is to accept the challenge posed formally in the literature, and less formally by practitioners and sceptics, and seek the kind of quantitative evidence that will prove convincing enough to justify the humanities' place in the medical curriculum. For the purpose of illustration, in Box 16.4 we sketch briefly some recent research on specific outcomes relating to teaching with the humanities in medical education. Of course what might count as acceptable evidence of impact depends very much on the underlying rationales. An instrumental outlook would seek evidence that a student or trainee displayed better professional skills in some way, such as improvement on clinical assessments or empathy scores; an intrinsic perspective would likely derive reassurance from improved patient feedback of student characteristics; while an epistemological view would probably look further down the line at diagnostic accuracy or reduction of medical error. So not all the evidence will satisfy all advocates of the humanities, and more targeted studies might therefore be required.

Resisting the Calls

Some proponents of the medical humanities contest the demand for evidence as both unfair and misguided. For example, Belling has criticised the sampling methodology and interpretative framework of the Ousager and Johannessen study, resulting in apparent misreadings of some of the studies categorised as reflecting concerns about the role of the humanities in medicine [55]. She further criticises the approach that privileges quantitative data in relation to outcomes as 'reductionist', overlooking qualitative evidence of the value of the humanities. In this sense, Belling echoes the assertion, common in education, that such an approach attempts to 'measure the unmeasurable'

BOX 16.4 WHERE'S THE EVIDENCE: For the effectiveness of the medical humanities?

Impact on empathy

- A number of studies now strongly demonstrate a link between reading literature and the development of empathy (often referred to by the psychological construct known as 'theory of mind') [119–121].
- Studies emerging from psychology build on many smaller scale evaluations of teaching with literature, suggesting that students' capacity to understand and/or relate to the patient's perspective can be improved with even relatively short teaching with literature [47, 64, 122, 123].
- A systematic review examined the effects of reflective writing on empathy [68]. The authors found eight quantitative studies, all reporting a significant change in student empathy.

Impact on observation skills

A systematic review examining the effectiveness of arts-based education found a high-quality evidence base for improvement in students' general observation skills as a result of viewing and discussing artworks [124]. Other studies provide evidence that this learning subsequently transfers to the clinical context [125, 126].

Impact on emotions

Several studies have shown that engaging systematically with art and narrative can have an impact on students' emotional status, including:

- improving students' capacity to focus on and infer emotions [127]
- raising students' awareness and sensitivity to patients in medical contexts [128]
- increasing emotional awareness in others and themselves [129]
- greater emotional benefit from narrative medicine teaching in Asian students on a Chinese medicine course compared to those on a Western medicine course [114].

Impact on tolerance of ambiguity/uncertainty

The practice of medicine comprises inherent uncertainty, arising from knowledge limitations, diagnostic ambiguities, treatment complexities, and unpredictable outcomes [130]. The capacity to better tolerate ambiguity and uncertainty had been a long-standing but largely unverified claim by humanities advocates. However, graduates from a progressive curriculum in the UK with humanities as an integrated and core component were found to have higher scores on *coping with uncertainty* and *tolerance of ambiguity* items than graduates from more traditional medical schools, in a self-report post-graduation survey [131].

Impact on depression and anxiety

The importance of self-care in students and practitioners has been an important rationale for the use of humanities in medical training. One study reported that humanities sessions to help students connect humanism with professional practice benefited the self-perceived resiliency of students, including lower rates of depression and anxiety and greater sense of connection to classmates [132].

[55, 133]. In terms that recall the focus of epistemological rationales for the humanities, Belling further argues that:

> The humanities resist the homogenization of social science metrics, for our focus is on the specific and particular, exactly those aspects of human texts that resist reduction. We value fine distinctions, even at the risk of defaulting to an *n* value of 1. This is precisely why the humanities are so valuable to medicine, for we offer a counterpart to the necessary reductions of the natural sciences. The unit of medicine is the particular patient, always irreducible [55, p. 940].

It seems uncontroversial to say that health systems are not particularly designed to consider the patient as the prime *unit* of consideration, and this reflection gives a sense of both the radical nature of Belling's argument, as well as the far-reaching potential of the humanities.

Charon [17] similarly questioned Ousager and Johannessen's study in terms of the idiosyncratic categorising of the articles, the sampling frame, and the reductive evaluation processes of the study: in particular the way that the complex input and nuanced goals of the humanities were expected to be distilled down to measurable and significant changes in narrowly defined phenomena. Charon also pointed out the power differential in such demands for proof of efficacy for 'any newly introduced curricular material' [17, p. 936] arguing that it was too early (at that point) to expect to see significant changes resulting from humanities input into medical curricula; such changes would need to be sought a decade or so down the track, and in more qualitative and far-reaching ways such as patient empowerment and self-determination.

However, even this longer view may not prove particularly fruitful. As Cook and West point out, issues such as 'dilution, feasibility, failure to establish a causal link, potentially biased outcome selection, and teaching to the test all challenge the routine use of patient outcomes [for evaluation]' [134, p. 5].

In other words, there are simply too many uncontrollable variables in medical education to establish the kind of causal claims many seem to expect of the medical humanities. This is a more far-reaching issue than just for the medical humanities, well appreciated in both medical education [135] and in educational research more generally [136]. Not surprisingly, then, most areas of medical education appear to have limited evidence of their impact on the development of competent physicians [132]. Finally, some educators have associated the demands for evidence with the competency-based movement in medical education, and argued that a competency framework is misguided when evaluating the value of the humanities, especially its critical and emancipatory aspects [42, 137].

Reframing the Expectations

Reframing represents a compromise between rallying and resisting. Rather than rejecting outright the need for evidence of efficacy, educators and researchers adopting this approach argue that the nature of humanistic practice means that evidence must be sought in criteria and methods that differ from quantitative and positivist methodologies. This viewpoint is illustrated particularly well in the recent study by Dennhardt and colleagues [5]. In the introduction to their scoping review of quantitative outcome studies in the medical humanities, they state their sympathy for the position which rejects the demands for evidence from the humanities, yet then decide to accept the challenge to account for its 'effectiveness and utility', acknowledging the necessity in an era of outcomes-based medical education. The catch, however, is that they subsequently argue that such evidence cannot be found until we properly understand what the medical humanities teach – and why. Their review proceeds to show that the typical descriptive analyses of humanities inputs or programmes are not sensitive enough to do justice to the enormous variation in the form and measurement of medical humanities teaching – a finding that must be construed as a further critique of the methodology employed by Ousager and Johannessen. Instead, conceptual and discursive analyses are deemed necessary. This gives rise in their study to three main functions of the humanities in medical education: (i) as expertise, (ii) as dialogue and expression; and (iii) as personal growth and activism, or what many would call 'professional identity formation'.

This is where curriculum theory comes into play. As we have noted on several occasions in this chapter, and as discussed extensively in Chapter 5, curriculum decisions are not value free, and the decision to include or exclude the humanities tends to involve other factors unrelated to its value as a learning approach for medical students. Lee and colleagues articulate this particularly well when they write that 'curriculum inquiry … attends to issues about what is being conveyed (or is intended to be conveyed) within a curriculum and, in particular, *the choices that are made about values, emphases and directions* that are not simply derivable from "evidence" of what works' (our italics) [138, p. 68]. We still stand by the importance of being clear and explicit about the rationale being used to support the value of the humanities in medical education, even though ensuing debates will often play out politically, whether deliberate or not. This can also be seen in discussions around the 'hidden curriculum' in medicine, where the message received by students does not fully accord with the curriculum and objectives formally documented [139]. Being explicit, and realistic, about the rationales and value of the humanities in medical education can help tackle such curricular arguments and unintended effects; and they may even facilitate conditions where, when the rationales makes sense and the timing is right, the incontrovertible evidence can wait and the curriculum can change to reflect new emphases and new values. We believe such a climate is increasingly developing for the medical humanities, where the focus in research should now shift to how *best* to incorporate the humanities into medical education, rather than continuing to discuss and debate *whether* to incorporate them.

Conclusion

In the *British Medical Journal* letters section several years ago, readers were invited to articulate what a good doctor meant to them. The following response caught our eye:

> The truly good doctor must, of course, be technically proficient and know the craft of medicine. In addition, however, the good doctor must be able to understand patients in enough breadth to call on a community of skilled healers – nurses, social workers, insurance specialists, yoga teachers, psychotherapists, technicians, chaplains, whatever is necessary – to help restore the person to health (or perhaps, to support the person in their journey towards death). To do that, the doctor must be able to be touched by the patient's life as well as his or her illness. The doctor need not be an anthropologist but must know how to ask about a person's culture; he or she need not be a marriage counsellor but must be able to spot the signs of spousal abuse or the depression that may be the result of a failing union. Good doctors are humble doctors, willing to listen to their patients and gather together the full array of resources – medical, human, social, and spiritual – that will contribute to their patients' healing [140, p. 712].

We couldn't agree more. The *good doctor* (or rather the *effective* doctor?) doesn't need to be an anthropologist, or sociologist, or psychologist, or social worker, or ethicist, or even a humanities scholar. But for that matter they shouldn't need to be an anatomist, microbiologist, pharmacologist, or neuroscientist either. They simply need to know when such knowledge applies, and when such a perspective is relevant. They need to spot the subtle relational signs, know when and how to elicit the relevant information, and care enough to want to do it, time and time again. While perhaps the will to care may be difficult to instil, a close attention to words and meaning can be; based on, to borrow Belling's phrase, 'reading for meaning rather than for data' [55], or even an earlier pronouncement about the importance of words and meaning in medicine: 'what the scalpel is to the surgeon, words are to the clinician' [141]. Such a focus is ultimately what the humanities, in its enormous variety of content, method, and perspective, can offer medical education – be it from, through, with … across, amid, beside, between, inside, upon, or via, or in any other prepositional way educators see fit.

References

1 Saunders, J. (2001). The practice of clinical medicine as an art and as a science. *Western Journal of Medicine* 174: 137–141.

2 Wong, J.G. (2015). The art and science of medical education. *The American Journal of the Medical Sciences* 349: 191.

3 Tooke, J. (2016). The science (and art) of medicine. *Lancet* 387: S6–S7.

4 Gordon, J.J. and Evans, H.M. (2010). Learning medicine from the humanities. In: *Understanding Medical Education* (ed. T. Swanwick), 83–98. Wiley-Blackwell.

5 Dennhardt, S., Apramian, T., Lingard, L. et al. (2016). Rethinking research in the medical humanities: a scoping review and narrative synthesis of quantitative outcome studies. *Medical Education* 50: 285–299.

6 Crawford, P., Brown, B., Baker, C. et al. (2015). *Health Humanities*. London: Palgrave Macmillan.

7 Jones, T., Blackie, M., Garden, R., and Wear, D. (2017). The almost right word: the move from medical to health humanities. *Academic Medicine* 92 (7): 932–935.

8 Tsai, D.J. (2008). Community-oriented curriculum design for medical humanities. *Kaohsiung Journal of Medical Sciences* 24 (7): 373–379.

9 Song, P. and Tang, W. (2017). Emphasizing humanities in medical education: promoting the integration of medical scientific spirit and medical humanistic spirit. *Bioscience Trends* 11 (2): 128–133.

10 Knight, L.V. (2006). A silly expression: consultants' implicit and explicit understanding of Medical Humanities. A qualitative analysis. *Medical Humanities* 32: 119–124.

11 Evans, M. and Finlay, I.G. ed. (2001). *Medical Humanities*, 7–11. London: BMJ Books.

12 Kirklin, D. (2003). The Centre for Medical Humanities, Royal Free and University College Medical School, London, England. *Academic Medicine* 78: 1048–1053.

13 Bod, R. (2013). *A New History of the Humanities: The Search for Principles and Patterns from Antiquity to the Present*. Oxford: Oxford University Press.

14 Arnott, R., Bolton, G., Evans, M. et al. (2001). Proposal for an academic Association for Medical Humanities. *Medical Humanities* 27: 104–105.

15 Cole, T.R., Carlin, N.S., and Carson, R.A. (2015). *Medical Humanities. An Introduction*. Cambridge: Cambridge University Press.

16 Evans, M. (2002). Reflections on the humanities in medical education. *Medical Education* 36: 508–513.

17 Charon, R. (2010). Calculating the contributions of humanities to medical practice – motives, methods, and metrics. *Academic Medicine* 85 (6): 935–937.

18 Kirklin, D. and Richardson, R. ed. (2001). *Medical Humanities: A Practical Introduction*. London: Royal College of Physicians of London.

19 Macnaughton, J. (2000). The humanities in medical education: context, outcomes and structures. *Medical Humanities* 26: 23–30.

20 Macnaughton, J. (2001). Why medical humanities now? In: *Medical Humanities* (ed. M. Evans and I.G. Finlay), 187–203. London: BMJ Books.

21 Greaves, D. (2001). The nature and role of the medical humanities. In: *Medical Humanities* (ed. M. Evans and I.G. Finlay), 13–22. London: BMJ Books.

22 Gordon, J.J. (2008). Humanising doctors: what can the medical humanities offer? *Medical Journal of Australia* 189: 420–421.

23 Warner, J.H. (2011). The humanising power of medical history: responses to biomedicine in the 20th century United States. *Medical Humanities* 37: 91–96.

24 Gordon, J. (2005). Medical humanities: to cure sometimes, to relieve often, to comfort always. *Medical Journal of Australia* 182: 5–8.

25 Spicer, J., Harrison, D., and Winning, J. (2013). The humanities in medical education. In: *Oxford Textbook of Medical Education* (ed. K. Walsh), 233–254. Oxford: Oxford University Press.

26 Curry, R.H. and Montgomery, K. (2010). Toward a liberal education in medicine. *Academic Medicine* 85: 283–287.

27 Blease, C. (2016). In defence of utility: the medical humanities and medical education. *Medical Humanities* 42: 103–108.

28 Wear, D. (2009). The medical humanities: toward a renewed praxis. *Journal of Medical Humanities* 30: 209–220.

29 Bishop, J.P. (2008). Rejecting medical humanism: medical humanities and the metaphysics of medicine. *Journal of Medical Humanities* 29 (1): 15–25.

30 Montgomery, K. (2006). *How Doctors Think. Clinical Judgment and the Practice of Medicine*. Oxford: Oxford University Press.

31 Ousager, J. and Johannessen, H. (2010). Humanities in undergraduate medical education: a literature review. *Academic Medicine* 85: 988–998.

32 Wershof Schwartz, A., Abramson, J.S., Wojnowich, I. et al. (2009). Evaluating the impact of the humanities in medical education. *Mount Sinai Journal of Medicine* 76: 372–380.

33 Macneill, P.U. (2011). The arts and medicine: a challenging relationship. *Medical Humanities* 37 (2): 85–90.

34 Fitzgerald, D. and Callard, F. (2016). Entangling the medical humanities. In: *The Edinburgh Companion to the Critical Medical Humanities* (ed. A. Whitehead, A. Woods, S. Atkinson, et al.), 35–49. Edinburgh: Edinburgh University Press.

35 Bleakley, A. (2015). *Medical Humanities and Medical Education: How the Medical Humanities Can Shape Better Doctors*. London and New York: Routledge.

36 Viney, W., Callard, F., and Woods, A. (2015). Critical medical humanities: embracing entanglement, taking risks. *Medical Humanities* 41: 2–7.

37 Wear, D. and Aultman, J.M. (2007). Creating difficulties everywhere: medical humanities as radical hermeneutics text. *Perspectives in Biology and Medicine* 50: 348–362.

38 Kumagai, A.K. and Wear, D. (2014). 'Making strange': a role for the humanities in medical education. *Academic Medicine* 89: 973–977.

39 Jones, T., Wear, D., and Friedman, L.D. ed. (2014). *Health Humanities Reader*. New Brunswick: Rutgers University Press.

40 Bleakley, A. (2015). When I say … the medical humanities in medical education. *Medical Education* 49 (10): 959–960.

41 Kuper, A. and D'Eon, M. (2011). Rethinking the basis of medical knowledge. *Medical Education* 45: 36–43.

42 Kumagai, A.K. (2014). From competencies to human interests: ways of knowing and understanding in medical education. *Academic Medicine* 89: 978–983.

43 Boudreau, J.D. and Fuks, A. (2015). The humanities in medical education: ways of knowing, doing and being. *Journal of Medical Humanities* 36: 321–336.

44 Chiavaroli, N. (2017). Knowing how we know: an epistemological rationale for the medical humanities. *Medical Education* 51: 13–21.

45 Kuper, A., Veinot, P., Leavitt, J. et al. (2017). Epistemology, culture, justice and power: non-bioscientific knowledge for medical training. *Medical Education* 51 (2): 158–173.

46 Bruner, J. (1985). Narrative and paradigmatic modes of thought. In: *Learning and Teaching the Ways of Knowing. 84th Yearbook of the National Society for the Study of Education* (ed. E. Eisner), 97–115. Chicago: University of Chicago Press.

47 Shapiro, J. (2008). Walking a mile in their patients' shoes: empathy and othering in medical students' education. *Philosophy, Ethics, and Humanities in Medicine* 3: 10.

48 Pattison, S. (2003). Medical humanities: a vision and some cautionary notes. *Medical Humanities* 29: 33–36.

49 Chiavaroli, N. and Ellwood, C. (2012). The medical humanities and the perils of curricular integration. *Journal of Medical Humanities* 33: 245–254.

50 Moore, A.R. (1975). The art of medicine: a missing subject? *Medical Journal of Australia* 2 (1): 27–28.
51 Moore, A.R. (1976). Medical humanities: a new medical adventure. *New England Journal of Medicine* 295 (26): 1479–1480.
52 Evans, M. and Greaves, D. (1999). Exploring the medical humanities. *British Medical Journal* 319: 1216.
53 Oyebode, F. (2010). The medical humanities: literature and medicine. *Clinical Medicine* 10: 242–244.
54 Peterkin, A. (2016). Curating the medical humanities curriculum: twelve tips. *Medical Humanities* 42 (3): 147–148.
55 Belling, C. (2010). Sharper instruments: on defending the humanities in undergraduate medical education. *Academic Medicine* 85 (6): 938–940.
56 Shapiro, J. (2012). Whither (whether) medical humanities? The future of humanities and arts in medical education. *Journal for Learning through the Arts* 8 (1): http://escholarship.org/uc/item/3x2898ww#page-4 (accessed 19 July 2017).
57 Gull, S.E. (2005). Embedding the humanities into medical education. *Medical Education* 39: 235–236.
58 Evans, H.M. and Macnaughton, R.J. (2006). A 'core curriculum' for the medical humanities? *Medical Humanities* 32: 65–66.
59 Polianski, I.J. and Fangerau, H. (2012). Toward 'harder' medical humanities: moving beyond the 'two cultures' dichotomy. *Academic Medicine* 87: 121–126.
60 Gurtoo, A., Ranjan, P., Sud, R., and Kumari, A. (2013). A study of acceptability and feasibility of integrating humanities based study modules in undergraduate curriculum. *The Indian Journal of Medical Research* 137: 197–202.
61 Supe, A. (2012). Medical humanities in the undergraduate medical curriculum. *Indian Journal of Medical Ethics* 9: 263–265.
62 Grant, V.J. (2002). Making room for medical humanities. *Medical Humanities* 28: 45–48.
63 Bleakley, A., Marshall, R., and Bromer, R. (2006). Toward an aesthetic medicine: developing a core medical humanities undergraduate curriculum. *Journal of Medical Humanities* 27: 197–213.
64 Chen, J.Y., Salter, D.J., and Chan, L.C. (2010). Pen, brush and camera: outcomes-based medical humanities. *Medical Education* 44: 1139.
65 Kitzes, J.A., Savich, R.D., Kalishman, S. et al. (2007). Fitting it all in: integration of 12 cross-cutting themes into a School of Medicine curriculum. *Medical Teacher* 29: 437–442.
66 Sklar, D., Doezeman, D., McLaughlin, S., and Helitzer, D. (2002). Teaching communications and professionalism through writing and humanities: reflections of ten years of experience. *Academic Emergency Medicine* 9: 1360–1364.
67 Choudhary, A. and Gupta, V. (2015). Teaching communications skills to medical students: introducing the fine art of medical practice. *International Journal of Applied and Basic Medical Research* 5: S41–S44.
68 Chen, I. and Forbes, C. (2014). Reflective writing and its impact on empathy in medical education: systematic review. *Journal of Educational Evaluation for Health Professions* 11: 20.
69 Muszkat, M., Yehuda, A.B., Moses, S., and Naparstek, Y. (2010). Teaching empathy through poetry: a clinically based model. *Medical Education* 44: 503–503.
70 Garden, R.E. (2007). The problem of empathy: medicine and the humanities. *New Literary History* 38 (3): 551–567.
71 Rees, G. (2010). The ethical imperative of medical humanities. *Journal of Medical Humanities* 31: 267–277.
72 del Pozo, P.R. and Fins, J.J. (2005). The globalization of education in medical ethics and humanities: evolving pedagogy at Weill Cornell Medical College in Qatar. *Academic Medicine* 80: 135–140.
73 Kanoti, G.A. (1990). Ethics and medical humanities education: the Cleveland Clinic Foundation. *Journal of Clinical Ethics* 1: 294–297.
74 Marini, M.G. (2016). *Narrative Medicine: Bridging the Gap between Evidence-Based Care and Medical Humanities*. Springer International Publishing.
75 Charon, R. (2006). *Narrative Medicine: Honoring the Stories of Illness*. Oxford: Oxford University Press.
76 Arntfield, S.L., Slesar, K., Dickson, J., and Charon, R. (2013). Narrative medicine as a means of training medical students toward residency competencies. *Patient Education and Counseling* 91: 280–286.
77 Bolton, G. (2001). *Reflective Practice: Writing and Professional Development*. London: Paul Chapman Publishing.
78 Branch, W.T. (2010). The road to professionalism: reflective practice and reflective learning. *Patient Education and Counseling* 80 (3): 327–332.
79 Liu, G.Z., Jawitz, O.K., Zheng, D. et al. (2016). Reflective writing for medical students on the surgical clerkship: oxymoron or antidote? *Journal of Surgical Education* 73 (2): 296–304.
80 Coulehan, J. (2005). Today's professionalism: engaging the mind but not the heart. *Academic Medicine* 80: 892–898.
81 Coulehan, J. (2007). Written role models in professionalism education. *Medical Humanities* 33: 106–109.
82 Shapiro, J., Nixon, L.L., Wear, S.E., and Doukas, D.J. (2015). Medical professionalism: what the study of literature can contribute to the conversation. *Philosophy, Ethics, and Humanities in Medicine* 10: 1–8.
83 Rees, C.E. and Knight, L.V. (2007). 'The stroke is eighty nine': understanding unprofessional behaviour through physician-authored prose. *Medical Humanities* 33: 38–43.
84 McCaffrey, G., Venturato, L., Patterson, J.D. et al. (2017). Bringing a novel to practice: an interpretive study of reading a novel in an undergraduate nursing practicum course. *Nurse Education in Practice* 24: 84–89.
85 Monrouxe, L.V. and Rees, C.E. (2017). Hero, voyeur, judge: understanding medical students' moral identities through professionalism dilemma narratives. In: *The Self, Social Identity and Education* (ed. K.I. Mavor, M. Platow and B. Buzumic), 297–319. Psychology Press.
86 Rees, C., Knight, L., and Cleland, J. (2009). Medical educators' metaphoric talk about their assessment relationships with students: 'you don't want to sort of be the one who sticks the knife in them'. *Assessment & Evaluation in Higher Education* 34: 455–467.
87 Rees, C., Knight, L., and Wilkinson, C. (2007). Doctors being up there and we being down here: a metaphorical analysis of talk about student/doctor-patient relationships. *Social Science & Medicine* 65: 725–737.
88 Monrouxe, L.V. and Rees, C.E. (2017). *Healthcare Professionalism: Improving Practice through Reflections on Workplace Dilemmas*. Oxford: Wiley.
89 Monrouxe, L.V., Rees, C.E., and Bradley, P. (2009). The construction of patients' involvement in hospital bedside teaching encounters. *Qualitative Health Research* 19: 918–930.
90 Lake, J., Jackson, L., and Hardman, C. (2015). A fresh perspective on medical education: the lens of the arts. *Medical Education* 49: 759–772.
91 Hodgson, K. and Thomson, R. (2000). What do medical students read and why? A survey of medical students in Newcastle-upon-Tyne, England. *Medical Education* 34: 622–629.
92 Chapple, S. (2015). Medical humanities and narrative medicine. *Australian Medical Student Journal* 6: 63–65.
93 Han, M. Artbeat: Reflections of a 4th year medical resident. Why our Companion Curriculum matters. (Undated). https://utmedhumanities.wordpress.com/reflections-of-a-4th-year-medical-resident-why-our-companion-curriculum-matters-mihan-han (accessed 19 July 2017).
94 Wear, D. and Aultman, J.M. (2005). The limits of narrative: medical student resistance to confronting inequality and oppression in literature and beyond. *Medical Education* 39: 1056–1065.
95 Birden, H.H. and Usherwood, T. (2013). 'They liked it if you said you cried': how medical students perceive the teaching of professionalism. *Medical Journal of Australia* 199: 406–409.
96 Chiavaroli, N. and Trumble, S. (2016). Not so funny: prompting reflection through single-panel cartoons. In: *Keeping Reflection*

Fresh: A Practical Guide for Clinical Educators (ed. A. Peterkin and P. Brett-MacLean), 107–110. Kent State University Press.

97 Green, M.J. and Myers, K.R. (2010). Graphic medicine: use of comics in medical education and patient care. *British Medical Journal* 340.

98 Green, M., Myers, K., George, D.R. (2017). The comics revealing medical school's hidden flaws and hard lessons. *The Atlantic*, March 24.

99 Monrouxe, L.V., Rees, C.E., Dennis, I., and Wells, S.E. (2015). Professionalism dilemmas, moral distress and the healthcare student: insights from two online UK-wide questionnaire studies. *BMJ Open* 5: e007518.

100 Bleakley, A., Brice, J., and Bligh, J. (2008). Thinking the post-colonial in medical education. *Medical Education* 42 (3): 266–270.

101 Hooker, C. and Noonan, E. (2011). Medical humanities as expressive of Western culture. *Medical Humanities* 37: 79–84.

102 Tsai, S.L., Ho, M.J., Hirsh, D., and Kern, D.E. (2012). Defiance, compliance, or alliance? How we developed a medical professionalism curriculum that deliberately connects to cultural context. *Medical Teacher* 34 (8): 614–617.

103 Looseley, D. (2011). Making an 'impact': some personal reflections on the Humanities in the UK. *Arts and Humanities in Higher Education* 10 (1): 9–18.

104 Tseng, F.Y., Shieh, J.Y., Kao, T.W. et al. (2016). Developing and evaluating medical humanities problem-based learning classes facilitated by the teaching assistants majored in the liberal arts: a longitudinal crossover study. *Medicine (Baltimore)* 95 (6): e2765.

105 Abdel-Halim, R.E. and AlKattan, K.M. (2012). Introducing medical humanities in the medical curriculum in Saudi Arabia: a pedagogical experiment. *Urology Annals* 4: 73–79.

106 Guo, Z. (1995). Chinese Confucian culture and the medical ethical tradition. *Journal of Medical Ethics* 21: 239–246.

107 Koh, E.K. and Koh, C.K. (2008). Caring for older adults: the parables in Confucian texts. *Nursing Science Quarterly* 21: 365–368.

108 Chen, L.C., Reich, M.R., and Ryan, J. (2017). *Medical Education in East Asia: Past and Future*. Indiana University Press.

109 Kozu, T. (2006). Medical education in Japan. *Academic Medicine* 81 (12): 1069–1075.

110 Tang, W. (2017). Propelling medical humanities in China. *Bioscience Trends* 11 (2): 125–127.

111 Chou, J.Y., Chiu, C.H., Lai, E. et al. (2012). Medical education in Taiwan. *Medical Teacher* 34: 187–191.

112 Wong, Y.F., Lin, S.J., Cheng, H.C. et al. (2012). The formation and performance of medical humanities by interns in a clinical setting. *Tzu Chi Medical Journal* 24 (1): 5–11.

113 Wang, Y., Kao, P., and Liao, H. (2015). Integrating narrative medicine storytelling into the study of the medical humanities: enhancing medical care students' sense of meaning in life and critical thinking capacity. *Journal of Medical Education* 19: 53–64.

114 Huang, C.D., Liao, K.C., Chung, F.T. et al. (2017). Different perceptions of narrative medicine between Western and Chinese medicine students. *BMC Medical Education* 17 (1): 85.

115 Weisbuch, R. (1999). Six proposals to revive the humanities. *Chronicle of Higher Education* B4–B5.

116 Gilman, S.L. (2000). *The Fortunes of the Humanities*. Stanford: Stanford University Press.

117 Nussbaum, M.C. (2010). *Not for Profit: Why Democracy Needs the Humanities*. Princeton: Princeton Unviersity Press.

118 Small, H. (2013). *The Value of the Humanities*. Oxford: Oxford University Press.

119 Kidd, D.C. and Castano, E. (2013). Reading literary fiction improves theory of mind. *Science* 342: 377–380.

120 Djikic, M., Oatley, K., and Moldoveanu, M.C. (2013). Reading other minds: effects of literature on empathy. *Scientific Study of Literature* 3: 28–47.

121 Mar, R.A., Oatley, K., and Peterson, J.B. (2009). Exploring the link between reading fiction and empathy: ruling out individual differences and examining outcomes. *Communications* 34 (4): 407–428.

122 Shapiro, J., Morrison, E., and Boker, J. (2004). Teaching empathy to first year medical students: evaluation of an elective literature and medicine course. *Education for Health (Abingdon)* 17: 73–84.

123 Graham, J., Benson, L.M., Swanson, J. et al. (2016). Medical humanities coursework is associated with greater measured empathy in medical students. *The American Journal of Medicine* 129: 1334–1337.

124 Perry, M., Maffulli, N., Willson, S., and Morrissey, D. (2011). The effectiveness of arts-based interventions in medical education: a literature review. *Medical Education* 45: 141–148.

125 Naghshineh, S., Hafler, J.P., Miller, A.R. et al. (2008). Formal art observation training improves medical students' visual diagnostic skills. *Journal of General Internal Medicine* 23: 991–997.

126 Braverman, I.M. (2011). To see or not to see: how visual training can improve observational skills. *Clinics in Dermatology* 29: 343–346.

127 Bardes, C.L., Gillers, D., and Herman, A.E. (2001). Learning to look: developing clinical observational skills at an art museum. *Medical Education* 35: 1157–1161.

128 Reilly, J.M., Ring, J., and Duke, L. (2005). Visual thinking strategies: a new role for art in medical education. *Family Medicine* 37: 250–252.

129 Shapiro, J., Rucker, L., and Beck, J. (2006). Training the clinical eye and mind: using the arts to develop medical students' observational and pattern recognition skills. *Medical Education* 40: 263–268.

130 Hancock, J., Roberts, M., Monrouxe, L., and Mattick, K. (2015). Medical student and junior doctors' tolerance of ambiguity: development of a new scale. *Advances in Health Sciences Education: Theory and Practice* 20: 113–130.

131 Bleakley, A. and Brennan, N. (2011). Does undergraduate curriculum design make a difference to readiness to practice as a junior doctor? *Medical Teacher* 33: 459–467.

132 Frank, C. and Martin, R.E. (2015). Humanities and geriatric education: a strategy for recruitment? *Canadian Geriatrics Journal* 18: 37–41.

133 Unterhalter, E. (2017). Negative capability? Measuring the unmeasurable in education. *Comparative Education* 53: 1–16.

134 Cook, D.A. and West, C.P. (2013). Reconsidering the focus on 'outcomes research' in medical education: a cautionary note. *Academic Medicine* 88: 162–167.

135 Norman, G.R. and Schmidt, H.G. (2000). Effectiveness of problem-based learning curricula: theory, practice and paper darts. *Medical Education* 34: 721–728.

136 Berliner, D.C. (2002). Educational research: the hardest science of all. *Educational Researcher* 31 (8): 18–20.

137 Wear, D. and Zarconi, J. (2011). The treachery of images: how René Magritte informs medical education. *Journal of General Internal Medicine* 26: 437–439.

138 Lee, A., Steketee, C., Rogers, G., and Moran, M. (2013). Towards a theoretical framework for curriculum development in health professional education. *Focus on Health Professional Education: A Multi-disciplinary Journal* 14: 64–77.

139 Hafferty, F.W. and Franks, R. (1994). The hidden curriculum, ethics teaching, and the structure of medical education. *Academic Medicine* 69 (11): 861–871.

140 Letters, B.M.J. (2002). What's a good doctor and how do you make one? *British Medical Journal* 325: 711–715.

141 Tumulty, P. (1970). What is a clinician and what does he do? *New England Journal of Medicine* 283 (1): 20–24.

Further Reading

Bates, V., Bleakley, A., and Goodman, S. (2013). *Medicine, Health and the Arts: Approaches to the Medical Humanities*. Routledge Chapman & Hall.

Belfiore, E. and Upchurch, A. ed. (2013). *Humanities in the Twenty-First Century: Beyond Utility and Markets*. Cambridge: Palgrave Macmillan.

Bleakley, A. (2015). *Medical Humanities and Medical Education: How the Medical Humanities Can Shape Better Doctors*. London and New York: Routledge.

Cole, T.R., Carlin, N.S., and Carson, R.A. (2015). *Medical Humanities. An Introduction*. Cambridge: Cambridge University Press.

Crawford, P., Brown, B., Baker, C. et al. (2015). *Health Humanities*. London: Palgrave Macmillan.

Dolan, B. (2015). *Humanitas: Readings in the Development of the Medical Humanities*. University of California Medical Humanities Press.

Jones, T., Wear, D., and Friedman, L.D. (2014). *Health Humanities Reader*. New Brunswick and London: Rutgers University Press.

Marcum, J.A. (2008). *An Introductory Philosophy of Medicine: Humanising Modern Medicine*. Dordrecht: Springer.

Monrouxe, L.V. and Rees, C.E. (2017). *Healthcare Professionalism: Improving Practice through Reflections on Workplace Dilemmas*. Oxford: Wiley.

Montgomery, K. (2006). *How Doctors Think. Clinical Judgment and the Practice of Medicine*. Oxford: Oxford University Press.

Schleifer, R. and Vannatta, J. (2013). *The Chief Concern of Medicine: The Integration of the Medical Humanities and Narrative Knowledge into Medical Practices*. University of Michigan Press.

Whitehead, A., Woods, A., Atkinson, S. et al. (2016). *The Edinburgh Companion to the Critical Medical Humanities*. Edinburgh: Edinburgh University Press.

17 The Development of Professional Identity

Sylvia R. Cruess and Richard L. Cruess
The Centre for Medical Education, McGill University, Montreal, Quebec, Canada

KEY MESSAGES

- Although the issue has been addressed explicitly only recently in medical education, physicians have acquired a professional identity or identities throughout the ages.
- The acquisition of a professional identity is essential if an individual is to 'think, act, and feel' like a physician.
- Professional identity or identities are acquired through the process of socialisation within medicine's community of practice at the undergraduate and postgraduate levels and throughout practice.
- The factors that support or inhibit the acquisition of a professional identity are largely understood, as are the responses of learners to the process.
- It is time to establish the acquisition of a professional identity as a principal objective of medical education and to develop instructional theories and strategies to support this objective.

Introduction

All the world's a stage
And all the men and women merely players;
And one man in his time plays many parts,
His acts being seven ages.
William Shakespeare, *As You Like It*, Act II, Scene 7 [1].

We begin this chapter with the familiar quotation from Shakespeare, as did the Eriksons in a summary of their life's work on identity formation [2]. We do so to emphasise that it has long been known that humans pass through recognisable developmental stages throughout their lives, during which their unique identity or identities are strongly linked with the roles that they play. Understanding this has relevance to those involved in medical education for two reasons. In the first place, the educational process, including the education of professionals, is both superimposed upon normal identity development and has a profound impact on it [3]. It is therefore necessary that we understand this. Second, we, along with Merton, believe that the dual objective of medical education has always been to provide future physicians with the necessary knowledge and skills for practice, but of equal importance, with a professional identity so that they come to 'think, act, and feel like physicians' [4]. While Shakespeare as a dramatist may have believed that those he described were playing their roles, it is the hope of those involved in medical education that learners at all levels will cease having to play the role of physician because that role will have become who they are [5] – they will have developed a professional identity.

The objective of this chapter is to address the issue of the formation of a professional identity in physicians. In doing so, it is important to stress that the current emphasis on identity formation has grown out of what has been termed the 'professionalism movement' [6] in medical education. It thus represents an evolution in our understanding [7, 8] of how best to produce individuals who 'think, act, and feel like physicians' [4]. This allows us to build upon what has been learned in the past two decades as virtually every medical school and postgraduate programme moved to teach and assess professionalism [7, 8]. As we directly address identity formation, it will become apparent that as individuals wish to join medicine's community of practice, they acquire the identity expected of members of the community and accept its norms [9]. The nature of these norms is determined by society and the medical profession and is encompassed in the word 'professional'. Thus in considering professional identity formation, it becomes necessary to understand what it means to be a professional, an issue that has been clarified for medicine through the development of programmes on teaching and assessing professionalism. Furthermore, many of the educational strategies created to ensure that learners understand professionalism and internalise its values can be reoriented to support professional identity formation [10–12].

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

Profession, Professionalism, and Professional Identity

Explicitly addressing the issues of professionalism and professional identity formation in the medical curriculum is a recent phenomenon [7, 8, 10–12]. This is somewhat surprising as the term 'profession' has been in use in medicine since at least Roman times and there are frequent references through the ages to the importance of acting like a professional [13]. Historically the characteristics of the 'good doctor' have been linked with the word professional [14].

While early social scientists studied the professions [15], there appears to have been little interest within the medical profession in outlining the nature of professionalism until recent times when medicine's professional status was threatened by the development of contemporary health care systems and the perception that many physicians were exhibiting unprofessional behaviour [16–18]. One response of the profession was to emphasise the teaching of medical professionalism throughout the continuum of medical education [19, 20]. This necessitated defining what was to be taught, leading to the emergence of several accepted definitions. While there are frequent statements that profession and professionalism are difficult to define, the core content of most definitions is remarkably uniform [21]. Because of its origins in the act of 'professing', we have chosen to define and use the word profession for teaching purposes. It is also the root of the terms professional and professionalism. The following has served as the basis for teaching professionalism and professional identity at our own university for over two decades:

> An occupation whose core element is work based upon the mastery of a complex body of knowledge and skills. It is a vocation in which knowledge of some department of science or learning or the practice of an art founded upon it is used in the service of others. Its members are governed by codes of ethics and profess a commitment to competence, integrity and morality, altruism, and the promotion of the public good within their domain. These commitments form the basis of a social contract between a profession and society, which in return grants the profession a monopoly over the use of its knowledge base, the right to considerable autonomy in practice, and the privilege of self-regulation. Professions and their members are accountable to those served, to the profession, and to society [22].

A professional is a member of a profession and is expected to demonstrate professional behaviour in his or her daily activities. Professionalism as defined by the Royal College of Physicians of London is 'a set of values, behaviours and relationships that underpins the trust the public has in doctors' [23].

Professionalism is a social construct and therefore the nature of the professionalism, and hence the professional identity of physicians, in any country or culture will be congruent with the national and cultural values of the society served [24]. The early literature on professionalism reflected the cultural values of western society and of the Anglo-Saxon world with its roots in Judeo-Christian morality [14]. With the increased mobility of individuals and ideas, it has become apparent that, while some core values such as competence, caring, compassion, honesty, and integrity appear to be universal, other aspects of professionalism can vary significantly in, for example, Asian [25], or Muslim societies [26]. This influences the structure and organisation of health care, societal expectations of the profession [27, 28] and the nature of the professional identity of physicians serving in these cultures [24].

Throughout the latter part of the nineteenth and into the mid-twentieth century the medical profession enjoyed a dominant role in health care and society [16, 17, 27]. It was trusted by both individual patients and society, based on the assumption that the profession would be altruistic [17]. Its autonomy was unquestioned, as was the principle of self-regulation [18]. Medicine exerted hegemony over other health care professions and enjoyed significant influence over health policy [16]. The economic burden of health care on society was not great as it consumed a small percentage of the gross domestic product of developed countries. After World War II this situation changed dramatically. Modern biomedical science transformed health care, making access to it essential to the well-being of individual citizens and society [27]. In most countries, national health services were established, while in the United States the marketplace was allocated significant responsibility for the delivery of health care [16, 29]. Costs escalated and providing health care became a lucrative enterprise for physicians, other health care professions, and a growing health care industry. Medicine's dominance, while present, was greatly diminished throughout the world as the state or corporate sector assumed responsibility for payment [16, 18]. The financial opportunities available to physicians led to increased opportunities for conflicts of interest and instances of unprofessional behaviour were well-documented, leading to the opinion that self-regulation was lax [16, 18]. Many came to agree with Shaw's description of the professions as constituting a 'conspiracy against the laity' [30]. Both society and the leadership of the medical profession realised that action was required and one of medicine's responses was to emphasise the teaching of professionalism to learners in hopes that, if they understood the nature of professionalism and the obligations necessary to sustain it, they would behave professionally [19, 20].

Teaching Medical Professionalism

This represented a major change. Professionalism had not been taught explicitly as individuals were expected to become professional by patterning their behaviour on that of respected role models [19]. While little documentation is available, the system did appear to function reasonably well, as it was based in part on the shared values of a relatively homogeneous medical profession serving a homogeneous population. This situation no longer exists in our multicultural world [24–26]. A consensus grew that professionalism must be taught explicitly and this became an obligation as accrediting and certifying bodies at the undergraduate and postgraduate level required that professionalism be taught and assessed [31–34].

In spite of the presence of multiple definitions [21] and differing pedagogic approaches to teaching professionalism [31], agreement emerged on some overarching principles [20, 35].

First, professionalism should be explicitly addressed throughout the continuum of medical education, with the strategies and material selected being appropriate for the stage of the learner.

Second, there is a cognitive base that serves as the basis of teaching and assessment, consisting of definitions of profession and professionalism and the importance of professionalism as the basis of medicine's relationship with society – its social contract [27, 28].

Third, the attributes expected of a physician must be communicated to learners [36]. We have chosen to separate the role of the healer and the professional for pedagogic purposes, understanding that they must be served simultaneously. While healers have existed in society since before recorded history, the modern professional only appeared late in the nineteenth century. This approach allows one to identify the attributes of the healer and the professional. Box 17.1, based on the literature, documents those attributes which describe the societal expectations of medicine under the social contract and are the foundation of the norms of the professional identity of a physician [7].

Fourth, it is not sufficient to merely communicate the cognitive base. Learners must internalise the value systems of the medical profession to 'inculcate the broader dimensions of competence and perspective' [37]. This process depends on role models and mentors [38] and is facilitated by reflection [39].

Fifth, faculty development, while important in its own right, becomes essential in the teaching of information whose exact nature is relatively unfamiliar to the faculty [40].

Finally, if professionalism is to be taught, methods to assess the professionalism of learners are required [41].

Teaching and assessing the professionalism of learners became an aspirational goal in medicine that, while difficult to achieve, led to the development of coherent approaches. The move to emphasise professional identity formation grew out of this approach, as some of the limitations of 'teaching professionalism' became apparent.

From the time that teaching professionalism explicitly was proposed, an existential question was always present: 'can professionalism be taught?' [42]. Or, as Hafferty succinctly asked, does medical practice require 'a professional presence that is best grounded in what one is rather than

BOX 17.1 Attributes of the healer and the professional

Attributes of the healer

- *Caring and compassion*: a sympathetic consciousness of another's distress together with a desire to alleviate it
- *Insight*: self-awareness; the ability to recognise and understand the patient's and one's actions, motivations, and emotions
- *Openness*: willingness to hear, accept, and deal with the views of others without reserve or pretence
- *Respect for the healing function*: the ability to recognise, elicit, and foster the power to heal inherent in each patient
- *Respect for patient dignity and autonomy*: the commitment to respect and ensure subjective wellbeing and sense of worth in others and recognise the patient's personal freedom of choice and right to participate fully in his/her care
- *Presence:* to be fully present for a patient without distraction and to fully support and accompany the patient throughout care

Attributes of both the healer and the professional

- *Competence*: to master and keep current the knowledge and skills relevant to medical practice
- *Commitment*: being obligated or emotionally impelled to act in the best interest of the patient; a pledge given by way of the Hippocratic Oath or its modern equivalent
- *Confidentiality*: to not divulge patient information without just cause
- *Autonomy*: the physician's freedom to make independent decisions in the best interest of the patients and for the good of society
- *Altruism*: the unselfish regard for, or devotion to, the welfare of others; placing the needs of the patient before one's self-interest
- *Integrity and honesty*: firm adherence to a code of moral values; incorruptibility
- *Morality and ethical conduct*: to act for the public good; conformity to the ideals of right human conduct in dealings with patients, colleagues, and society
- *Trustworthiness*: worthy of trust, reliable

Attributes of the professional

- *Responsibility to the profession*: the commitment to maintain the integrity of the moral and collegial nature of the profession and to be accountable for one's conduct to the profession
- *Self-regulation*: the privilege of setting standards; being accountable for one's actions and conduct in medical practice and for the conduct of one's colleagues
- *Responsibility to society*: the obligation to use one's expertise for, and to be accountable to, society for those actions, both personal and of the profession, which relate to the public good
- *Teamwork*: the ability to recognise and respect the expertise of others and work with them in the patient's best interest

what one does?' [43], shifting the emphasis from 'doing' to 'being'. Haidet also wondered whether professionalism had become like a white coat that can be donned or discarded at will [44]. Against this background, the concept of professional identity emerged, first as educators within medicine described the nature of professional identity, and subsequently as an educational objective.

From Teaching Professionalism to Supporting Professional Identity Formation

Understanding gradually emerged that, as had been proposed earlier by Merton [4], Becker [45], and Bosk [46], medical students and residents, during the course of their educational experiences, come to acquire the identity or identities of a physician [3, 7, 11, 47–59]. This new understanding was based upon the work of a small group of observers who applied the rich literature on identity formation, found largely in developmental psychology, to learners and practitioners in medicine. The concept did not have a significant impact until the Carnegie Foundation reports on the future of professional education recommended that identity formation become a foundational element of the education of all professions [37], including medicine [10]. This required a reassessment of the movement to teach professionalism. It became apparent to many that one of the ultimate objectives of medical education had always been to support individuals as they develop their professional identities, making the teaching of professionalism a means to an end rather than an end in itself [7].

Shifting the focus of medical education from teaching professionalism to supporting individuals as they develop their own professional identities has advantages. If the objective of teaching professionalism is to assist in the development of professional identities, the educational strategies derived from this approach can address the objective of professional identity formation directly. In addition, there is a fundamental shift in emphasis from professors teaching and students learning to one whose aim is engaging learners in the development of their own unique identities and supporting them through the process [7, 11]. However, it does lead medical educators into less well-charted waters. The nature of identities in general and medical professional identities must be understood. Since identities are formed through the process of socialisation, an understanding of this process is required. Socio-cultural theories of learning, such as those of Lave and Wenger [9] concerning 'participation' and 'communities of practice' are particularly relevant here and are discussed in detail in Chapters 2 and 12 of this book.

Some of the 'teaching principles' listed above remain relevant. The cognitive base, in addition to a knowledge of professions and professionalism, remains an essential element but now should contain explicit reference to identity formation, socialisation, and communities of practice [60]. Because self-perception is fundamental to one's identity, reflection, guided by mentors and role models, remains fundamental [55]. Faculty development becomes more important, as identity formation and socialisation are largely unfamiliar to most clinical teachers [61]. Finally, while assessment of professionalism is difficult, assessing progress towards acquiring a professional identity poses problems of a different magnitude [61, 62].

Box 17.2 summarises the evidence on professional identity formation in medicine.

Personal and Professional Identity

Personal Identity

It is important to realise that the development of a professional identity takes place within the context of individual identity formation, a process that, as Shakespeare noted, begins at birth and continues throughout life [1]. Therefore, knowledge of personal identity formation is necessary if professional identity formation is to be placed in its proper context. Shoemaker [63] defines identity in generic terms. 'A set of traits, capacities, and attitudes that an individual normally retains over a considerable period of time that distinguishes that individual from others and represents the individual's conception of self and is recognized as such by others.' Personal identity is based on an individual's concept of who they are and how they are perceived by others.

Our understanding of identity formation is built upon a theoretical framework whose roots lie in psychoanalysis [64]. During the past half-century, Piaget [64], Erikson [2], Marcia [65], Kohlberg [66], Kegan [67], and others have been major contributors to our understanding of the various developmental stages through which each individual passes. Piaget established the fact that development proceeds in stages [64], preparing the ground for subsequent advances. Erikson's pioneering studies detailed eight developmental stages, stressing the role of crises in development, characterising adolescence as a time of 'role confusion' [2]. Erikson's work is relevant to the education of physicians as they enter medical school during or immediately after adolescence and may still exhibit considerable 'role confusion' [68].

BOX 17.2 WHERE'S THE EVIDENCE: Professional identity formation in medicine

- It is well established in developmental psychology that every human being develops a personal identity in stages
- It has been recognised for half a century that learners in medicine develop the identity of a physician and that this also occurs in stages
- A substantial body of literature exists in medicine outlining the nature of the identity of physicians
- This literature also documents many of the factors that can either promote or inhibit the development of a professional identity
- While still difficult, assessment of progress towards the acquisition of a professional identity is possible

Building on Erikson's ideas, Marcia traced a path beginning in adolescence with a 'diffuse' and poorly developed identity that, if identity formation is successful, results in 'identity achievement' during the post-adolescent period [65]. Some individuals fail to progress to identity achievement as they declare a moratorium, deferring decisions and action. Furthermore, he postulated that identity achievement will not occur if premature 'foreclosure' occurs, as an individual conforms to traditional norms rather than developing their own value system. Marcia's observations also provide some insight for medical education [68].

Kohlberg documented the interdependency of moral development and identity formation, believing that individuals move from wishing to be perceived as good by conforming to socially acceptable norms of right and wrong to developing an internalised set of moral principles, basing their behaviour on strong personal beliefs [66].

The formative elements impacting on identity are a classic mix of nature and nurture. Genetic inheritance contributes significantly [2, 43]; however, a profound influence is exerted by life experiences, including the multiple social interactions in which every individual engages in their respective communities [67, 69]. Psychological theory proposes that these forces impact on each individual as they journey through life, attempting to organise their experiences into a meaningful whole that incorporates their personal, private, public, and professional selves [2, 43, 49, 55, 57, 67]. As they pass through each stage, from infancy to childhood and beyond, individuals gain experience and become capable of constructing more complex personae [43, 64–67].

The theoretical approach to identity formation suggests three domains, all relevant to medical education, through which identity is influenced and developed: individual, relational, and collective [69]. While the influences impacting upon these domains may be somewhat independent of each other, they clearly are related, as are the identity or identities resulting from their influence.

- The *individual* domain includes genetic composition, self-chosen or mandated commitments, beliefs about oneself, and the impact of personal life experiences.
- The *relational* domain expresses the influence on the identity of significant individuals, such as family members, friends, and co-workers, including role models and mentors.
- The *collective* domain reflects the impact of the social groups to which an individual belongs or wishes to join. Individual status within the group and the group's status within society are important contributors to one's identity [70, 71].

A well-developed identity provides a sense of continuity, uniqueness, and belonging.

Some aspects of identity remain relatively stable throughout life, while others change as an individual passes through various developmental stages. Their individual, relational, and collective relationships are altered. Some changes are conscious, while others are more 'automatic and implicit' [69]. While identity stabilises in early adulthood, transformation continues throughout life, with an enduring core of 'self' being ever present [2, 57, 69]. While most changes are gradual, it must be recognised that seminal events, such as a major personal event, a change of career, or a religious conversion, represent significant alterations in an individual's life and can lead to a rapid transformation in the sense of self [63].

While the term 'identity' is frequently used in the singular, every individual has multiple identities that are context dependent [57, 69]. A person can be a son or daughter, married or single, a member of an ethnic, religious, or national grouping, and many others, in addition to being a physician. Furthermore, the physician can have a professional identity that reflects the entire medical profession as well as a specialty and be further identified as a practitioner, researcher, teacher, or any distinct role. Each identity is associated with a community and the individual will share traits and values with others in each community [9, 72]. In most instances, emergence of an identity takes place at an unconscious level, depending on need and place [69, 70].

Professional Identity

Individuals usually enter medical school in late adolescence or early adulthood with identities that have been developing since birth and that would have continued to develop in other directions had they not chosen medicine as a career. Superimposed on this trajectory is the development of a physician's identity: 'a representation of self, achieved in stages over time during which the characteristics, values, and norms of the medical profession are internalized, resulting in an individual thinking, acting, and feeling like a physician' [7].

Robert Kegan [67], building upon the work of those who came before, has had a significant impact on our understanding of identity formation in the education of professionals [3, 7, 43]. He proposed a framework for the longitudinal development of the self into a moral and meaning-making entity [8]. His classification consists of six stages (0–5), beginning in childhood and extending into adult life, which he termed:

0 incorporative
1 impulsive
2 imperial
3 interpersonal
4 institutional
5 inter-individual

Kegan's early and final stages are not pertinent to the development of a young adult such as a medical student or resident but Box 17.3 summarises stages 2–4 as they apply to professional identity formation in medicine. In Kegan's *imperial* stage, learners take on a professional role but it is not fully integrated into their identity. They may act like a professional but are still searching. In the *interpersonal* stage individuals begin to identify with the profession, to the point where they become immersed in and integrated with it, as the norms of the profession take hold. Those who reach the *institutional* stage are characterised as the self-defining professional. They can negotiate conflicts between professional values and their core beliefs and criticise or challenge aspects of the profession. Their reason is in control of their emotions and desires. Deep authentic and unshakable incorporation occurs, with professional identity

BOX 17.3 Stages of personal and professional identity formation

Stage	Personal characteristics	Manifestation in a professional context
Imperial	An individual who takes into account the views of others but whose own needs and interests predominate.	An individual who can assume professional roles but is primarily motivated to follow rules and to be correct; self-reflection is low. Emotions can overwhelm reason.
Interpersonal	An individual who is able to view multiple perspectives simultaneously and subordinate self-interest; who is concerned about how she or he is perceived by others.	An individual who can assume professional roles and is oriented towards sharing obligations; tends to seek out those to emulate; is idealistic and self-reflective. Emotions are generally under control, and she or he generally does the right thing.
Institutional	An individual who can assume a role and enter into relationships while assessing them in terms of self-authored principles and standards; the self is defined independently of others.	An individual who is able to understand relationships in terms of different values and expectations. The external values of the professional become internal values. Reason is in full control over needs, desires, and passion.

Source: Adapted from Kegan [67]; reproduced with the permission of *Academic Medicine*.

and the other enduring identities defining the self. Those who transition to the *inter-individual* stage do not perceive themselves as having a single identity and are open to, and accepting of, many other value systems. While it would be comforting to think that every incoming medical student will achieve Stage 5, research indicates that few of us actually progress beyond Stage 4 [43].

Bebeau, a respected investigator influenced by Kegan, summarises this transition in this way: 'Individuals move from self-centered conceptions of identity through a number of transitions, to a moral identity characterized by the expectations of a profession to put the interests of others before the self, or to subvert one's own ambitions to the service of society' [48].

The Formation of a Professional Identity

Since Lave and Wenger [9] developed their theory of situated learning (see Chapters 2 and 12 in this book) it has been difficult to discuss identity formation without alluding to 'communities of practice' as, from a conceptual point of view, the two subjects are intertwined. Medicine is identifiable as a distinct community of practice [70] and individuals wishing to become physicians voluntarily join the community by learning to carry out the day-to-day activities of the physician [70, 71, 73–75]. In the process, each individual moves from legitimate peripheral participation in medicine's community to full membership. In so doing, he or she acquires the identity of a physician, accepting the norms, values, and structural organisation of the medical profession. Early membership is termed 'legitimate' because the individual has been accepted as a novice as they enter medical school. Full membership requires a demonstration of competence within the domain, with the standards being set and assessed by the community. Learning is a social activity, depending on interactions with fellow students, physicians, and other health care professionals within the community. Much of it occurs at the unconscious level, resulting in the acquisition of a large body of both explicit and tacit knowledge. The learning is 'situated' within medicine's community, giving its content authenticity as it is acquired in the same context in which it is applied [76]. Medicine's community is dynamic, with individuals moving from peripheral to full participation and to senior positions, followed by retirement, thus making room for new members. Because each individual must re-create medicine's knowledge base as they acquire it, the knowledge base is also dynamic and ever-changing as novices challenge existing practices.

Rather than a single community of practice, medicine offers opportunities to belong to several groupings. The Wenger-Trayners describe the medical profession as consisting of 'a complex landscape of differing communities of practice – involved not only in practising the occupation, but also in research, teaching, management, regulation, associations, and other relevant dimensions' [70]. These communities can be examined at the macro, meso, and micro levels. At the *macro* level, there is the healing profession with its roots in antiquity and emphasis on well-recognised universal values such as competence, caring, compassion, confidentiality, honesty, and integrity – the 'good doctor' [14, 77]. At the *meso* level are the medical specialties whose collective impact on identity formation during postgraduate training and in later practice is extraordinarily strong [78]. The *micro* level encompasses many small communities such as hospital or university departments, research units, teaching units, and the many activities in which individual learners and physicians engage with other physicians and health care professionals [70]. Thus, physicians can simultaneously possess multiple professional identities whose expression depends on the

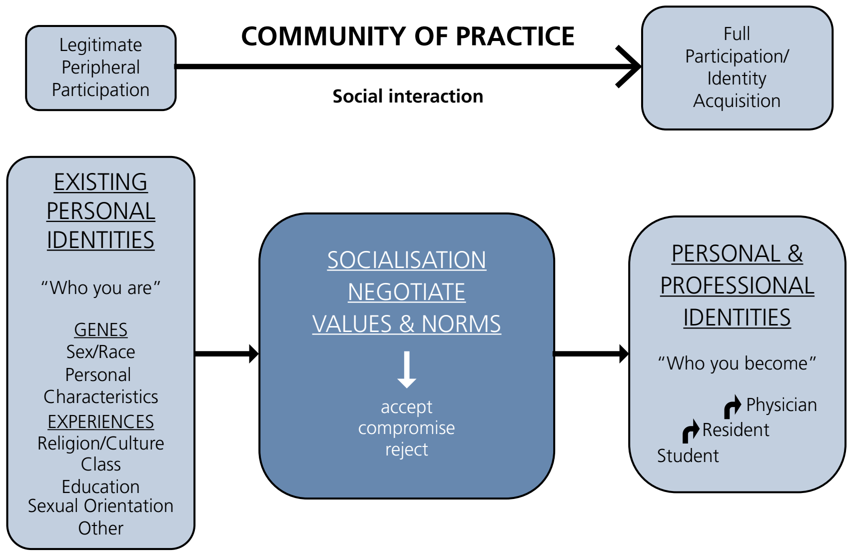


Figure 17.1 The role of socialisation in identity formation. *Source*: Reproduced with the permission of *Academic Medicine*.

context. Furthermore, during their professional lives they may discard identities as they acquire new ones [63].

Figure 17.1 offers a schematic representation of the process through which professional identities are formed. Its upper portion shows the movement from legitimate peripheral participation in medicine's community or communities of practice to the acquisition of a professional identity and full membership [79]. The lower half of the figure links this to the process of socialisation through which professional identities are formed [54, 63, 71]. Socialisation is 'the process by which a person learns to function within a particular society or group by internalizing its values and norms' [80]. It involves a personal transformation, thus differentiating it from training. Hafferty stated: 'while any occupational training involves learning new knowledge and skills, it is the melding of knowledge and skills with an altered sense of self that differentiates socialization from training' [43].

Incoming medical students arrive as adolescents or young adults, having been socialised from birth. While some role 'confusion' [2] or 'diffusion' [65] may be present, the core of an individual identity is present. The nature of this identity depends in part upon the genetic makeup of the individual that determines certain physical characteristics, including gender and race. However, each individual's personal experiences have had a profound effect, with culture, religion, class, education, and a host of other factors contributing to the sense of self of each medical student [69, 81].

Each learner voluntarily embarks upon the journey from peripheral to full participation, and is exposed to the norms of the community that they wish to join – in this case medicine. In accepting and internalising these norms they will, both consciously and unconsciously, acquire in stages the identity expected of a physician [45, 46, 69, 81]. Medical students have a distinct identity, with a major transformation occurring as they have increasing contact with patients [3, 43–50]. As they progress to postgraduate training, they acquire the identity of a postgraduate trainee or resident. Although data supporting this statement is sparse, most believe that the impact of postgraduate training is extremely strong. Most individuals regard their specialty affiliation as leading to their strongest professional identity [78]. Learners enter practice, where they may belong, as noted above, to multiple communities [70]. It is also important to understand that during their professional lifetimes, physicians' identities may undergo several major transformations depending on their personal and professional situations [63, 70].

Each individual must respond to the pressures on their personal identities resulting from the process of socialisation [3, 43, 67]. A major part of this response follows exposure to and coming to terms with the norms of medicine's community [43]. There is general agreement that this involves negotiation with both the 'self' and the community at both the conscious and unconscious level [68, 81]. Each learner can accept the norms, attempt to negotiate so that some norms are altered, or can reject selected norms. Some norms are so central to the identity of the profession that they are non-negotiable [43]. Caring, compassion, honesty, integrity, altruism, commitment, and competence are examples. If a learner is felt to be rejecting essential norms, consequences will occur, including sanctions or exclusion from the community [43].

Progress towards full participation in the community and the acquisition of a professional identity is not linear [68, 69, 81]. There are times of real movement, often linked to periods of transition (such as the first contact with a cadaver or death), the beginning of significant patient contact, or movement from undergraduate to postgraduate education [43, 50]. Each major transitional period offers opportunity for progress, often associated with significant levels of stress as major changes in identity are entailed.

Internalising the norms and behaviours impacts an individual's existing identity. Changing one's identity is difficult and is inevitably associated with significant degrees of discomfort and uncertainty. Erikson felt that it was necessary to 'suppress' one's existing identity in order to effect change [2] and Monrouxe believes that change can lead to 'identity dissonance' [50], with increased levels of stress. There is constant tension between the need to maintain the central core of 'who one is' as each individual is obliged to come to terms with both the norms of the community and the reality of 'who one wishes to become' [59]. In addition, experiences such as personal failures or a hostile learning environment, that makes the community seem unwelcome, can retard progress [68, 81]. If the identity of the learner is congruent with the identity of the community, dissonance can be minimised. Thus selecting individuals with identities similar to those of a physician can ease the transition from layperson to professional [82].

The norms of a professional identity in medicine that must be made explicit to learners [36] are encompassed in the definitions of professionalism, including the obligations of individual students, postgraduate trainees, and physicians to their patients and to society. In addition, each individual student, postgraduate trainee, and physician has obligations to other members of the community and to the community itself [7, 60]. Finally, medicine's community of practice enjoys a social contract with society, the details of which must be understood by all [27] and the attributes of 'the healer' outlined in Box 17.1 delineate the norms expected of members of medicine's community of practice.

Factors Impacting on Socialisation

The imposition of medical education upon the normal process of identity development exposes individuals to a series of factors that directly impact this process. Figure 17.2 attempts to summarise these factors.

Role Models, Mentors, and Experiential Learning

The most powerful forces impacting on professional identity formation in medicine are role models and mentors [38, 50, 51, 69], and the multiple personal and group interactions that occur in the course of both clinical and nonclinical experiences in medicine's multiple learning environments [43, 50, 51]. These social mechanisms lead to the acquisition of knowledge and skills, and have a profound impact on the attitudes and values of learners at all stages. They exert their influence through both consciously reflective and unconscious pathways [83, 84], that lead to the presence of a large body of explicit [85] and tacit knowledge [86].

Role models are 'individuals admired for their ways of being and acting as professionals' [87], while *mentors* have closer and more prolonged contact with learners, serving as 'experienced and trusted counsellors' [80]. Both, as members of the community, provide a template for the identity to which learners aspire [88]. Becoming like them in action, appearance, and beliefs becomes a goal for learners that is both conscious and unconscious. Their impact can be, depending upon their actions and behaviour, either positive or negative [88]. Role modelling of unprofessional behaviour is unfortunately common, and, by leading to role confusion, can impede professional identity formation [52, 81, 88, 89].

Of equal importance is experiential learning – the end result of the multiple clinical and nonclinical experiences to which learners are exposed on a daily basis [43, 85]. However, merely exposing learners to experiences is no longer deemed sufficient. For experiences to have a maximum impact upon identity formation, reflection on the experiences is necessary [82, 83]. Kolb tells us that following an experience, observation and conscious reflection

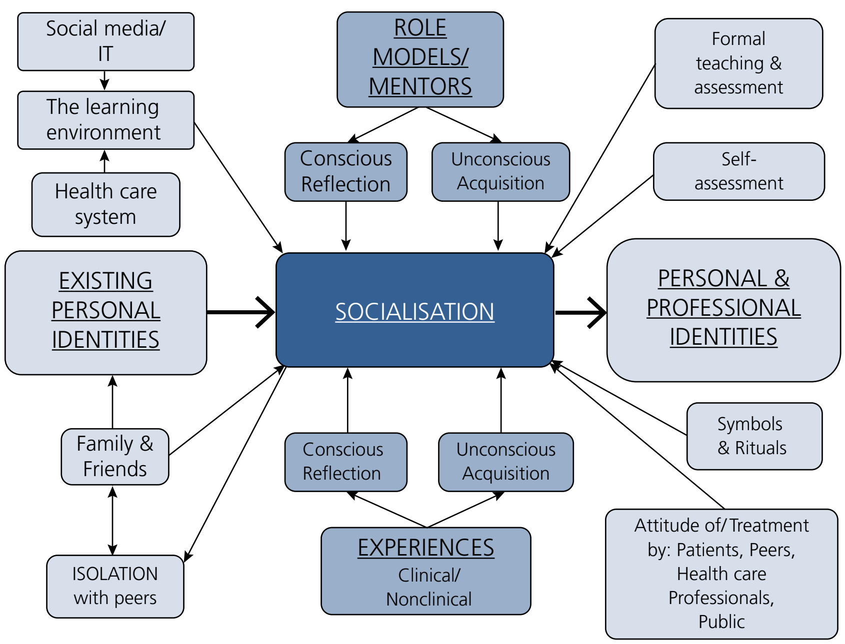


Figure 17.2 Factors impacting on socialisation. *Source*: Reproduced with the permission of *Academic Medicine*.

must take place, which will lead to the assimilation of the experience into an individual's beliefs, followed by actions based upon the changed knowledge base [90]. To ensure that reflective exercises are directed towards professional identity formation, they should be guided by knowledgeable individuals, primarily role models and mentors, and opportunities provided in the curriculum for reflection.

Since the time of Dewey we have known that not all experiences are appropriate or result in learning devoted to a specific educational objective [91]. It is necessary to ensure that learners are exposed to and reflect on a broad variety of experiences specific to the development of their professional identity. Of all of the experiences to which medical students are exposed, their contact with patients has the greatest impact on their professional identity formation [3, 50, 55, 56]. Early and ongoing clinical contact followed by reflection on these experiences is central [84, 85]. The medical curriculum has an obligation to expose students to the wide variety of clinical issues that they will face, but to engage them in the process of their own identity development. Points of transition or tension offer an opportunity to both advance identity formation and, through open discussion, diminish the stress resulting from identity changes [3, 59]. Entry into medical school, exposure to cadavers and experience with death, transitions from classroom to clinical experiences and on to residency, are all examples of such opportunities [59, 81].

A final point is important. Medicine's knowledge base consists of both explicit and tacit knowledge, and it is believed that the vast majority is tacit [92]. The volume of tacit knowledge is so great that it cannot all be made explicit [91, 93]. Medical educators must select the items of knowledge to be made explicit for individuals to be consciously engaged in altering who they are [94]. There is also agreement that knowledge is constantly being exchanged between the tacit and explicit categories [95]. Much of the tacit knowledge is acquired through unconscious routes, with the individual being unaware of its presence [95]. It is knowledge that 'one knows but cannot tell' [96]. As an example, as students join medicine's community of practice, they often unconsciously change their behaviour, their dress, and their ways of communicating [68]. The importance of this route is such that faculty must be aware of the fact that they are role models at all times, and not just during time devoted to formal teaching [88]. Tacit knowledge is constantly being acquired. Informal learning and the complexities of work-based learning are discussed in greater depth in Chapter 12 of this book.

The Teaching Environment

In addition to the influence of role models, mentors, and experience, Figure 17.2 also describes the multiple factors in the teaching environment that can impact upon the process of socialisation, outlining many of the elements of the formal, the informal, and the hidden curriculum [97].

The formal curriculum includes the elements that the faculty hopes that learners will deem important [96]. If professional identity formation becomes an educational objective, it must be included in the formal curriculum as a part of the cognitive base that is explicitly presented to learners at all levels [98]. As 'assessment drives learning', for this information to have an impact, assessment of the acquisition of this knowledge base must take place and multiple methods are readily available for this task [62]. However, proceeding to the next step of attempting to assess progress towards the acquisition of a professional identity is more difficult as practical methods are not yet available [62, 99].

Because identity involves the creation and presentation of the 'self', self-assessment, assisted by an informed faculty member, becomes both desirable and valuable. Accompanied by feedback, learners can be encouraged to examine their own progress from layperson to professional and to determine factors that are both encouraging and inhibiting this journey [59, 60, 99].

The learning environment, which includes both the informal and hidden curriculum, can have a profound impact on socialisation and identity formation [47, 49, 51, 97]. A welcoming community that recognises learners as future colleagues from the beginning of their educational experience is invaluable [3, 81]. Institutional policies that honour and reward 'the good doctor' contribute to this sense of community [97]. On the other hand, a corrosive learning environment that demeans or humiliates learners will inhibit proper identity formation [47, 49, 81].

The nature of the health care system has an impact on this learning environment as learning is 'situated' in elements of the system and learners are preparing themselves to function within it [100, 101]. While it is not feasible to attempt to tailor health care systems so that they support identity formation, reflection on the impact of health care systems on each individual's identity can provide insight to learners on how best to cope with complex systems while remaining 'who they wish to become' [79].

Social media is a factor that has emerged in recent times whose impact on professional identity formation is as yet unclear For recent generations, social media has become a natural way of communicating with colleagues, friends, family, and indeed the world [102]; even though its use frequently violates professional norms [103], it now constitutes a commonly used means of projecting who individuals believe themselves to be. While evidence is sparse, it is probable that social media will have an impact on the professional identity of individual learners because of the large amount of information in circulation that is of relevance to that individual. It also represents a powerful means through which others can indicate how individuals are perceived. As such, social media can have a positive or negative effect on identity formation.

Other Factors

The remaining portions of Figure 17.2 outline other factors that impact on socialisation. How medical students are treated by a wide variety of individuals has a profound impact on how they regard themselves [3, 43, 45, 50]. The most important influence is exerted by patients, followed by peers and other health care professionals, family, and friends [53, 55, 56, 104]. As learners are treated as either doctors in training or as doctors, they come to feel like doctors. As pointed out by Goffman, as they play the role, the role becomes who they are [5].

Symbols and rituals also have a long and important history in medical education and impact identity formation [55, 57, 104–106]. Symbolic events such as acquiring a stethoscope, 'white coat ceremonies', and reciting the Hippocratic Oath all contribute to a sense of belonging in medicine's community of practice [55, 57, 105].

Finally, because of the demanding and time-consuming nature of medical education, isolation in the learning environment with like-minded individuals has been a fact of life for generations [43, 49, 51]. This isolation from one's previous environment including family and friends, enhances the impact of the multiple factors influencing socialisation [43, 68] but is currently a major source of tension. The current generation of students, postgraduate trainees, and young practitioners strongly object to the impact of isolation on their private lives [107, 108]. As they negotiated the norms of practice within the community, they have actually changed the norms, with the objective of establishing a proper work/life balance. The result has been a restriction in work hours and significant changes in patterns of practice [108].

In concluding this section it must be emphasised that while the factors listed impact on each learner, the response of each individual to each factor will not be identical [3, 51, 59]. Thus, isolation may enhance identity formation in some, while causing stress in others. Self-assessment may be easy for one individual and either difficult or ineffective in others. However, the sum total of these factors results in individuals who become socialised to acquire a changed identity composed of portions of who they were and who they have become.

The Response of Learners to Socialisation

The journey from layperson to professional entails a series of responses that, while unique to each individual, are reasonably well documented. These responses do not result from our current focus on identity formation, but have long been an inherent part of medical education. Figure 17.3 attempts to summarise them.

The wider literature on identity formation stresses the powerful formative impact of occupation on an individual's identity [72]. This is true of medicine as, with increasing levels of competence and hence confidence, a learner's sense of self as a physician becomes more secure [50, 52, 54]. This occurs more rapidly in clinical situations, hence the emphasis on early and continued clinical contact for medical students [54, 57, 95, 108]. A sense of competence leads to both satisfaction and pleasure, enhancing motivation [55, 57]. However, if competence is questioned, doubt may result in a loss of confidence accompanied by shame or guilt [109]. Progress towards the acquisition of a professional identity may be inhibited by the resulting stress that can lead to anxiety or fear [55, 57, 72]. The importance of positive feedback to avoid such a situation cannot be overemphasised.

Another constant in medical education is stress [110]. Whether the acquisition of a new identity actually requires suppression of an existing identity or if identity dissonance is the issue, all are agreed that some level of stress is inevitable [47, 53–55, 69, 72]. This stress is added to that derived from a demanding curriculum involving the acquisition of an enormous body of knowledge, the exposure to new experiences involving pain and suffering, exposure to many difficult situations, including death, and personal involvement with other human beings [54, 56, 68]. However, it is important to note that we have understood for over a century that stress, up to a certain level, is beneficial to performance and learning [43, 45, 47, 49, 110, 111]. The stress–performance curve tells us that to a certain level stress is beneficial but that, beyond a critical point, it becomes corrosive. Stress will never be eliminated from medical education but learners must be supported during stressful periods and attempts made to ensure that stress remains within the beneficial range.

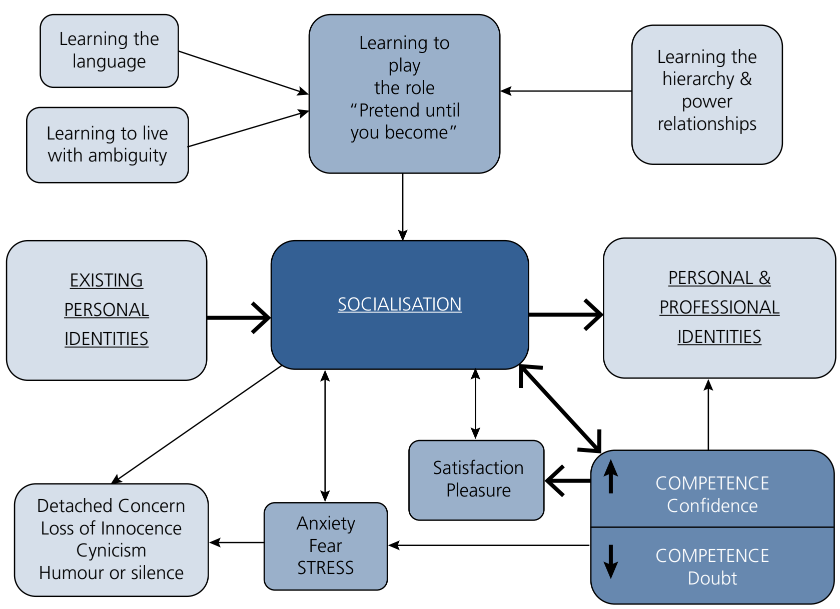


Figure 17.3 Learner responses to the process of socialisation. *Source*: Reproduced with the permission of *Academic Medicine*.

Another response that has been noted through the ages is the necessity for learners, when they are faced with a new role, to 'pretend' competence [5, 50–53]. In clinical situations their instructors, patients, peers, and friends and family expect them to comport themselves like physicians and they do so, even when they understand that they are pretending. Each change of role, from the classroom to the clinic, from the undergraduate and postgraduate level and into practice necessitates a repeat of this [50, 55, 58]. However, by playing the expected role over and over again it becomes incorporated into themselves and they come to 'think, act, and feel like physicians' [4].

Learners must acquire the language of medicine that has developed through centuries and learn to live with the ambiguity and uncertainty inherent in the practice of medicine [43, 45, 47, 49]. They are exposed to the hierarchy and power relationships of faculties of medicine, hospital departments and administrative units, regional health structures, and the wider health care systems [49, 89, 97, 112]. It is important that they understand the impact of these relationships on both their community and on themselves [112, 113].

Finally, the varying responses of medical students to the stressful experiences encountered in medical school have been well documented through the ages. Gallows humour or silence are normal human responses to stress and appear to have been present since formal instruction in medicine began [114, 115]. Loss of innocence, detached concern, or cynicism occur as learners come to understand that the reality of medical practice differs from early idealistic aspirations [89, 116]. This is accentuated if learners are exposed to unacceptable practices or behaviours that are, unfortunately, frequent experiences [47, 51, 53]. These situations pose moral dilemmas to learners, thus further contributing to the stress involved in identity transformation [113].

Implications for Medical Education

Using various educational theories and strategies, medical education throughout the ages has produced physicians who have acquired the knowledge and skills necessary for practice as well as a professional identity. Therefore, even if professional identity formation is made an explicit educational objective, it may not lead to a radical transformation of the curriculum. However, many strategies, used for generations, may be altered as they are examined through the lens of identity formation. In addition, some new activities may be developed to support learners as they develop their professional identities.

The following curricular recommendations, summarised in Box 17.4, are based upon the limited literature on identity formation found in medicine [3, 8, 11, 55, 57, 59, 81], our institutional experience [117, 118], and on the wider literature outside of medicine devoted to the educational aspects of identity formation, communities of practice, and workplace learning [119–123].

BOX 17.4 HOW TO: Support professional identity formation in the medical curriculum

1 Make professional identity formation an educational objective
2 Establish a cognitive base
3 Engage students in the development of their own identities
4 Create and maintain a welcoming community of practice
5 Explicitly address the major factors impacting on identity formation (e.g. role modelling, mentoring, experiential learning, reflection)
6 Recognise periods of transition as opportunities
7 Provide faculty development to support the programme
8 Chart progress towards the development of a professional identity
9 Identify learners who struggle and provide remediation when required

1 *Establish professional identity formation as a principal educational objective* [7]
We believe that identity formation has always been an implicit objective of medical education and that many educational strategies used in the past, including the teaching of professionalism, represented a means to an end [7]. Making professional identity formation an educational objective establishes it as a foundational element of medical education. This should be reflected in mission statements and other public documents outlining the objectives and values of educational institutions.

2 *Establish a cognitive base for the teaching of professionalism* [31]
The concepts of professional identity formation, communities of practice, and socialisation should be added to this cognitive base in the formal curriculum. The aim is to ensure that the process of professional identity formation is understood by teachers and learners so that all understand the nature of the journey undertaken by those wishing to become physicians, along with the various factors that can promote or inhibit individuals through the journey. The schematic representations of identity formation and socialisation provided in this chapter are designed to categorise and organise these factors.

3 *Engage students in the development of their own identities* [3, 10, 11, 55, 57, 59, 81]
Supporting learners as they develop their own identities as an educational objective shifts the emphasis from faculty teaching and students learning, to students consciously addressing the issue of 'who they wish to become'. Combined with knowledge on the nature of identity formation and socialisation, individuals can better understand their journey, trace its trajectory, and assume some control over the acquisition of a professional identity. Students should be encouraged to retain the core elements of their personal identities as they become physicians [59, 81].

4 *Create a welcoming community of practice*
Medicine has always been a community but is not always welcoming [47, 49, 54, 68, 112, 113]. The learning environment requires critical examination through the lens of identity formation. The acquisition of a professional identity by joining a community of practice is conceptually important because it dictates that incoming members, or peripheral participants, represent the future of the community [70]. It is therefore incumbent upon the community to be welcoming and supportive of their future colleagues. In spite of progress that has been made, access to medicine's community remains difficult for individuals from minority groups, a situation that must be addressed [47, 49]. The learning environment, including elements of the hidden and informal curricula, can positively influence identity formation by being explicitly supportive, open to negotiating the norms for each individual, and providing feedback [43, 49, 51, 97]. On the other hand, a corrosive learning environment that is hostile and unwelcoming can undermine confidence and inhibit identity formation [97, 113]. Social activities involving members of the community from junior to senior can have a powerful impact on creating a sense of belonging, as can the various rituals that have characterised the medical profession for generations [43, 50, 51, 57].

5 *Address the major factors impacting on identity formation*
Because role modelling and mentoring have always been, and remain, fundamental to the acquisition of a professional identity, they require special attention as the curriculum is revised [75, 83, 88]. Role models and mentors must be aware of the cognitive base and of the fact that they transmit explicit and tacit knowledge at all times, not just during formal teaching [98]. In addition, they must be aware of the role of experiential learning and reflection on the creation of professional identities [83, 84, 88, 90]. The curriculum must ensure exposure to experiences believed to be fundamental to the development of a professional identity and contain protected time for guided reflection on these experiences, understanding that group sessions are particularly beneficial [123].

6 *Recognise periods of transition as opportunities*
Identity formation does not proceed in a linear fashion throughout medical education [57, 68]. Times of transition to new roles are particularly stressful as they frequently require alteration of existing identities [81]. They present both challenges and opportunities, as they represent periods when individuals' identities undergo significant change. Entry into medical school, the beginning of full-time clinical rotations, transition to residency, and entry into practice represent major challenges [68]. These should be exploited as opportunities for each individual, guided by a role model or mentor, to reflect on their progress towards feeling like a physician. To an increasing degree, time is being made in the curriculum to accommodate transitional periods, and viewing these activities through the lens of identity formation has been found particularly beneficial.

7 *Provide faculty development to support the programme*
Faculty development is particularly important to support a curriculum devoted to identity formation [98]. The concepts of identity formation, communities of practice, and socialisation are not widely known to clinical educators and, if they are to actively and explicitly support students in their journey, they must understand these processes and their roles. Faculty development provides knowledge and guidance to role models and clinical educators as well as causing them to reflect on their own identities [117].

8 *Chart progress towards the development of a professional identity*
The standards of the community of practice of medicine are established by collaboration with society. The community is responsible for determining whether aspiring members meet those standards [70, 72, 75]. As the acquisition of the identity of the community is indivisibly linked with membership, assessing progress towards acquiring a professional identity becomes necessary. Some methods are available to assess the status of a professional identity but are not yet feasible on a large scale [61, 99]. Because the presentation of self is so central to professional identity [70, 81], self-assessment is important. An individual's assessment of their own progress towards 'feeling like a doctor', carried out in collaboration with a mentor, is an important stimulus for formative feedback and feasible methods of doing so are now available [61]. Positive feedback about progress is an important factor in the development of a professional identity, engendering a feeling of confidence and a sense of belonging [3, 68, 81]. Summative feedback is difficult. However, the assessment of professional behaviours can be regarded as a surrogate for assessing professional identity and unprofessional behaviour will always be with us [61]. Therefore, methods developed to assess professional behaviours can indirectly measure the status of professional identity, and serve as the basis of further action, including remediation or removal [41].

9 *Provide remediation where required*
Methods of remediation as seen through the lens of professional identity have only recently been considered [61]. They 'apply social science theories about social and psychological development in adult conceptualizing and implementing remediation programs' [124]. They look upon remediation as life-long continuous quality improvement and are heavily dependent upon involving the individual in establishing and meeting aspirational goals.

Conclusion

This chapter is based upon the premise that professional identity formation should be 'the central focus in educating tomorrow's physicians' [125]. Professional identity formation as an educational movement evolved as it became apparent that an approach to improving physicians' professional performance based on teaching professionalism, while a step forward, had inherent contradictions that were difficult to overcome. A feeling developed that producing graduates who merely acted like physicians somehow lacked authenticity. The behaviour of the 'good physician' should spring from within, being based on 'who they are'. As educators and investigators in medicine provided more information about identity formation in general and professional identity formation, it became possible to envision a curriculum that would be devoted to supporting individuals as they develop their own identities. A theoretical base for the educational activities in such a curriculum is provided by social learning theories such as communities of practice and situated learning in addition to the well-developed theoretical basis of identity formation. It is our hope that this chapter can point the way for the development of curricula whose primary objective is the support of individuals as they develop their identities or, as an alternative, to activities within a traditional curriculum that will recognise the importance of professional identity formation and address it specifically in programmes of instruction. We close, as we began, with a quote that encapsulates the ideas that we have attempted to develop within this chapter – 'the central issue in learning is becoming a practitioner, not learning about practice' [126].

References

1 Shakespeare, W. As You Like It. Act II, Scene 7.
2 Erikson, E.H. (1982). *The Lifecycle Completed*. New York: Norton.
3 Jarvis-Selinger, S., Pratt, D.D., and Regehr, G. (2012). Competency is not enough: integrating identity formation into the medical education discourse. *Academic Medicine* 87: 1185–1191.
4 Merton, R.K. (1957). Some preliminaries to a sociology of medical education. In: *The Student Physician: Introductory Studies in the Sociology of Medical Education* (ed. R.K. Merton, L.G. Reader and P.L. Kendall), 3–79. Cambridge, MA: Harvard University Press.
5 Goffman, E. (1959). *The Presentation of Self in Everyday Life*. New York: Doubleday.
6 Wear, D. and Kuczewski, M.G. (2004). The professionalism movement: can we pause? *American Journal of Bioethics* 4: 1–10.
7 Cruess, R.L., Cruess, S.R., Boudreau, J.D. et al. (2014). Reframing medical education to support professional identity formation. *Academic Medicine* 89: 1446–1451.
8 Irby, D.M. and Hamstra, S.J. (2016). Parting the clouds: three professionalism frameworks in medical education. *Academic Medicine* 91: 1606–1611.
9 Lave, J. and Wenger, E. (1991). *Situated Learning. Legitimate Peripheral Participation*. Cambridge, UK: Cambridge University Press.
10 Cooke, M., Irby, D.M., and O'Brien, B.C. (2010). *Educating Physicians: A Call for Reform of Medical School and Residency*. San Francisco, CA: Jossey-Bass.
11 Holden, M., Buck, E., Luk, J. et al. (2015). Professional identity formation: creating a longitudinal framework through TIME (transformation in medical education). *Academic Medicine* 90: 761–767.
12 Cruess, R.L., Cruess, S.R., and Steinert, Y. ed. (2016). *Teaching Medical Professionalism: Supporting the Development of a Professional Identity*. Cambridge, UK: Cambridge University Press.
13 Pellegrino, E.D. and Pellegrino, A.A. (1988). Humanism and ethics in roman medicine: translation and commentary on a text of Scribonius Largus. *Literature and Medicine* 7: 22–38.
14 Pellegrino, E. (2006). Character formation and the making of good physicians. In: *Lost Virtue: Professional Character Development in Medical Education* (ed. N. Kenny and W. Shelton), 1–17. Oxford, UK: Elsevier.
15 Martimianakis, M.A., Maniate, J.M., and Hodges, B.D. (2009). Sociological interpretations of professionalism. *Medical Education* 43: 829–837.
16 Starr, P. (1982). *The Social Transformation of American Medicine*. New York: Basic Books.
17 Krause, E. (1996). *Death of the Guilds: Professions, States and the Advance of Capitalism, 1930 to the Present*. New Haven, CT: Yale University Press.
18 Freidson, E. (2001). *Professionalism: The Third Logic*. Chicago: University of Chicago Press.
19 Cruess, S.R. and Cruess, R.L. (1997). Professionalism must be taught. *British Medical Journal* 315: 1674–1677.
20 Cohen, J.J. (2006). Professionalism in medical education, an American perspective: from evidence to accountability. *Medical Education* 40: 607–617.
21 Birden, H., Glass, N., Wilson, I. et al. (2014). Defining professionalism in medical education: a systematic review. *Medical Teacher* 36: 47–61.
22 Cruess, S.R., Johnston, S., and Cruess, R.L. (2004). Profession: a working definition for medical educators. *Teaching and Learning in Medicine* 16: 74–76.
23 Royal College of Physicians of London (2005). *Doctors in Society: Medical Professionalism in a Changing World*. London: Royal College of Physicians of London.
24 Helmich, E., Yeh, H.-M., Kalet, A., and Al-Araky, M. (2017). Becoming a doctor in different cultures: toward a cross-cultural approach to supporting professional identity formation in medicine. *Academic Medicine* 92: 58–62.
25 Ho, M.-J. (2013). Culturally sensitive medical professionalism. *Academic Medicine* 88: 1014.
26 Al-Eraky, M.M. and Chandratilake, M. (2012). How medical professionalism is conceptualized in Arabian context: a validation study. *Medical Teacher* 34: S90–S95.
27 Sullivan, W. (2005). *Work and Integrity: The crisis and promise of professionalism in North America*, 2e. San Francisco, CA: Jossey-Bass.
28 Cruess, R.L. and Cruess, S.R. (2008). Expectations and obligations: professionalism and medicine's social contract with society. *Perspectives in Biology and Medicine* 51: 579–559.
29 Stevens, R. (2001). Public roles for the medical profession in the United States: beyond theories of decline and fall. *Milbank Quarterly* 79: 327–353.
30 Shaw, G.B. (1946). *Introduction to The Doctor's Dilemma*, 75. Harmondsworth: Penguin Books (original published 1911).
31 Birden, H., Glass, N., Wilson, I. et al. (2013). Teaching professionalism in medical education: A Best Evidence Medical Education (BEME) systematic review. *Medical Teacher* 35: 569–574.
32 Frank, J.R. and Danoff, D. (2007). The CanMEDS initiative: implementing an outcomes-based framework of physician competencies. *Medical Teacher* 29: 642–647.
33 Accreditation Council for Graduate Medical Education (ACGME) and the American Board of Medical Specialists (ABMS) (2000). *The Outcome Project*. Iowa: Accreditation Council for Graduate Medical Education (ACGME).
34 General Medical Council (2013). *Good Medical Practice*. London: General Medical Council.
35 Cruess, R. and Cruess, S. (2006). Teaching professionalism: general principles. *Medical Teacher* 28: 205–208.
36 Stern, D.T. and Papadakis, M. (2006). The developing physician – becoming a professional. *New England Journal of Medicine* 355: 1794–1799.
37 Sullivan, W. (2005). Markets vs. professions: value-added? *Daedalus* 134: 10–26.

38 Kenny, N.P., Mann, K.V., and MacLeod, H. (2003). Role modeling in physicians' professional formation: reconsidering an essential but untapped educational strategy. *Academic Medicine* 78: 1203–1210.
39 Mann, K., Gordon, J., and MacLeod, A. (2009). Reflection and reflective practice in health professions education: a systematic review. *Advances in Health Sciences Education: Theory and Practice* 14: 595–621.
40 Steinert, Y., Cruess, R.L., Cruess, S.R. et al. (2007). Faculty development as an instrument of change: a case study on teaching professionalism. *Academic Medicine* 82: 1057–1064.
41 Hodges, B.D., Ginsburg, S., Cruess, R.L. et al. (2011). Assessment of professionalism: Recommendations from the Ottawa 2010 Conference. *Medical Teacher* 33: 354–363.
42 Coulehan, J. and Williams, P.C. (2001). Vanquishing virtue: the impact of medical education. *Academic Medicine* 76: 598–605.
43 Hafferty, F.W. (2016). Professionalism and the socialization of medical students. In: *Teaching Medical Professionalism* (ed. R.L. Cruess, S.R. Cruess and Y. Steinert), 54–68. Cambridge: Cambridge University Press.
44 Haidet, P. (2008). Where we are headed: a new wave of scholarship on educating medical professionalism. *Journal of General Internal Medicine* 23: 1118–1119.
45 Becker, H.S., Geer, B., Hughes, E.C., and Strauss, A.L. (1961). *Boys in White: Student Culture in Medical School*. Chicago, IL: University of Chicago Press.
46 Bosk, C. (1979). *Forgive and Remember*. Chicago, IL: University of Illinois Press.
47 Beagan, B.L. (2005). Everyday classism in medical school: experiencing marginality and resistance. *Medical Education* 39: 777–784.
48 Bebeau, M.J. (2006). Evidence based character development. In: *Lost Virtue: Professional Character Development in Medical Education* (ed. N. Kenny and W. Shelton), 47–86. Oxford, UK: Elsevier.
49 Brosnan, C. (2009). Pierre Bourdieu and the theory of medical education: thinking 'rationally' about medical students and medical curricula. In: *Handbook of the Sociology of Medical Education* (ed. C. Brosnan and B.S. Turner), 51–68. London: Routledge.
50 Monrouxe, L.V. (2010). Identity, identification and medical education: why should we care? *Medical Education* 44: 40–49.
51 MacLeod, A. (2011). Caring, competence and professional identities in medical education. *Advances in Health Sciences Education: Theory and Practice* 16: 375–394.
52 Monrouxe, L.V., Rees, C.E., and Hu, W. (2011). Differences in medical students' explicit discourses of professionalism: acting, representing, becoming. *Medical Education* 45: 585–602.
53 Weaver, R., Peters, K., Koch, J., and Wilson, I. (2011). 'Part of the team': professional identity and social exclusivity in medical students. *Medical Education* 45: 1220–1229.
54 Burford, B. (2012). Group processes in medical education: learning from social identity theory. *Medical Education* 46: 143–152.
55 Goldie, J. (2012). The formation of professional identity in medical students: considerations for educators. *Medical Teacher* 34: e641–e648.
56 Helmich, E., Bolhuis, S., Dornan, T. et al. (2012). Entering medical practice for the very first time: emotional talk, meaning and identity development. *Medical Education* 46: 1074–1086.
57 Monrouxe, L.V. (2013). Identities, self and medical education. In: *Oxford Textbook of Medical Education* (ed. K. Walsh), 113–123. Oxford: Oxford University Press.
58 Wilson, I., Cowin, L.S., Johnson, M., and Young, H. (2013). Professional identity in medical students: pedagogical challenges to medical education. *Teaching and Learning in Medicine* 25: 369–373.
59 Frost, H.D. and Regehr, G. (2013). 'I am a doctor': negotiating the discourses of standardization and diversity in professional identity construction. *Academic Medicine* 88: 1570–1577.
60 Cruess, S.R. and Cruess, R.L. (2016). General principles for establishing programs to support professionalism and professional identity formation at the undergraduate and postgraduate levels. In: *Teaching Medical Professionalism: Supporting the Development of a Professional Identity* (ed. R.L. Cruess, S.R. Cruess and Y. Steinert), 113–124. Cambridge: Cambridge University Press.
61 Bebeau, M.J. and Faber-Langendoen, K. (2014). Remediating lapses in professionalism. In: *Remediation in Medical Education: A Mid-Course Correction* (ed. A. Kalet and C.L. Chou), 103–127. New York: Springer.
62 Norcini, J.J. and Shea, J.A. (2016). Assessment of professionalism and progress in the development of a professional identity. In: *Teaching Medical Professionalism: Supporting the Development of a Professional Identity* (ed. R.L. Cruess, S.R. Cruess and Y. Steinert), 155–169. Cambridge: Cambridge University Press.
63 Shoemaker, S. (2006). Identity and identities. *Daedalus* 135: 40–48.
64 Piaget, J. and Inhelder, B. (1969). *The Psychology of the Child*. New York: Basic Books.
65 Marcia, J.E. (1966). Development and validation of ego identity status. *Journal of Personality and Social Psychology* 3: 551–558.
66 Kohlberg, L. (1984). *The Psychology of Moral Development: The Nature and Validity of Moral Stages*. San Francisco, CA: Harper and Row.
67 Kegan, R. (1982). *The Evolving Self: Problem and Process in Human Development*. Cambridge, MA: Harvard University Press.
68 Cohen, M.J.M., Kay, A., Youakim, J.M., and Balicuis, J.M. (2009). Identity transformation in medical students. *American Journal of Psychoanalysis* 69: 43–52.
69 Vignoles, V.L., Schwartz, S.J., and Luyckx, K. (2011). Toward an integrative view of identity. In: *Handbook of Identity Theory and Research* (ed. S.J. Schwartz, K. Luyckx and V.L. Vignoles), 1–27. New York: Springer.
70 Wenger-Trayner, E. and Wenger-Trayner, B. (2015). Learning in a landscape of practice. In: *Learning in a Landscape of Practice: A Framework* (ed. E. Wenger-Trayner, M. Fenton-O'Creevy, S. Hutchison, et al.), 13–31. London: Routledge.
71 Kaufman, D.M. and Mann, K.V. (2013). Teaching and learning in medical education: how theory can inform practice. In: *Understanding Medical Education: Evidence, Theory and Practice*, 2e (ed. T. Swanwick), 7–29. Oxford: Wiley.
72 Skorikov, V.B. and Vondracek, F.W. (2011). Occupational identity. In: *Handbook of Identity Theory and Research* (ed. S.J. Schwartz, K. Luyckx and V.L. Vignoles), 693–714. New York: Springer.
73 Wenger, E. (1998). *Communities of Practice: Learning, Meaning, and Identity*. Cambridge: Cambridge University Press.
74 Cruess, R.L., Cruess, S.R., and Steinert, Y. (2018). Medicine as a community of practice: Implications for medical education. *Academic Medicine* 93 (2): 185–191.
75 Lave, J. (1998). Situated learning in communities of practice. In: *Communities of Practice: Learning, Meaning, and Identity* (ed. E. Wenger), 63–82. Cambridge: Cambridge University Press.
76 Wenger, E. (2010). Communities of practice and social learning systems: the career of a concept. In: *Social Learning Systems and Communities of Practice* (ed. C. Blackmore), 179–199. London: Springer.
77 Bennett, D., Solomon, Y., Bergin, C. et al. (2017). Possibility and agency in a figured world: becoming a 'good doctor'. *Medical Education* 51: 248–257.
78 Ludmerer, K.M. (2015). *Let Me Heal: The Opportunity to Preserve Excellence in American Medicine*. New York: Oxford University Press.
79 Cruess, R.L., Cruess, S.R., Boudreau, D.B. et al. (2015). A schematic representation of the professional identity formation and socialization of medical students and residents: a guide for medical educators. *Academic Medicine* 90: 718–725.
80 (1989). *Oxford English Dictionary*, 2e. Oxford: Clarendon Press.
81 Monrouxe, L. (2016). Theoretical insights into the nature and nurture of professional identities. In: *Teaching Medical Professionalism:*

Supporting the Development of a Professional Identity (ed. R.L. Cruess, S.R. Cruess and Y. Steinert), 37–54. Cambridge: Cambridge University Press.

82 Kinghorn, W.A. (2010). Medical education as moral formation: an Aristotelian account of medical professionalism. *Perspectives in Biology and Medicine* 53: 87–105.

83 Mann, K., Gordon, J., and MacLeod, A. (2009). Reflection and reflective practice in health professions education: a systematic review. *Advances in Health Sciences Education: Theory and Practice* 14: 595–621.

84 Epstein, R.M. (2008). Reflection, perception and the acquisition of wisdom. *Medical Education* 42: 1048.

85 Billet, S. (2004). Conceptualizing workplaces as learning environments. *Journal of Workplace Learning* 16: 312–324.

86 Eraut, M. (2000). Non-formal learning and tacit knowledge in professional work. *British Journal of Educational Psychology* 70: 113–136.

87 Cote, L. and Leclere, H. (2000). How clinical teachers perceive the doctor-patient relationship and themselves as role models. *Academic Medicine* 75: 1117–1124.

88 Mann, K.V. and Gaufberg, E. (2016). Role modelling and mentoring in the formation of a professional identity. In: *Teaching Medical Professionalism: Supporting the Development of a Professional Identity* (ed. R.L. Cruess, S.R. Cruess and Y. Steinert), 4–97. Cambridge: Cambridge University Press.

89 Benbassat, J. (2014). Changes in well-being and professional values among medical undergraduate students: a narrative review of the literature. *Advances in Health Sciences Education: Theory and Practice* 19: 597–610.

90 Kolb, D. (1984). *Experiential Learning*. Englewood Cliffs: Prentice-Hall.

91 Dewey, J. (1963). *Experience and Education*. New York: Collier Books.

92 Mann, K., Dornan, T., and Teunissen, P.W. (2011). Perspectives on learning. In: *Medical Education: Theory and Practice* (ed. T. Dornan, K. Mann, A. Scherpbier and J. Spencer), 17–39. London: Churchill Livingstone.

93 Wenger, E. (2010). Communities of practice and social learning systems: the career of a concept. In: *Social Learning Systems and Communities of Practice* (ed. C. Blackmore), 179–199. London: Springer.

94 Billet, S. (1996). Situated learning: bridging sociocultural and cognitive theorizing. *Learning and Instruction* 6: 263–280.

95 Eraut, M. (2000). Non-formal learning and tacit knowledge in professional work. *British Journal of Educational Psychology* 70: 113–136.

96 Polanyi, M. (1958). *Personal Knowledge: Towards a Post-critical Philosophy*. Chicago, IL: University of Chicago Press.

97 Hafferty, F.W. and Hafler, J.P. (2011). The hidden curriculum, structural disconnects, and the socialization of new professionals. In: *Extraordinary Learning in the Workplace, Innovation and Change in Professional Education* (ed. J.P. Hafler), 17–35. New York: Springer.

98 Steinert, Y. (2016). Faculty development to support professionalism and professional identity formation. In: *Teaching Medical Professionalism: Supporting the Development of a Professional Identity* (ed. R.L. Cruess, S.R. Cruess and Y. Steinert), 124–140. Cambridge: Cambridge University Press.

99 Cruess, R.L., Cruess, S.R., and Steinert, Y. (2016). Amending Miller's pyramid to include professional identity formation. *Academic Medicine* 91: 155–169.

100 Relman, A.S. (2007). Medical professionalism in a commercialized health care market. *Journal of the American Medical Association* 298: 2668–2670.

101 Lesser, C.S., Lucey, C.R., Egener, B. et al. A behavioural and systems view of professionalism. *Journal of the American Medical Association* 304: 2732–2737.

102 Chretien, K.C. and Kind, T. (2013). Social media and clinical care: ethical professional and social implications. *Circulation* 127: 1413–1421.

103 Ryan Greyson, K.C., Chretien, K.C., and Kind, T. (2012). Physician violations of online professionalism and disciplinary actions: a national survey of state medical boards. *Journal of the American Medical Association* 307: 1141–1142.

104 Graungaard, A.H. and Anderson, J.S. (2014). Meeting real patients: a qualitative study of medical students' experiences of early patient contact. *Education for Primary Care* 25: 132–139.

105 Wald, H.S., Reis, S.P., Monroe, A.D., and Borkan, J.M. (2010). 'The loss of my elderly patient': interactive reflective writing to support medical students' rites of passage. *Medical Teacher* 32: e178–e184.

106 Yardley, S., Littlewood, S., Margolis, S.A. et al. (2010). What has changed in the evidence for early experience? Update of a BEME systematic review. *Medical Teacher* 32: 740–760.

107 Borges, N.J., Stephen, M.R., Elam, C., and Jones, B.J. (2006). Comparing millennial and generation X medical students at one medical school. *Academic Medicine* 81: 571–576.

108 Watson, D.E., Slade, S., Buske, L., and Tepper, J. (2006). Intergenerational differences in workloads: a ten year population-based study. *Health Affairs* 25: 1620–1628.

109 Bynum, W.E. and Goodie, J.L. (2014). Shame, guilt, and the medical learner: ignored connections and why we should care. *Medical Education* 48: 1045–1054.

110 LeBlanc, V. (2009). The effects of acute stress on performance: implications for health professions education. *Academic Medicine* 84: S25–S23.

111 Yerkes, R.M. and Dodson, J.D. (1908). The relation of strength of stimulus to the rapidity of habit-formation. *Journal of Comparative Neurology* 18: 459–482.

112 Janss, R., Rispens, S., Segers, M., and Jehn, K.A. (2012). What is happening under the surface? Power, conflict and the performance of medical teams. *Medical Education* 46: 828–849.

113 Benbassat, J. and Baumal, R. (2005). Enhancing self-awareness in medical students: an overview of teaching approaches. *Academic Medicine* 80: 156–161.

114 Wear, D., Aultman, J.M., Varley, J.D., and Zaeconi, J. (2006). Making fun of patients: medical students' perceptions and use of humor in clinical settings. *Academic Medicine* 81: 854–862.

115 Lingard, L. (2012). Language matters: towards an understanding of silence and humor in medical education. *Medical Education* 47: 40–48.

116 Feudtner, C., Christakis, D.A., and Christakis, N.A. (1994). Do clinical clerks suffer ethical erosion? Student perceptions of their ethical environment and personal development. *Academic Medicine* 69: 670–679.

117 Boudreau, J.D., MacDonald, M.E., and Steinert, Y. (2014). Affirming professional identities through an apprenticeship: insights from a four-year longitudinal case study. *Academic Medicine* 89: 1038–1045.

118 Boudreau, J.D. (2016). The evolution of an undergraduate medical programme on professionalism and identity formation. In: *Teaching Medical Professionalism in Support of Professional Identity Formation* (ed. R.L. Cruess, S.R. Cruess and Y. Steinert), 217–231. Cambridge: Cambridge University Press.

119 Wenger, E. (2010). Communities of practice and social learning systems: the career of a concept. In: *Social Learning Systems and Communities of Practice* (ed. C. Blackmore), 179–199. London: Springer.

120 Stoll, R.B., McMahon, A., Wallace, M., and Thomas, S. (2006). Professional learning communities: a review of the literature. *Journal of Educational Change* 7: 221–258.

121 Billett, S. (2002). Workplace pedagogical practices: co-participation and learning. *British Journal of Educational Studies* 50: 457–481.

122 Eraut, M. (2000). Non-formal learning and tacit knowledge in professional work. *British Journal of Educational Psychology* 70: 113–136.
123 Imel, S. (1999). Using groups in adult learning: theory and practice. *Journal of Continuing Education in the Health Professions* 19: 54–61.
124 Arnold, L., Sullivan, C., and Quaintance, J. (2016). Remediation of unprofessional behaviour. In: *Teaching Medical Professionalism in Support of Professional Identity Formation* (ed. R.L. Cruess, S.R. Cruess and Y. Steinert), 169–186. Cambridge: Cambridge University Press.
125 Sullivan, W.M. (2016). Foreword. In: *Teaching Medical Professionalism in Support of Professional Identity Formation* (ed. R.L. Cruess, S.R. Cruess and Y. Steinert), ix–xvi. Cambridge: Cambridge University Press.
126 Brown, J.S. and Duguid, P. (1991). Organizational learning and communities of practice: toward a unified view of working, learning and innovation. *Organization Science* 2 (1): 40–57.

Further Reading

Frost, H.D. and Regehr, G. (2013). 'I am a doctor': negotiating the discourses of standardization and diversity in professional identity construction. *Academic Medicine* 88: 1570–1577.

Goldie, J. (2012). The formation professional identity in medical students: considerations for educators. *Medical Teacher* 34: e641–e648.

Helmich, E., Yeh, H.-M., Kalet, A., and Al-Araky, M. (2017). Becoming a doctor in different cultures: toward a cross-cultural approach to supporting professional identity formation in medicine. *Academic Medicine* 92: 58–62.

Jarvis-Selinger, S., Pratt, D.D., and Regehr, G. (2012). Competency is not enough: integrating identity formation into the medical education discourse. *Academic Medicine* 87: 1185–1191.

18 Portfolios in Personal and Professional Development

Erik Driessen[1] and Jan van Tartwijk[2]

[1]Department of Educational Development and Research, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands

[2]Department of Education, Utrecht University, Utrecht, The Netherlands

KEY MESSAGES

- Portfolios are a useful vehicle for supporting and assessing learning in the clinical workplace.
- Portfolios serve three main goals in medical education: monitoring and planning learner development, assessing performance, and stimulating reflection.
- Depending on their purpose, portfolios may differ significantly in scope, structure, and content.
- The assessment of portfolios requires a qualitative interpretative approach.
- Mentors are a crucial factor in portfolio effectiveness.

Introduction

In 1990, Miller described the challenges of assessing 'clinical skills/competence/performance'. At that time, he signalled that in medical education instruments were available for assessing knowledge, skills, and competence, but not for assessing what a graduate does when functioning independently in a clinical practice [1]. Because of their ability to fill this gap in assessment, in recent decades portfolios have gained a prominent position in medical education [2, 3].

The portfolio concept is borrowed from the arts and architecture, where work samples and evidence of quality were traditionally kept in a portable case, a portfolio. Today, portfolios are used in many educational programmes aimed at developing competences and most educational portfolios are digital (see Box 18.2). Content may be prescribed or left to the discretion of individual learners, and the portfolio can report on work done, feedback received on this work, progress made, plans for improving competence, and reflections on performance and development [4, 5]. What makes the portfolio eminently suitable for supporting and assessing learning in the clinical workplace is its ability to accommodate non-standardised information about performance, and thereby do justice to the characteristics and challenges of individual learners and specific workplaces [6]. As a result, the portfolio is in perfect alignment with recent developments in education with a strong focus on learning in practice, such as outcomes-based education and competency-based learning.

This chapter focuses on the following topics:

- diversity of portfolios
- use of portfolios for the monitoring and planning of competency development
- portfolio assessment
- use of portfolios to stimulate reflection.

The evidence base for portfolios in medical education is summarised in Box 18.1 [2, 5, 7, 8]. For the rest of this chapter we will use the term 'learner' to refer to any individual in undergraduate or postgraduate medical education or practitioners undertaking continuing professional development.

Diversity of Portfolios

Scope

Portfolios can differ substantially in scope [9]. They may vary from being very limited (such as a portfolio for presentation skills only) focused on one single skill, competency domain, or curricular component, to being very broad, covering the learner's development across all relevant competency domains over a prolonged period of time.

Closed or Open

Learners composing a portfolio can be offered different degrees of structuring or guidance, with clear consequences for the content and structure of the portfolio.

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

BOX 18.1 WHERE'S THE EVIDENCE? Factors promoting portfolio success [2, 5, 7, 8]

Factor	Recommendation
Goals	Clearly explain the goals of working with the portfolio. Combine goals (learning and assessment).
Introducing the portfolio	Provide clear guidelines about the procedure, the format, and the content of the portfolio.
	Be on guard against problems with information technology.
	Use a hands-on introduction with a briefing on the objective of the portfolio and the procedures used.
Mentoring/interaction	Arrange for mentoring by teachers, trainers, supervisors, or peers.
Assessment	Incorporate safeguards in the assessment procedure, like intermittent feedback cycles, involvement of relevant resource persons (including the learner), and a sequential judgement procedure.
	Use assessment panels of two to three assessors depending on the stakes of the assessment.
	Train assessors.
	Use holistic scoring rubrics (global performance descriptors).
Portfolio format	Keep the portfolio format flexible.
	Avoid being overly prescriptive with regard to portfolio content.
	Avoid excessive paperwork.
Position in the curriculum	Integrate the portfolio with other educational activities in the curriculum.
	Be moderately ambitious with regard to early-undergraduate portfolio use.

A 'closed' portfolio with detailed guidelines and strict regulations allows learners relatively little freedom to determine the format and content of their own portfolios. Closed portfolios are easy to compare and navigate, which is an advantage for large-scale portfolio assessment. The downside is that the closed portfolio cannot really do justice to the characteristics of individual learners and specific workplaces, which is one of the most important reasons for using a portfolio.

A more 'open' portfolio results when directions are rather loose and general, allowing learners considerable freedom with respect to portfolio content and format. As a consequence, learners can provide richer descriptions of their individual learning processes and pay attention to specific characteristics of the workplaces in which they have worked. Here, the other side of the coin is that navigating such a portfolio is more difficult and reliable assessment of such a portfolio is a challenge.

Goals and their Relation to Portfolio Design

In medical education, portfolios serve three main goals: learner development, assessment, and reflection. Which goal or goals predominate drives the structure and content of the portfolio, as summarised in Figure 18.1 [2]. In portfolios that are used to *monitor and plan development*, the main features are overviews of achievements and targets. In portfolios used for *assessment*, the evidence of competency attainment takes centre stage. In portfolios primarily aimed at *stimulating reflection*, the core of the portfolio consists of written evaluations and performance analyses to direct performance improvement. There is overlap in assessment and reflection in the Learning portfolio, as shown in Figure 18.1. In practice, most portfolios combine all or some of these goals, and the goal mix determines what the portfolio looks like (see Box 18.2). There is no 'one size fits all' and in the next part of this chapter we focus on the three main goals of portfolio use: monitoring and planning development, assessment, and stimulating reflection.

Portfolios for Monitoring and Planning Development

From an educational perspective, workplaces are not the most suitable environment to enable learners to comply with the demands of a structured curriculum. While learning outcomes and competencies can be determined in advance, whether or not the workplace offers opportunities to achieve them depends on the presence of patients with different pathologies and the presence of clinical

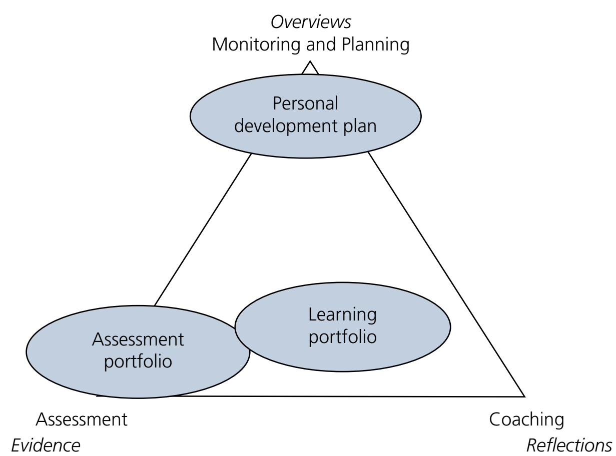


Figure 18.1 Purposes and content of portfolios.

BOX 18.2 Electronic portfolios

Electronic portfolios (e-portfolios) are widely used in medical education [9]; in most cases a vendor-supplied platform is used, but using standard applications is possible too [2]. An e-portfolio can have three functions:

- **i** provide a repository for all the materials (dossier)
- **ii** facilitate the administrative and logistical aspects of the assessment process (i.e. direct online loading of assessment and feedback forms via multiple platforms, regulation of who has access to which information and connecting information pieces to the overarching framework)
- **iii** enable a quick overview of aggregated information (such as overall feedback reports or overview of competency growth).

The (e-)portfolio should be easily accessible to any stakeholder who has access to it. The portfolio should combine good accessibility with safety by blocking unauthorised access. A user friendly and feasible platform is vital.

teachers to supervise learners. To put it differently, learning in the workplace depends on the availability of learning opportunities, and this inevitably varies from learner to learner [10]. On top of this, it is rare in undergraduate education (perhaps less so in postgraduate training) for students to have the same supervisor for more than one or two weeks [11]. The clinical workplace is thus by nature an erratic environment in which it is difficult for students to direct their own learning. While being immersed in clinical practice, students have a hard time perceiving exactly how their experiences can contribute to the overarching learning objectives and competency achievements required by the curriculum. Moreover, lack of continuity of supervision and limited observation of student activities stand in the way of effective monitoring of learner development. These problems can be solved by including in the portfolio a systematic overview of tasks undertaken to obtain specific competencies, the competency levels achieved, and areas where more work is needed [12] (see Box 18.3).

Setting Learning Goals

For such a portfolio to be effective, it is essential that well-defined learning goals are set for a specific period. Purposeful activities aimed at achieving these goals are one of the pillars of workplace-based learning [15]. Learning goals are often incorporated in professional development plans, which are included in the portfolio and used to guide progress interviews. Learning goals can be determined based on the following [16]:

- the programme requirements and the availability of a placement for the upcoming period
- the analysis of the portfolio and the progress interview – defining aspects that require special attention
- the learner's personal learning objectives – elective subjects, special interests, etc.

BOX 18.3 FOCUS ON: Entrustable professional activities (Figure 18.2)

Educational frameworks to structure workplace-based learning are often constructed using competencies. Competency frameworks emphasise that the focus of medical education is not exclusively on medical knowledge and skills but on non-technical competencies as well. Additionally, competencies are used to guide learning and assessment in the workplace. Learning activities in the workplace can be linked to competencies. However, in practice, linking abstract competencies to clinical work can be problematic. While records of learner performance may suggest adherence to the formal curriculum, in reality the connection between the formal curriculum and what learners actually learn in the workplace may be paper-thin. To bridge the gap between abstract competencies and clinical practice, ten Cate and Scheele [13] have introduced the concept of 'entrustable professional activities' (EPA). EPAs are tasks that are considered to be crucial to a certain profession and which every student must have mastered at the end of the course or curriculum. Given their importance, the EPAs are given special attention during the programme. Scheele et al. [14] developed three criteria to define EPAs: a task of high importance for daily practice; a high-risk or error-prone task; and a task that is exemplary of specific competencies.

In the portfolio the learner can collect materials in evidence of attainment of competency in one of the EPAs. The schematic below offers an example of how an EPA works in postgraduate medical training.

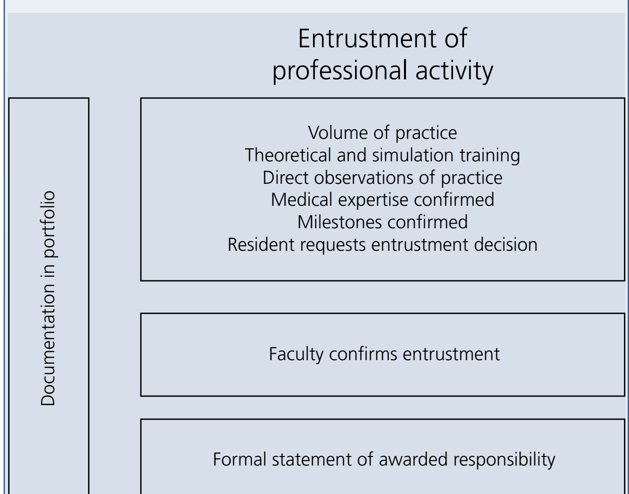


Figure 18.2 Example of an Entrustable Professional Activity (EPA).

For learning goals to be effective in steering development, it is important that both learner and teacher are committed to achieving them. The teacher should see to it that objectives are concrete and that a feasible plan is drawn up to achieve them. A useful aid for this is the SMART model: objectives should be Specific, Measurable, Achievable, Realistic, and

Time-bound, for only if these criteria are met is there a real chance that objectives will actually be achieved [17].

Portfolio Structure and Content

In portfolios that are used as instruments to promote and monitor development, overviews of what has been mastered and what remains to be achieved are important. Many portfolios provide overviews to be completed by learners to show what they have done, where they have done it, what they have learned as a result, and how they plan to proceed [12].

Such overviews could contain the following information:

- *Procedures or patient cases* Which procedures? What was the level of supervision? Which types of patients? What was learned? Were the activities assessed? Plans?
- *Prior work experience* Where? When? Which tasks? Strengths and weaknesses? Which competencies or skills were developed? Evaluation by the learner?
- *Prior education and training* Which courses or programmes? Where? When? What was learned? Completed successfully? Evaluation by the learner?
- *Experiences within and outside the course/programme* Where? When? Which tasks? What was done? Strengths and weaknesses? Which competencies or skills were developed? Evaluation by the learner? Plans?
- *Components of the course/programme* Which components have been attended so far? Which remain to be attended? When? What was learned? Completed successfully? Evaluation by the learner? Plans?
- *Competencies or skills* Where addressed? Level of proficiency? Plans? Preferences?

Portfolios for Assessment

In recent years, there has been a marked change in the thinking about ways of assessing portfolios. The traditional psychometric approach, characterised by a focus on objective judgement based on standardisation and analytical assessment criteria, has been found to be incompatible with the essentially non-standardised nature of many portfolios centred around the individual characteristics and challenges of individual learners and specific workplaces [18, 19]. The psychometric quantitative approach does not quite fit with portfolios containing a variety of qualitative information in addition to numerical information (scores) [6, 20]. Portfolios are not used to assess technical skills only: they are especially suited to assess non-technical skills, including professionalism [21]. This kind of assessment task cannot be translated into an analytical procedure with a standardised checklist and a list of strictly defined criteria [22]. Consequently, due to the presence of diverse qualitative information, when assessors weigh the information in a portfolio to assess competency they inevitably have to rely on their personal judgement [19, 23, 24].

To achieve a match between portfolio assessment and portfolio characteristics, we advocate an approach that leans heavily on the methodology of qualitative research [18, 19]. As is the case with most portfolios, qualitative research requires interpretation of different kinds of qualitative information to arrive at meaningful statements about ill-defined problems. The strategies listed below can be useful when assessing portfolios [2, 18].

Strategies in Portfolio Assessment

Arrange for Feedback Cycles

Conduct periodic feedback cycles to ensure that learners are not taken by surprise when the final judgement arrives. Since portfolio contents are usually compiled over a long period of time, it is ill advised to wait until the end of the period to make pronouncements about the quality of the portfolio. Intermediate formative assessments, such as feedback from a mentor,[1] are useful to allow learners to adapt and improve their portfolio. Regular feedback at different stages of portfolio development is advisable not only from an assessment perspective but from a learning perspective as well [12, 25].

Involve Multiple Informants

In addition to the assessors who judge the completed portfolio at the end of the period, different people who are in some way involved in the portfolio process can also make a valuable contribution to the assessment. The mentor is usually the first to comment on the quality of the portfolio [26]. He or she often knows the learner best, is in a position to ascertain the authenticity of the materials, and is familiar with the learner's work habits [27]. See Box 18.4 for more information on how to combine the role of mentor and assessor. Peers are another group that can contribute to the assessment. The advantages of peer assessment are twofold: peers know from experience what it means to produce a portfolio and by engaging in peer assessment they can familiarise themselves with the portfolio's assessment standards. Finally, learners can also self-assess the quality of their portfolios – for instance, by responding to the mentor's comments and/or by self-assessing their competencies. The literature shows that self-assessments tend to be biased [28]. To mitigate this bias, Eva and Regehr [29] recommended that learners should be encouraged to actively seek external information about their performance to arrive at well-validated self-assessments. In a similar vein, learners' self-assessments of their portfolios could be supported by mentor judgements to arrive at valid self-assessments [30].

Train Assessors

Organise a meeting (before the final assessment round and at an intermediate stage during the portfolio period) in which assessors can calibrate their judgements and discuss the assessment procedure and its results. Assessing vast amounts of highly diverse information in personalised portfolios requires professional judgement. Assessors inevitably use assessment criteria idiosyncratically, with judgement depending, for instance, on prior experiences and individual notions and beliefs about education and the competencies to be judged [31]. Differences between assessors can be reduced

[1] We use the term 'mentor' for a teacher or peer that supports the learner's development over a certain period of time. Other terms that could be used are tutor, coach, (clinical) teacher, or supervisor.

Box 18.4 HOW TO: Combine the role of mentor and assessor

Teachers commonly fulfil the combined role of mentor and assessor for their students. It has been argued, however, that combining these roles threatens the safety of the learning environment [43]. Elsewhere we have described alternative scenarios for the role of the mentor in assessment [2].

The PhD supervisor

In some scenarios the role of the mentor in the assessment procedure of the portfolio resembles that of supervisors of PhD students. In many countries, dissertations are formally assessed by a committee. When the supervisor considers the dissertation to be up to standard, he/she invites peers with relevant expertise to sit on the assessment committee, of which the supervisor is not a member. As a negative assessment would be harmful to the supervisor's reputation, supervisors are highly unlikely to convene a committee unless they are convinced the dissertation meets the criteria. In this type of procedure, mentors and learners have a shared interest: to produce a dissertation or portfolio that merits a positive judgement.

The driving instructor

In this model the roles of the mentor and the assessor are strictly separate. The mentor/driving instructor coaches the learner in achieving the required competencies, shown in the portfolio. When the mentor thinks the learner is sufficiently competent, he or she invites an assessor from the relevant professional body (i.e. the Driver and Vehicle Licensing Agency) to assess the competencies of the learner. It is also possible for learners to take the initiative to approach the licensing agency.

The coach

In this model, the learners take the initiative. They may ask a senior colleague to coach them until they have achieved the required level of competence. This scenario is appropriate for instance when a professional wants to obtain an additional qualification. The assessor would be someone from an external body.

by engaging them in a discussion of the judgement process [32]. After discussing a benchmark portfolio, for example, assessors' interpretations of assessment criteria may converge and a joint understanding of the procedure to be followed can be built. Such discussions should preferably be scheduled not only immediately before an assessment round but also at an intermediate stage of the portfolio period when assessors can compare their own portfolio judgements with those of their colleagues and discuss differences of interpretation [33]. After the final assessment, information about all the assessments should be communicated to assessors to help improve their understanding of the entire process.

Develop Sequential Assessment

At Maastricht Medical School a procedure has been developed to optimise efficient use of the time available for assessment in which conflicting information triggers the gathering of more information [18]. Mentors make a recommendation for the assessment of the portfolios of the students under their guidance. Individual students and an assessor decide whether they agree with the mentor's recommendation. Agreement signals the completion of the assessment procedure. In cases where there is no agreement, the portfolio is submitted to a larger group of assessors. In this way it is ensured that portfolios causing doubt are judged more carefully than portfolios where judgement is unanimous. As more judges are consulted, the trustworthiness of the assessment increases. Additionally, the discussions between the assessors will enhance clarity with respect to the application of the criteria (see also the section 'Train assessors' above).

Include Narrative Information

Incorporate in the portfolio requests for qualitative, narrative feedback, and give this information substantial weight in the assessment procedure. Narrative comments offer learners and assessors much richer information than quantitative, numerical feedback [15]. A score of 7 on a 10-point scale gives little insight into what a learner has and has not done well. Only when strengths and weaknesses of performance are explicated in narrative feedback does assessment become truly informative. A related problem in workplace assessment is rater's leniency. For various reasons, low scores are a rarity in practice, and consequently scores generally do not discriminate very well [34]. Narrative feedback, however, often provides more detailed and discriminative information about learner performance. Assessors can be encouraged to give narrative feedback by providing dedicated spaces in the assessment form.

Use Clear Rubrics or Descriptors

Education institutions often put a great deal of energy into generating competency profiles. The important thing is to strike a balance between very long lists of concrete criteria detailing everything a learner must be able to do ('can do-statements') and global descriptions offering a general outline but little practical guidance for assessors. In other words, the trick is to strike the right balance between analytical and global criteria. To achieve this, one might give learners and assessors an idea of the level to be attained for each global competency. A very useful instrument for this is rubrics or descriptors, which typically contain descriptions of expectations for each competency at different levels, such as a novice, a competent professional, and an expert [35].

Portfolio Structure and Content

Portfolios were first introduced into education for assessment purposes. These portfolios were basically nothing more than containers for storing various types of evidence of quality of performance. However, experiments with portfolios revealed that for evidence in a portfolio to be meaningful to assessors, it should be organised to reflect the competencies learners wish to demonstrate or the tasks they wish to illustrate [36]. To this end, captions should be attached to the evidence in the portfolio explaining what the evidence is supposed to show. The nature and diversity of the materials determine the richness of the picture the portfolio paints of the learner's activities and achievements. Although it may be tempting for learners to include a vast amount of material, leaving it up to the assessor to determine their value, this strategy is to be discouraged. It increases the assessors' workload and can also cause confusion by preventing assessors from seeing the forest for the trees. It is therefore important for learners to be selective, and an excellent selection criterion is that materials should provide insight into the learner's development and progress.

Portfolios to Stimulate Reflection

Cycles of Reflection

Elsewhere we have defined 'reflection' as 'letting future behaviour be guided by a systematic and critical analysis of past actions and their consequences' [37]. Learning from participating in the workplace is the process of transforming experiences into knowledge, skills, attitudes, and values, a process that can be represented graphically by experiential learning cycles, such as those proposed by Korthagen and colleagues and by Kolb and Fry [38, 39]. Concrete experiences, evaluation, analysis, formulation of abstract concepts and generalisations, and testing the implications in new situations are the stations of these cyclical models. At several of these stages, portfolios have a contribution to make. We will illustrate this using Korthagen's ALACT model (see Figure 18.3) [30, 37, 38].

Figure 18.3 A cyclical model of reflection [37].

Action

The cycle kicks off with action. To enable learners to improve their existing competencies while concurrently acquiring new ones, it is important to pre-select a task mix covering all the competencies required.

Looking back on action: Evaluation

Because unguided self-assessment is generally of (quite) poor quality, Eva and Regehr [29] proposed 'self-directed assessment seeking' as an alternative for this stage in professional development cycles. Following Boud [40], they describe self-directed assessment seeking as a process of taking personal responsibility for looking outward by explicitly seeking feedback and information from external sources. At this stage, the portfolio would be the 'folder' in which the information is stored and organised in line with the competencies to be attained and with captions indicating what the evidence shows and the conclusions to be drawn about the level of performance.

Awareness of essential aspects: Analysis

In the next step, the analysis, data are examined, patterns detected, and cause and effect associations identified. At this stage, theory can be helpful to identify patterns and causal associations. Research shows that it is not self-evident that learners are able to analyse their own performance appropriately [41]. In view of this, Korthagen et al. [38] recommend that mentors should ask questions to stimulate learners to discover and explicate the reasons underlying their own and others' actions and to pinpoint any inconsistencies in the analysis.

Creating or identifying alternative methods of action: Change

Following and based on the analysis, alternative methods of action should be selected. It is the role of the mentor to encourage the learner to consider alternative courses of action, to decide which one to use, and to justify that choice. A SMART (Specific, Measurable, Achievable, Realistic, Time-bound) action plan initiates the next cycle of reflective learning.

For more on reflection and reflective practice, see Chapters 4 and 25 in this book.

Portfolio Structure and Content

In portfolios aimed at stimulating reflection, written contributions feature prominently. These reflections can relate to the competencies the learner wishes to acquire, and the learner will generally also evaluate performance, analyse what has already been mastered, and determine which competencies need further development [30]. The reflections can also pertain to the learner's motivation for attending the programme, and/or how the learner views his/her self as a doctor/professional. These reflections can serve as a long-term agenda for learners.

In portfolios that are specifically aimed at stimulating reflection, the reflections are central in the portfolio structure, with learners supporting their reflections by referring to materials and overviews in the portfolio [42]. This helps to focus the reflections, because learners are likely to aim for consistency of reflections and evidential materials. The requirement that reflections be supported by evidence helps to make reflections less non-committal. It is, for instance, not acceptable for learners to simply state that they have learned how to give a clinical presentation; they have to substantiate this statement by evidential materials and overviews demonstrating why and how they have done this.

Conclusions

The portfolio approach has theoretical as well as practical merits. It can capture performance and development in the workplace using qualitative information that can take into account unique characteristics of specific workplaces. In this way, the portfolio completes the assessment landscape by enabling assessment at Miller's 'does' level [1]. Portfolios that include reflective writing require learners to engage in a 'conversation with self', which can be enhanced by reflective discussions with another person and by aiming for consistency with the evidence in the portfolio. Reflection provides learners and mentors a means to keep an overview of what has already been achieved and what still remains to be done.

Portfolios do not work of and by themselves. For a portfolio to be effective, certain conditions must be fulfilled. Probably the most crucial factor is the mentor: a person with whom the learner discusses the content of his/her portfolio (see Box 18.4).

Acknowledgement

We would like to acknowledge John Pitts, whose chapter in the first edition of this book informed our writing of this one.

References

1 Miller, G.E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine* 65 (9 Suppl): S63–S67.
2 Van Tartwijk, J. and Driessen, E.W. (2009). Portfolios for assessment and learning: AMEE guide no. 45. *Medical Teacher* 31 (9): 790–801.
3 Driessen, E.W. (2017). Do portfolios have a future? *Advances in Health Science Education* 22 (1): 221–228.
4 Paulson, F.L., Paulson, P.R., and Meyer, C.A. (1991). What makes a portfolio a portfolio. *Educational Leadership* 48 (5): 60–63.
5 Driessen, E., van Tartwijk, J., van der Vleuten, C., and Wass, V. (2007). Portfolios in medical education: why do they meet with mixed success? A systematic review. *Medical Education* 41 (12): 1224–1233.
6 Snadden, D. (1999). Portfolios – attempting to measure the unmeasurable? *Medical Education* 33 (7): 478–479.
7 Buckley, S., Coleman, J., Davison, I. et al. (2009). The educational effects of portfolios on undergraduate student learning: a Best Evidence *Medical Education* (BEME) systematic review. BEME guide no. 11. *Medical Teacher* 31 (4): 282–298.
8 Tochel, C., Haig, A., Hesketh, A. et al. (2009). The effectiveness of portfolios for post-graduate assessment and education: BEME guide no. 12. *Medical Teacher* 31 (4): 299–318.
9 JChertoff, J., Wright, A., Novak, M. et al. (2015). Status of portfolios in undergraduate medical education in the LCME accredited US medical school. *Medical Teacher* doi:10.3109/0142159X.2015.11145952.
10 Billet, S. (2006). Constituting the workplace curriculum. *Journal of Curriculum Studies* 38 (1): 31–48.
11 Jones, M.D. Jr., Rosenberg, A.A., Gilhooly, J.T., and Carraccio, C.L. (2011). Perspective: competencies, outcomes, and controversy–linking professional activities to competencies to improve resident education and practice. *Academic Medicine* 86 (2): 161–165.
12 Driessen, E.W., van Tartwijk, J., Govaerts, M. et al. (2012). The use of programmatic assessment in the clinical workplace: a Maastricht case report. *Medical Teacher* 34 (3): 226–231.
13 ten Cate, O. and Scheele, F. (2007). Competency-based postgraduate training: can we bridge the gap between theory and clinical practice? *Academic Medicine* 82 (6): 542–547.
14 Scheele, F., Teunissen, P., Van Luijk, S. et al. (2008). Introducing competency-based postgraduate medical education in the Netherlands. *Medical Teacher* 30 (3): 248–53.
15 Ericsson, K.A. (2015). Acquisition and maintenance of medical expertise: a perspective from the expert-performance approach with deliberate practice. *Academic Medicine* 90 (11): 1471–1486.
16 Driessen, E., Kenter, G., de Leede, B. et al. (2011). Richtlijn voortgangs-gesprek in de medische vervolgopleiding. *Tijdschrift voor Medisch Onderwijs* 30 (6 Suppl 3): 51–62.
17 Overeem, K., Driessen, E.W., Arah, O.A. et al. (2010). Peer mentoring in doctor performance assessment: strategies, obstacles and benefits. *Medical Education* 44 (2): 140–147.
18 Driessen, E., van der Vleuten, C., Schuwirth, L. et al. (2005). The use of qualitative research criteria for portfolio assessment as an alternative to reliability evaluation: a case study. *Medical Education* 39 (2): 214–220.
19 van der Vleuten, C.P., Schuwirth, L.W., Scheele, F., Driessen, E.W. and Hodges, B. (2010). The assessment of professional competence: building blocks for theory development. *Best Practice and Research. Clinical Obstetrics and Gynaecology* 24 (6): 703–19.
20 Pitts, J. and Coles, C. (2003). The challenge of non-standardised assessment of professionals-the need for a paradigm shift. *Education for Primary Care* 14: 397–405.
21 Driessen, E. (2009). Portfolio critics: do they have a point? *Medical Teacher* 31 (4): 279–281.
22 Kuper, A., Reeves, S., Albert, M., and Hodges, B.D. (2007). Assessment: do we need to broaden our methodological horizons? *Medical Education* 41 (12): 1121–1123.
23 van der Vleuten, C.P., Schuwirth, L.W., Driessen, E.W. et al. (2012). A model for programmatic assessment fit for purpose. *Medical Teacher* 34 (3): 205–214.
24 Eva, K.W. and Hodges, B.D. (2012). Scylla or Charybdis? Can we navigate between objectification and judgement in assessment? *Medical Education* 46 (9): 914–919.
25 Altahawi, F., Sisk, B., Poloskey, S. et al. (2012). Student perspectives on assessment: experience in a competency-based portfolio system. *Medical Teacher* 34 (3): 221–225.
26 Dekker, H., Driessen, E., Ter Braak, E. et al. (2009). Mentoring portfolio use in undergraduate and postgraduate medical education. *Medical Teacher* 31 (10): 903–9.

27 Driessen, E., van Tartwijk, J., Vermunt, J.D., and van der Vleuten, C.P. (2003). Use of portfolios in early undergraduate medical training. *Medical Teacher* 25 (1): 18–23.

28 Davis, D.A., Mazmanian, P.E., Fordis, M., Van Harrison, R., Thorpe, K.E. and Perrier, L. (2006). Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *Journal of the American Medical Association* 296 (9): 1094–102.

29 Eva, K.W. and Regehr, G. (2008). 'I'll never play professional football' and other fallacies of self-assessment. *Journal of Continuing Education in the Health Professions* 28 (1): 14–19.

30 Driessen, E., Overeem, K. and Van Tartwijk, J. (2010). Learning from practice: mentoring, feedback, and porfolios. In: Dornan T, Mann K, Scherpbier A and Spencer J (eds) *Medical Education: Theory and Practice*, 211–27. Edinburgh: Churchill Livingstone Elsevier.

31 Govaerts, M.J., Van de Wiel, M.W., Schuwirth, L.W., Van der Vleuten, C.P. and Muijtjens, A.M. (2013). Workplace-based assessment: raters' performance theories and constructs. *Advances in Health Sciences Education* 18 (3): 375–96.

32 Pitts, J., Coles, C., Thomas, P., and Smith, F. (2002). Enhancing reliability in portfolio assessment: discussions between assessors. *Medical Teacher* 24 (2): 197–201.

33 Rees, C.E. and Sheard, C.E. (2004). The reliability of assessment criteria for undergraduate medical students' communication skills portfolios: the Nottingham experience. *Medical Education* 38 (2): 138–144.

34 Williams, R.G. and Dunnington, G. (2004). Prognostic value of resident clinical performance ratings. *Journal of the American College of Surgeons* 199 (4): 620–627.

35 Meade, L.B., Borden, S.H., Mcardle, P. et al. (2012). From theory to actual practice: creation and application of milestones in an internal medicine residency program, 2004–2010. *Medical Teacher* 34 (9): 717–723.

36 Bird, T. (1990). The schoolteacher's portfolio: an essay on possibilities. In: *The New Handbook of Teacher Evaluation: Assessing Elementary and Secondary School Teachers* (ed. J. Millman and L. Darling-Hammond), 241–256. Newbury Park, CA: Corwin Press Inc.

37 Driessen, E., van Tartwijk, J., and Dornan, T. (2008). The self critical doctor: helping students become more reflective. *British Medical Journal (Clinical Research Edition)* 336 (7648): 827–830.

38 Korthagen, F.A.J., Kessels, J., Koster, B. et al. (2001). *Linking Theory and Practice: The Pedagogy of Realistic Teacher Education*. Mahwah, NY: Lawrence Erlbaum Associates.

39 Kolb, D.A. (1984). *Experiental Learning: Experience as the Source of Learning and Development*. Englewood Cliffs, NJ: Prentice Hall.

40 Boud, D. (1999). Avoiding the traps: seeking good practice in the use of self assessment and reflection in professional courses. *Social Work in Education* 18: 121–132.

41 Mansvelder-Longayroux, D.D., Beijaard, D., and Verloop, N. (2007). The portfolio as a tool for stimulating reflection by student teachers. *Teaching and Teacher Education* 23 (1): 47–62.

42 Van Tartwijk, J., Van Rijswijk, M., Tuithof, H. and Driessen, E.W. (2008). Using an analogy in the introduction of a portfolio. *Teaching and Teacher Education* 24: 927–38.

43 Tigelaar, D., Dolmans, D., Wolfhagen, H. and van der Vleuten, C. (2004). Using a conceptual framework and the opinion of portfolio experts to develop a teaching portfolio prototype. *Studies in Educational Evaluation* 30: 305–21.

Further Reading

Driessen, E., van der Vleuten, C., Schuwirth, L. et al. (2005). The use of qualitative research criteria for portfolio assessment as an alternative to reliability evaluation: a case study. *Medical Education* 39 (2): 214–220.

Driessen, E.W. (2017). Do portfolios have a future? *Advances in Health Science Education* 22 (1): 221–228.

Franco, R.S., Giuliani dos Santos Franco, C.A., Pestana, O. et al. (2016). The use of portfolios to foster professionalism: attributes, outcomes, and recommendations. *Assessment & Evaluation in Higher Education* doi: 10.1080/02602938.2016.1186149.

19 Continuing Professional Development

Simon Kitto[1,2], David Price[3,4], Dahn Jeong[1], Craig Campbell[5], and Scott Reeves[6]

[1]Office of Continuing Professional Development, Department of Innovation in Medical Education, Faculty of Medicine, University of Ottawa, Ottawa, Ontario, Canada
[2]Department of Surgery, University of Toronto, Toronto, Ontario, Canada
[3]Deparment of Family Medicine, University of Colorado School of Medicine, Denver, CO, USA
[4]ABMS Research and Education Foundation, American Board of Medical Specialties, Chicago, IL, USA
[5]Office of Specialty Education, Royal College of Physicians and Surgeons of Canada, Ottawa, Ontario, Canada
[6]Faculty of Health, Social Care and Education, Kingston University and St George's University of London, London, UK

KEY MESSAGES

- Physicians' professional development and lifelong learning model has shifted to a broader model of Continuing Professional Development (CPD), which encompasses diverse activities, addresses diverse aspects of physicians' competency, and includes diverse professionals, highlighting the importance of team-based interprofessional education.
- Self-Directed Learning (SDL) is a CPD activity that allows physicians to have control over their own learning and manage their time efficiently. Individual and environmental factors need to be taken into consideration to maximise the effectiveness of SDL for CPD.
- Contemporary forms of competency-based CPD emphasise the impact of learning on performance, quality of care, and health outcomes.
- Maintenance of Certification provides a framework for physicians' continuous professional development and improvement across the competencies throughout the course of their careers.
- Team-based learning is a CPD activity that can improve individuals' ability to work in teams and has a positive impact on organisation of clinical practice and patient outcomes.

Introduction

Recently, the emphasis of physicians' professional development has been shifting toward the broader concept of continuing professional development (CPD) and away from the narrower traditional view where it was perceived as a didactic and clinically oriented activity [1]. According to Sargeant and colleagues [1], CPD 'encompasses multiple educational and developmental activities physicians undertake to maintain and enhance their knowledge, skills, performance and relationships in the provision of health care' (p. S54). CPD is also broad in the sense that it not only comprises diverse activities but also addresses diverse aspects of physicians' competency [1]. It has been argued that the ultimate goal of CPD is to improve the quality of care and health outcomes of the public [1].

Several future trends in CPD have been predicted [2]: a growing attention to interprofessional and team-based learning through a greater focus on interprofessional education; increased use of longitudinal and multidimensional educational interventions; more frequent use of research that looks at what works, under what conditions, and why in CPD; a need for more knowledge on how teams change their practice; a shift away from a narrow focus on evidence-based clinical practices and recognition of other ways of knowing; emphasis on peer-to-peer learning; tension between accreditation policies that are predicated on simplistic individualist models of CPD and the realities of current CPD practice; better integration of CPD activities and the workplace, etc. Inspired by Olson's predictions for the future of CPD [2], we outline a broad concept of CPD consisting of a range of diverse elements including:

- self-directed learning
- the competencies across the medical education continuum
- the relationship between maintenance of certification and professional development
- team-based education in CPD.

Self-directed Learning in Continuing Professional Development

With new medical and scientific information emerging faster than ever, the field of medicine is characterised by rapid advances in clinical, scientific, and technological knowledge. This exponential rise of new medical knowledge

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

has emphasised the importance of lifelong learning of health care professionals to keep up with the current and pertinent clinical knowledge to maintain their competence. Around the world, lifelong learning of physicians is encouraged, supported, and managed by CPD programmes. The Physicians' Charter, which was endorsed by over 120 international organisations, advocates that 'physicians must be committed to lifelong learning' [3, 4]. In addition, in many countries, including Canada and the United States, self-directed lifelong learning is now required by medical professional and regulatory associations for board certification and maintenance of certification [5–7].

Self-directed learning (SDL) is a set of skills that can be taught, learned, and acquired [8] and is considered in CPD to be a critical part of the process that allows physicians to stay up-to-date with the current knowledge [9]. SDL is a central concept in adult learning that has a long history prior to its adoption [10] as one of the central elements of CPD. Educational psychologists view SDL 'as a complex cycle comprising intermingling elements such as psychological characteristics (e.g. self-efficacy, motivation, beliefs and learning style), personal choices (effort expended, learning strategies), judgments (self-assessment, attributions) and personal actions (e.g. goal setting, adjustment)' [11, p. 201]. There are many different definitions, models, and way of theorising SDL. Readers who are interested in learning more about the different theoretical approaches to SDL can refer to Chapter 4 in this book.

In medicine, there is no one, shared definition of SDL. Many medical associations have their own definition of SDL, for example, in Canada, the Royal College of Physicians and Surgeons defines SDL as 'activities planned to address specific needs, enhance awareness of new evidence potentially relevant to practice or enhance the quality of multiple systems' [5]. Within adult education in general, the most popular and frequently cited definition of SDL is from Knowles [12], in which SDL is described as 'a process in which individuals take the initiative, with or without the help of others, in diagnosing their learning needs, formulating learning goals, identifying human and material resources for learning, choosing and implementing appropriate learning strategies, and evaluating learning outcomes' (p. 18). Knowles also identified five components of cognitive activities undertaken during an SDL process: (i) diagnosing learning needs, (ii) formulating learning goals, (iii) identifying human and material resources for learning, (iv) choosing and implementing appropriate learning strategies, and (v) evaluating learning outcomes [12]. For CPD in medicine, Sargeant et al. suggest a conceptual framework of SDL that incorporates two elements, one being the five cognitive activities identified by Knowles, and the other being the willingness of learners (i.e. physicians) to drive their own learning [1, 13]. According to Sargeant et al. [1], physicians' willingness to self-direct their own learning is the key to lifelong learning and effective CPD.

A systematic review conducted by Murad et al. [9] found several factors associated with greater effectiveness of SDL, including: (i) learners' involvement in choosing the learning methods, strategies, and resources; (ii) engagement of more advanced learners rather than novices; and (iii) learners' discipline (e.g. nurses had a larger effect of SDL compared to other health professionals). In addition, Naumann et al. [14] conducted a literature review of SDL activities in Continuing Medical Education (CME). Box 19.1 presents some of the SDL strategies and factors that lead to most and least effective SDL outcomes (e.g. effect size of change in knowledge, effect size in physician performance and patient outcomes), adapted from a literature review by Naumann et al. [14].

Furthermore, Curran et al. [34] have identified barriers to SDL, especially in terms of SDL using social media and technologies.

Physicians' barriers to SDL included:

- lack of time
- lack of knowledge and/or understanding of what resources are available and how to use/find them
- preferences for more traditional approaches to collaboration
- information overload
- privacy and professionalism issues
- technical difficulties, such as inadequate wireless and Internet access

BOX 19.1 FOCUS ON: Factors contributing to most and least effective SDL outcomes [14]

Factors/programmes contributing to most effective SDL outcomes	Factors/programmes contributing to least effective SDL outcomes
• Social interaction [1, 15–17]	• Familiar content [28]
• Problem-based learning [18]	• Inaccurate needs assessment [29, 30]
• Point of care learning [19, 20]	• Informal 'scanning' approaches [31]
• Addressing appropriate learning objectives [21]	• Inexperience of learners [32]
• Guided self-assessment [22–24]	• Age of learners [31, 32]
• Reflection [15, 24]	• Unguided self-assessment [1]
• Blended learning [25–27]	• Lack of social interaction [31]
• Use of technology [28]	• Unstructured reflection or reflection on aggregate performance [24, 33]

- difficulties in trusting that information/content is accurate, peer reviewed, and/or credible.

Several studies have identified strategies for successful outcomes associated with SDL: dedicating time, creating accountability, and helping with goal generation through mentorship or relying more on external assessments than on self-assessment [35–37].

In addition, a scoping review published in 2018 detailed a comprehensive list of barriers and facilitators that influence the engagement of physicians in SDL in CPD, and identified current issues around SDL design and implementation in CPD programmes [38]. This review found that the barriers to SDL most frequently discussed in the literature were related to environmental context and resources, as previously shown [34], and the facilitators to SDL most frequently discussed in the literature were related to social aspects experienced by physicians such as teamwork and collaborative work [38]. The authors highlighted a gap in the knowledge about external factors that influence the design and implementation experienced by SDL developers [38].

These studies show that much of the current discussion around SDL focuses on learners' individual characteristics such as the ability to identify and address their own learning needs. However, Knowles' original concept characterised SDL as an activity in which external sources (i.e. various kinds of helpers, such as teachers, tutors, mentors, resource people, and peers) also played a highly important role [12, 39, 40]. Schumacher et al. [39] propose that, in the absence of external sources of feedback and information, SDL depends solely on the learners' ability to self-assess and this may, in turn, make effective SDL impossible [41]. The authors also identify three principles for creating 'master learners' in the context of a competency-based education approach [39]: (i) while learners must take responsibility for their learning, they must also seek external information to guide their efforts and calibrate their self-assessments; (ii) while teachers must allow learners to assume responsibility for their learning, they must provide the role modelling, support, and feedback necessary to guide learning and assessment; and (iii) while the learning environment impacts learners and teachers through defining the culture and context in which professional formation occurs, learners and teachers must attend to the reciprocal impact they have on the learning environment, so as to ensure they are also creating a meaningful learning environment for others. The authors define the 'master learner' as 'a learner who demonstrates the most advanced level of lifelong learning skills' [39, p. 1635]. By creating 'master learners', these three principles can also help inform CPD: master learners who are able to take part in SDL effectively and independently are an essential component to successful CPD.

Moreover, studies from a variety of fields indicate that self-directed learners: can cope better with change [42], have the ability to adapt [43], and have greater access to new knowledge and skills than passive learners [44]. Moreover, SDL produces better results in learning and retention through a longitudinal learning process and by giving learners more responsibility for their own learning [12]. Among practising physicians, studies have shown several benefits of SDL – such as increased performance, maintenance of competence, and improved patient outcomes and quality of care [45–49]. Furthermore, SDL encourages physicians to adapt to the rapid emergence and constant evolution of evidence-based practice [12].

Physicians' busy schedules and heavy workloads are two of the main challenges to SDL. They find it difficult to adopt and incorporate new evidence into their busy and demanding practice. However, SDL has flexibility; learners have control over their learning [50]. SDL enables physicians to undertake the learning process at their own pace and time, formulate their own goals, and choose the learning strategies that accommodate their work-context pressures.

Conversely, some study findings raise questions about the effectiveness of SDL. For example, there are potential risks for learner misunderstandings, overestimation of preparedness for practice, and development of unchecked bad habits [14, 51]. Additionally, SDL and lifelong learning rely heavily on accurate self-assessment, but several studies have found inaccuracies in physicians' self-assessments of their knowledge and skills. Physicians with lowest external performance and with less experience tend to overestimate their competence [21, 52–55]. Some suggest the contradictory evidence surrounding SDL may be caused by inconsistent definitions and interpretations of SDL, particularly as applied to CPD programmes in medicine [8, 56, 57]. Other authors argue for the importance of employing a consistent and evidence-based model of SDL within and across the continuum of medical education and across programmes in order to explore the effectiveness of this approach [8].

Hiemstra and Brockett [58] propose an updated model of SDL from the Person, Responsibility, Orientation (PRO) model [59] based on the enhanced understanding of the concept of self-direction in learning. The updated model emphasises the equal importance of three dimensions to SDL: Person, Process, and Context (PPC). Box 19.2, adapted from Hiemstra and Brockett's PPC model, shows the elements of the three dimensions. With this updated model, the authors conclude that previous models of SDL under-emphasised the importance of context in SDL and further emphasise that SDL is a complex process which is most effective when person, process, and context are in balance [58].

As such, we suggest that to maximise the potential of SDL for CPD, both learners and educators must consider and address the various contributing factors.

Competencies Across the Continuum of Medical Education

Stimulated by the Flexner report in 1910 [60], the structures and processes defined for training doctors have remained largely unchanged. As previously noted in this chapter, the exponential growth in knowledge and the evidence that informs practice have challenged medical schools and residency education programmes to ensure their graduates have acquired the requisite knowledge, skills, and abilities to enable them to address the health needs of patients and

BOX 19.2 FOCUS ON: The 'Person, Process, Context' (PPC) model [58]

Person	Process	Context
Individual characteristics: • Creativity • Critical reflection • Enthusiasm • Life experience • Life satisfaction • Motivation • Previous education • Resilience • Self-concept	Teaching-learning transaction: • Facilitation • Learning skills • Learning styles • Planning • Organising • Evaluating abilities • Teaching styles • Technological skills	Environmental and sociopolitical climate: • Culture • Power • Learning environment • Finances • Gender • Learning climate • Organisational policies • Political milieu • Race • Sexual orientation

populations. Although the application of competency-based medical education has traditionally focused on residency education, competency frameworks such as the Scottish Doctor [61], Netherlands National Undergraduate Framework [62], and CanMEDS 2005 [63] have been used to guide curriculum design and assessment to enable students to assume greater responsibility to reflect on and 'monitor their progress toward stated goals and elect to focus on those activities that will assist them to manage any deficiencies' [64, p. 648]. This approach to learning and assessment in medical schools is viewed as helping students to develop the capacity for self-reflection, exert control over the learning process, and acquire the skills of lifelong learning. Interestingly, CME and CPD have only recently been incorporated into discussions of competency-based medical education. Traditionally, CME was organised around a short course model that used experts to disseminate medical knowledge or evidence to participants who remained largely passive. Over time, this didactic model was replaced with a more interactive, interprofessional competency-based model.

Although calls for the implementation of competency-based medical education are not new [65], the transition from a time-based approach (immersing learners in a specific context for a period of time) to a competency-based approach to residency education began with the introduction of *Tomorrow's Doctors* in 1993 [66]. This shift toward an outcomes-based education model was intended to replace the emphasis on prescribing the structure (blocks or rotations, academic half-days) and processes (creating rotational or curricular objectives, mandatory requirements, instructional processes designed to achieve the intended objectives) for how education will be implemented to focus more on how education and a programme of regular assessment with feedback can enable trainees to demonstrate the outcomes (competencies or abilities) the curriculum intended them to achieve. In competency-based medical education, time is a resource not a metric of ability or the achievement of competence. Similar initiatives in undergraduate medical education were initiated several years later [67] and over the course of the past two decades competency-based medical education has become a global movement.

Competency-based medical education has been defined by an international group of medical educators as 'an outcomes-based approach to the design, implementation, assessment and evaluation of medical education programmes, using an organizing framework of competencies' [68, p. 641]. This emphasis on defining the abilities graduates are expected to achieve was intended to promote greater learner engagement in the educational process by providing greater transparency on what must be learned and demonstrated by the end of the training programme. Competency-based medical education is shifting the focus on measuring the amount of time a trainee spent in a specific rotation or content domain (e.g. emergency room or endocrinology) toward ensuring all graduates are making successful progress toward the demonstration of competence across all essential domains.

The use of competencies as an organising framework for curriculum development has been typically expressed through milestones that describe the behaviours or abilities expected of a trainee at a specific point in time. Regardless of approach, milestones are intended to guide curricular development, teaching, and learner development and guide the monitoring of 'progression of competence' to 'improve resident physicians' ability to provide quality patient care and to work effectively in current and evolving health care delivery systems' [69, p. 648]. Although the impact of milestones for CPD remains unknown, milestones may serve to guide physicians in the selection of learning activities and the assessment of outcomes for practice.

Competency-based medical education also re-framed the traditional approach to assessment. There has been a shift from end-of-rotation in-training evaluation reports (ITERs) supplemented by other assessment strategies (for example MCQs, OSCEs, simulation, and other work-based assessment strategies) to a programmatic model for assessment organised around entrustable professional activities (EPAs).

Although there is significant heterogeneity on approaches to competency-based medical education, the use of milestones and EPAs are intended to define a set of outcomes that enable programmes to provide frequent feedback and monitor progression to achievement of the intended outcomes. The rationale for adoption of competency-based medical education has equally been based on the need to align education with societal and patient needs and to demonstrate greater accountability to the public [70]. For further reading about EPAs and milestones, see Chapters 5 and 22 in this book.

How competency-based medical education will influence CPD remains largely unknown. Most international systems or programmes of CPD require physicians to engage in and report on their participation in learning and assessment activities using credits (time spent) as a metric. The strategies supporting engagement of physicians in CPD has traditionally relied on the ability of physicians to self-assess their needs, despite the research evidence that unguided self-assessment is inaccurate [37]. As previously described in this chapter, engagement in assessments that provide individual physicians, groups, or health care teams with multiple sources of data, such as feedback from patients and colleagues, formal appraisals of performance in their workplace and identifying patient outcomes to guide future learning and continuously improve the quality of care provided to patients have numerous challenges. Competency-based medical education will shift CPD from its original 'roots' as a practice-based learning and self-assessment strategy to requiring physicians to demonstrate how engaging in learning and assessment has impacted performance, the quality of care, and health outcomes experienced by patients (outcomes-based). Competency-based CPD should enable physicians to answer questions such as:

- Has my growth in competence and performance improved?
- Is the care I provide to patients reflective of the best evidence?
- Are my patients better off?
- Is my workplace safe for patients?

The Royal College of Physicians and Surgeons of Canada's 'Competence by Design' project aims to express a competency-based medical education model across the continuum. This project defined a competence continuum that included two additional stages beyond residency training [71]. Milestones for the CPD stage were proposed during the revisions to the CanMEDS framework published in 2015 [72]. The development of a shared vision for competency-based CPD has described competency-based CPD as a 'system that uses competencies to continuously improve specialty practice, patient outcomes and the health system' [73, p. 2]. The model is still in development and is not planned for implementation until 2020. However, the transition to competency-based CPD aims to use competencies as the foundation for the development of lifelong learning strategies that are relevant to a physician's specialty and practice. Competency-based CPD will place a greater emphasis on demonstrating how learning impacts performance and the quality of care provided to patients. The creation of a learning plan based on a response to patient and population health needs is envisioned to require an integration of self-appraised needs with external measures of performance and health outcomes. Models of 'informed self-assessment' as described by Sargeant et al. [74], coupled with evidence-informed feedback strategies [21, 74–76], are viewed as important in 'making sense' of external sources of data particularly when they conflict with the physicians' views of their performance.

The use of internal and external sources of data as a foundation for developing a learning plan to continuously improve physicians' performance and practice is viewed as an expectation of medical regulatory authorities in the United Kingdom [77] and in Canada [78], as reflected in the Federation of Medical Regulatory Authorities of Canada's Physician Practice Improvement document. Despite the attempt to transition to an outcome-based approach to enable medical education to be more transparent, accountable, and socially responsive, numerous concerns and challenges have been expressed on this movement. Chief among these criticisms are concerns about inconsistency in the use of language (terms and definitions across the continuum), a drift to reductionist approaches to assessment, the lack of assessment strategies for some competencies or competency domains (for example Health Advocacy or Professionalism), increasing administrative costs, and the burden being placed on clinical faculty without adequate faculty development [79]. In addition, the focus on competence and the emphasis of assessment on 'measurable tasks' may miss the connectedness between the development of competence and identity formation that addresses a broader focus of 'being a physician' rather than simply focusing on 'doing the work of a physician' [80]. The intuitive appeal of emphasising the educational outcomes and how they will be achieved to address changing patient and public expectations and demands from practising as members of complex health systems [81] has served as a recurring critique that competency-based medical education has been 'adopted by consensus in the face of weak empirical evidence' [82, p. 851]. Some have recently expressed competency-based medical education as 'faith-based medical education' [83].

These legitimate concerns should challenge competency-based medical education to contribute to the empirical evidence on the impact of this transition and address the concerns through research and programme evaluation strategies. Research questions such as:

- What are the outcomes that must be achieved by an undergraduate [64] or postgraduate [84] or CME programme?
- How will these outcomes be assessed?
- What are the implications for competency-based medical education on faculty development [85]?
- How does a focus on the competence of an individual align with evolving notions of the collective competence of teams [86]?
- How will competency-based medical education contribute to clinical performance improvement and the achievement of better health outcomes for patients?
- How will competency-based CPD contribute to health systems improvement?

The next decade will provide numerous opportunities to explore the impact of competency-based medical education training on medical students, residents, and physicians in practice.

Physician Professional Development and Maintenance of Certification

Several competency frameworks that apply to Undergraduate Medical Education (UME) and Graduate Medical Education (GME) apply to practising physicians, including those developed by the American Board of Medical Specialties (ABMS)/Accreditation Council for

Graduate Medical Education (ACGME) [87], CanMEDS [63], Good Medical Practice [88], and the Scottish Doctor [61]. These frameworks emphasise that physician competence involves more than medical knowledge, and include the ability to: practise patient-centred medicine, engage in team-based care, practise in systems and participate in improving health care system performance and outcomes, improve patient safety, and function as a steward of health care resources by practising high value, cost-effective care.

Professionals possess knowledge and skills of great importance to society. Society grants professionals autonomy to determine educational standards, self-assess, and self-regulate. In return, through a social compact [89, 90], professionals commit to use their special knowledge and skills for the good of society and place societal needs over individual self-interest.

Board certification is a means of physician self-assessment and self-regulation; a public-facing designation that a physician has the knowledge, skills, and competencies to be deemed a specialist. It has existed in the United States for over 100 years. Initially, board certification was a capstone, diploma-like designation awarded at the completion of graduate medical training after passing a standardised written (and sometimes oral) knowledge assessment.

Family Medicine set a precedent for periodic board recertification in 1969 that instituted a process of periodic reassessment to ensure physicians were keeping up with rapid changes in medical science and medical practice. Subsequent changes in science (e.g. genomics, human immunodeficiency virus) and practice (e.g. changes in the care of peptic ulcer disease, myocardial infarction, new medical technologies) further strengthened the push for periodic recertification in other specialties. The continued increase in complexity of medicine, along with reports about persistent gaps in health care quality [90, 91] and safety [92], prompted questions about the adequacy of an every 6–10 year high stakes, secure examination as the sole mechanism for board specialty recertification.

These concerns, along with changing societal expectations of physicians (from beneficent paternalism toward shared care and patient autonomy) and shifts toward team-based care, helped catalyse the movement to a more continuous process of physician certification. The United States implemented this continuous certification process in 2003 as Maintenance of Certification (MOC) [94]. The ABMS MOC framework, based on the ABMS/ACGME competencies, has four components:

- Professionalism and Professional Standing
- Lifelong Learning and Self-Assessment
- Assessment of Knowledge, Judgement and Skills
- Improvement in Medical Practice.

Each ABMS specialty (Member Board) determines the mechanisms to satisfy each component based on the needs of specialty practice. Physicians spend 3–10 years in graduate training, then often spend 30 years or more in practice. During the course of a contemporary medical career, medical knowledge rapidly expands, new skills are needed, patients become more engaged in their care, and professionals work together across disciplines and teams to improve care. Concomitantly, there will be need for changes in the application of competencies within the contemporary medical career pathway. Maintenance of Certification provides a framework for physicians to continuously improve across these competencies throughout the course of their careers.

Other countries have also developed systems that view specialty certification as ongoing CPD instead of a one-time capstone event at completion of graduate medical training. The Royal College of Canada's Maintenance of Certification Program began in 1991. It currently exists as a three-part framework, incorporating group learning, self-learning (including individual and systems elements), and assessments of knowledge and performance [95, 96]. In 2012, the United Kingdom instituted a process of revalidation, tied to medical licensure rather than to specialty certification per se. Five-year cycles consist of annual reviews of physician portfolios by a senior physician that incorporate assessment of knowledge, skills, and performance; of safety and quality; of communication, partnership, and teamwork; and of maintaining trust [97, 98].

The Maintenance of Certification programmes in the United States and Canada and the revalidation process in the United Kingdom each reinforce the broader-based competency frameworks that are applicable across the continuum from UME to GME to practice, and move beyond medical knowledge. CME is incorporated into each framework. CME remains important, but alone it is insufficient as a mechanism for physician CPD. Given the difficulties, inaccuracies, and known unreliability of unguided self-assessment [37, 54], and that examination performance has been found to be related to measures of clinical performance in several studies [99], objective assessments of knowledge, judgement, and skills remain a core part of MOC. Traditionally, the secure, periodic examination has served this purpose in ABMS MOC; several ABMS boards are piloting (and studying) more frequent, regular, smaller volume, psychometrically sound longitudinal assessments as an alternative to assessing knowledge. More integration between the components of MOC is desirable to help create 'an integrated framework of continuous certification and professional development' [100, pp. 196–198].

Areas of relative strength as well as gaps in knowledge can be objectively identified using psychometrically sound objective tests and assessments. Gaps in performance (which may, in some circumstances, have components of gaps in knowledge) can be identified by patient and peer surveys and practice-based process and outcomes metrics contained in Electronic Health Records and reported to health systems, the government, insurers, or other entities. Continuing medical education activities can be used to address gaps in knowledge ('knows what'), competence ('knows how'), or conditional knowledge ('knows when') [101–103] and to help physicians develop strategies to modify their practices. Quality and performance improvement efforts can then be implemented where physicians (and their teams, as indicated) assess their baseline performance (with the help of pre-existing data), implement their learning and change strategies in practice, use data to assess the effect of their intervention, analyse successes and barriers to improvement, repeat until the desired level of improvement is reached, then follow-up to ensure sustainability of the change in practice. Physicians can then move on to address different gaps. Follow-up objective assessments of knowledge and judgement can then be used to identify whether knowledge gaps are closed and to assess for the emergence of new knowledge gaps.

Physicians face many competing demands for their time and attention in addition to caring for patients. For physicians to be truly engaged, CPD and MOC systems must be aligned with and relevant to practice needs and contexts. MOC systems need to be rigorous and based upon sound educational, systems, and improvement science, yet failure to create relevance and minimise burden risks physicians viewing these activities as 'one-offs', or resisting 'yet one more requirement.' Methods to align elements of MOC with the needs of health care system have the potential to create benefit for both physicians and systems, with systems resources (including provision of data, personnel support for implementing improvement efforts) helping to reduce some of the more administrative aspects of MOC. One such example is the American Board of Medical Specialties Multispecialty Portfolio Program™ [104], which allows physicians to receive MOC credit for meaningful involvement in rigorous, systems-based improvement activities of organisations with mature quality/performance improvement programmes. As of January 2017 more than 11 800 physicians have received MOC credit for over 16 000 improvement efforts in areas targeting specific diseases (hypertension, depression, diabetes mellitus), health promotion and disease prevention, patient safety, transitions between care settings, communication between physicians and patients, and other patient perceptions of the care experience.

Physicians vary in CPD needs and practice contexts; every physician represents an individual CPD 'use case'. While MOC programmes are continually being evaluated to help determine 'what works for whom under what circumstances' [105], these frameworks serve as a means of incorporating competencies beyond medical knowledge in a lifelong framework for physician CPD and improvement [106].

Team-based CPD

As noted above, Olson [2] suggests CPD will continue to become an increasingly team-based, interactive activity. In this chapter, we define team-based CPD as an activity 'undertaken after initial qualification when members of two or more health and/or social care professions learn with, from, and about each other to improve collaboration and the quality of care' [107, p. 143]. Of course, CPD can also be delivered to intraprofessional teams (involving learners from the same profession or speciality). However, given the increasing interest in interprofessional teams in the CPD literature, we only focus on this form of learning in this chapter. For additional reading on interprofessional education across the continuum, see Chapter 14 of this book.

Problems with collaboration and teamwork have been well documented in the literature, repeatedly demonstrating that such problems result in serious compromises in patient safety [108, 109]. Such research indicates that traditional (professionally isolated) approaches to delivering CPD has failed to support the development of abilities required to provide effective care [110, 111]. As a result, the education and training of health professionals, around the world, has re-focused attention on providing collaborative opportunities to develop the attitudes, knowledge, skills, and behaviours needed to work effectively together to deliver safe, high-quality care. Drawing upon this research, health care policymakers have identified the key role of team-based CPD in improving the organisation of health care systems and outcomes [112, 113]. As a result, we have witnessed an increasing number of team-based CPD programmes delivered throughout the world [109, 114].

In relation to the delivery of effective team-based CPD, strategies that enable interactivity are a key requirement. Barr and colleagues [114] have outlined a range of differing types of interactive learning methods than can be employed in this form of learning (see Box 19.3). The literature contains numerous examples of how these interactive learning activities have been employed in a range of team-based CPD activities – see for example Owen and colleagues' study [115] of an IPE programme to improve sepsis care by enhancing health care team collaboration and Luetsch and Rowett's study [116] on interprofessional communication skills development to improve interprofessional collaboration.

Facilitating team-based CPD activities can be challenging, and requires skill, experience, and preparation to deal with the various responsibilities and demands involved. There are a range of attributes required for this type of work: experience of teamwork, commitment to collaboration, understanding of interactive learning methods, knowledge of managing team dynamics, confidence in working with health care teams, flexibility, and approachability and a good sense of humour [117].

Similar to other small-group education, CPD facilitators need to: focus on team formation and team maintenance, create a non-threatening environment, and enable all

BOX 19.3 FOCUS ON: Types of interactive learning methods [114]

Interactive methods	Details
Exchange-based	Seminar-based discussions
Observation-based	Joint visits to patients'/clients' homes
Action-based	Problem-based or case-based learning
Simulation-based	Simulating clinical practice
Practice-based	Team-based clinical placements
e-learning	Online discussions
Blended learning	Combining e-learning with other traditional methods

learners to participate equally, but these aims are more challenging in a team-based context given the history of social and economic inequalities, and friction, that exist between the members of the health and social care professions [109].

During the past decade, a number of systematic reviews have been conducted to examine the growing evidence base related to team-based CPD as well as other forms of collaborative learning [118, 119]. One key feature of this work has been the use of a typology (see Box 19.4) to help classify the outcomes from team-based CPD activities.

As Box 19.4 indicates, there is a range of possible outcomes that can be generated from team-based CPD activities – from recording learner reactions to measuring how such activities can improve patient care.

Using this typology, a review of reviews was undertaken to offer a meta-synthesis of this growing evidence base [120]. Following a comprehensive search, six reviews were identified. Collectively, these reviews reported the effects of more than 200 studies spanning over 40 years – with around 60% of the included studies reporting team-based CPD activities. Of these studies, the synthesis revealed that team-based CPD was delivered in a variety of acute, primary, and community care settings and addressed a range of different clinical chronic or acute conditions. While different groups of professions were involved in this form of CPD, medicine and nursing were the core participants. It was found that quality improvement principles were often drawn upon within a team-based CPD activity. In general, interprofessional education programmes used formative assessments of learning, typically using assessment techniques in the form of individual written assignments and/or joint/team presentations which provided a collective account of learners' team-based experiences [120, 121].

BOX 19.4 HOW TO: Use a typology to classify the outcomes from team-based CPD [118, 119]

Outcome	Details
Level 1: Reaction	Learners' views on the team-based CPD experience
Level 2a: Modification of attitudes/perceptions	Changes in reciprocal attitudes or perceptions between professional groups
Level 2b: Acquisition of knowledge/skills	Gains of knowledge and skills linked to collaboration and teamwork
Level 3: Behavioural change	Transfer of an individual's learning from their team-based CPD activity to their professional performance
Level 4a: Change in organisational practice	Wider changes in the organisation and delivery of care resulting from the team-based CPD activity
Level 4b: Benefits to patients	Improvements in health or well-being of patients resulting from the team-based CPD activity

The synthesis found that most studies reported that team-based CPD resulted in positive learner reactions, positive changes in learner perceptions/attitudes, positive changes in views of teamwork, and positive changes in learner knowledge and skills of collaboration (Levels 1, 2a, and 2b). In contrast, fewer studies reported outcomes related to individual behaviour, organisational practice, or patient benefit. Of these studies, they tended to report positive change in individual practitioners' interactions (Level 3), positive changes to team-based referral practices/working patterns or improved documentation (i.e. guidelines, shared records) which supported improvements to the organisation of care (Level 4a), and/or positive changes to clinical outcomes (e.g. infection rates, clinical error rates), patient satisfaction scores, and/or length of patient stay. Recently, this work was updated and found eight additional reviews of interprofessional education [121]. Despite a growth of the evidence contained in this newer review, the key results in essence remained unchanged.

As discussed above, team-based activities have expanded across the globe over the past few decades in response to failures of teamwork that have undermined patient quality and safety. As the evidence syntheses have shown, this type of CPD can have positive outcomes in relation to participants' reactions, attitudes, knowledge/skills, as well as some improvement for collaborative behaviours, and organisational practice and patient benefit. Future investment in team-based CPD must be based on the growing evidence base. Further research needs to focus on addressing current gaps in knowledge relating to the longer term impact of team-based CPD. Economic analyses are also needed to help indicate the costs and benefits related to this form of CPD.

Conclusion

The model for physicians' professional development and lifelong learning has been shifting from the narrower traditional view of CME toward a broader and more encompassing model of CPD which includes multiple components and aspects such as SDL, continuous MOC programmes, competency, and team-based models [1]. In this chapter, we discussed four different elements that are considered pertinent, in our views, to the current model of CPD for physicians and other health care professionals. SDL is a learning process that is considered to be one of the most appropriate strategies in CPD for physicians [9] as the key advantages of this process allow physicians to have control over their learning and manage their time more efficiently. Adoption of a consistent and evidence-based model across medical education and consideration of the different factors that may influence SDL are crucial to development of effective CPD programmes generating positive outcomes.

Moreover, this chapter attempts to place current CPD approaches within the history of competency-based medical education. Based on the need to align education with societal and patient needs, competency-based medical education emphasises the abilities (outcomes) that learners are expected to achieve, instead of the amount of time spent in a specific context. In terms of CPD, competency-based CPD places a greater emphasis on demonstrating how learning impacts performance, quality of care, and health outcomes. To address some of the concerns and challenges that have been expressed about competency-based medical education, further research and programme evaluation are crucial to build empirical evidence on the impact of the transition to this approach.

Furthermore, maintenance of physician certification programmes has also shifted from a traditional one-time certification toward an ongoing lifelong process of maintaining certification. In recent years, many countries have established continuous MOC programmes to enable physicians to remain current with rapidly changing medical knowledge and changes in practice. The American, Canadian, and British MOC programme examples provided in this chapter demonstrate the promulgation of broader-based competency frameworks that are more than just medical knowledge focused, as they also combine CPD with patient safety and quality improvements activities.

Finally, we discussed the shift in focus of CPD from an individual activity to a team-based one. We emphasised the critical importance of providing health care professionals with collaborative opportunities to develop attitudes, knowledge, skills, and behaviours needed to work in teams to deliver safe and high-quality care now and into the future. Evidence of positive outcomes generated by team-based CPD activities included: positive learner reactions, perceptions, and attitudes; positive changes in views of teamwork, knowledge, and collaboration skills; as well as improved collaborative behaviours, organisational practice, and patient benefit.

This chapter has summarised what we believe to be four current and future key elements of CPD: self-directed learning; the relationship between maintenance of certification and professional development; the competencies across the medical education continuum; and team-based education in CPD. Collectively, these areas of CPD represent what we see as the need to view the concept of CPD in a broad and relational way. CPD sits within a nexus between the individual physician and health care team and the process and relationship between certification and professional development. Therefore, we believe that a macro view of the continuum of medical education needs to be taken and maintained by CPD scholars and education practitioners whereby CPD activities can be informed by undergraduate and postgraduate training and the workplace context in which CPD is delivered.

Acknowledgement

We would like to acknowledge Professor Scott Reeves, a friend, colleague, mentor and a global leader in health professions education for his co-authorship of this chapter. His kind manner, intellectual generosity, warmth and support will be sorely missed.

References

1 Sargeant, J., Bruce, D., and Campbell, C.M. (2013). Practicing Physicians' needs for assessment and feedback as part of professional development. *Journal of Continuing Education in the Health Professions* 33: S54–S62.

2 Olson, C.A. (2012). Twenty predictions for the future of CPD: implications of the shift from the update model to improving clinical practice. *Journal of Continuing Education in the Health Professions* 32: 151–152.

3 ABIM Foundation. American Board of Internal Medicine, ACP-ASIM Foundation. American College of Physicians-American Society of Internal Medicine, European Federation of Internal Medicine (2002). Medical professionalism in the new millennium: a physician charter. *Annals of Internal Medicine* 136: 243–246.

4 Blank, L. (2002). Medical professionalism in the new millennium: a physicians' charter. *Lancet* 359: 520–522.

5 The Royal College of Physicians and Surgeons of Canada. Accreditation of Continuing Professional Development (CPD) Activities. http://www.royalcollege.ca/rcsite/cpd/accreditation-continuing-professional-development-cpd-activities-e (accessed 3 March 2017).

6 American Board of Medical Specialties. Board Certification and Maintenance of Certification. http://www.abms.org/board-certification (accessed 3 March 2017).

7 Merkur, S., Mladovsky, P., Mossialos, E., and McKee, M. (2008). Do lifelong learning and revalidation ensure that physicians are fit to practise?. http://www.euro.who.int/__data/assets/pdf_file/0005/75434/E93412.pdf (accessed 28 June 2018).

8 White, C.B., Gruppen, L.D., and Fantone, J.C. (2014). Chapter 15: Self-regulated learning in medical education. In: *Understanding Medical Education: Evidence, Theory and Practice* (ed. T. Swanwick). Chichester, West Sussex: Wiley.

9 Murad, M.H., Coto-Yglesias, F., Varkey, P. et al. (2010). The effectiveness of self-directed learning in health professions education: a systematic review. *Medical Education* 44: 1057–1068.

10 Brockett, R. and Hiemstra, R. (1985). Bridging the theory-practice gap in self directed learning. *New Directions in Adult Continuing Education* (25): 31–40.

11 White, C.B. et al. (2014). Self-regulated learning in medical education. In: *Understanding Medical Education: Evidence, Theory and Practice* (ed. T. Swanwick). Chichester, West Sussex: Wiley.

12 Knowles, M.S. (1975). *Self-Directed Learning: A Guide for Learners and Teachers*. New York: Association Press.

13 Candy, P.C. (1991). *Self-direction for lifelong learning*. Wiley.

14 Naumann, D.N., Mascaro, C., and Smith, K. (2015). Self-Directed Learning in CME: A Literature Review.doi: 10.13140/RG.2.1.3729.3922; https://www.researchgate.net/publication/280723052_Self-Directed_Learning_in_CME_A_Literature_Review (accessed 25 June 2018).

15 Bernabeo, E.C., Holmboe, E.S., Ross, K. et al. (2013). The utility of vignettes to stimulate reflection on professionalism: theory and practice. *Advances in Health Sciences Education* 18: 463–484.

16 Hess, B.J., Johnston, M.M., Iobst, W.F., and Lipner, R.S. (2013). Practice-based learning can improve osteoporosis care. *Journal of the American Geriatric Society* 61: 1651–1660.

17 Fox, R.D. and Bennet, N.L. (1998). Learning and change: implications for continuing medical education. *BMJ* 316: 466–468.

18 Davis, N., Davis, D., and Bloch, R. (2008). Continuing medical education: AMEE education guide no 35. *Medical Teacher* 30: 652–666.

19 Ebell, M.H., Cervero, R., and Joaquin, E. (2011). Questions asked by physicians as the basis for continuing education needs assessment. *Journal of Continuing Education in the Health Professions* 31: 3–14.

20 Ebell, M.H. and Shaughnessy, A. (2003). Information mastery: integrating continuing medical education with the information needs

of clinicians. *Journal of Continuing Education in the Health Professions* 23: S53–S62.
21 Sargeant, J., Mann, K., Van Der Vleuten, C., Metsemakers, J. (2008). Directed self-assessment: practice and feedback within a social context. *Journal of Continuing Education in the Health Professions* 28: 47–54
22 Galbraith, R.M., Hawkins, R.E., and Holmboe, E.S. (2008). Making self-assessment more effective. *Journal of Continuing Education in the Health Professions* 28: 20–24.
23 Duffy, D.F., Holmboe, E.S., Duffy, F.D., and Holmboe, E.S. (2006). Self-assessment in lifelong learning and improving performance in practice. *Journal of the American Medical Association* 296: 1137.
24 Eva, K.W. and Regehr, G. (2007). Knowing when to look it up: a new conception of self-assessment ability. *Academic Medicine* 82: S81–S84.
25 Shaw, T., Long, A., Chopra, S., and Kerfoot, B.P. (2011). Impact on clinical behavior of face-to-face continuing medical education blended with online spaced education: a randomized controlled trial. *Journal of Continuing Education in the Health Professions* 31: 103–108.
26 Gyorki, D.E., Shaw, T., Nicholson, J. et al. (2013). Improving the impact of didactic resident training with online spaced education. *ANZ Journal of Surgery* 83: 477–480.
27 Kerfoot, B.P. and Baker, H. (2012). An online spaced-education game for global continuing medical education: a randomized trial. *Annals of Surgery* 256: 33–38.
28 Brown, J.A., Beaser, R.S., Neighbours, J., and Shuman, J. (2011). The integrated Joslin performance improvement/CME program: a new paradigm for better diabetes care. *Journal of Continuing Education in the Health Professions* 31: 57–63.
29 van de Wiel, M.W.J., van den Bossche, P., Janssen, S., and Jossberger, H. (2011). Exploring deliberate practice in medicine: how do physicians learn in the workplace? *Advances in Health Sciences Education* 16: 81–95.
30 Kruger, J. and Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology* 77: 1121–1134.
31 Goulet, F., Hudon, E., Gagnon, R. et al. (2013). Effects of continuing professional development on clinical performance. *Canadian Family Physician* 59: 518–525.
32 Abadel, F.T. and Hattab, A.S. (2013). How does the medical graduates' self-assessment of their clinical competency differ from experts' assessment? *BMC Medical Education* 13: 1.
33 Epstein, R.M. (1999). Mindful Practice. *Journal of the American Medical Association* 282: 833.
34 Curran, V., Fleet, L., Simmons, K. et al. (2016). Exploratory study of rural Physicians … self-directed learning experiences in a digital age. *Journal of Continuing Education in the Health Professions* 36: 284–289.
35 Li, S.T., DA, P., Co, J.P.T., and West, D.C. (2010). Successful self-directed lifelong learning in medicine : a conceptual model derived from pediatric residents. *Academic Medicine* 85: 1229–1236.
36 Stuart, E., Sectish, T.C., and Huffman, L.C. (2005). Are residents ready for self-directed learning ? A pilot program of individualized learning plans in continuity clinic. *Ambulatory Pediatrics*. *Program* 5: 298.
37 Davis, D.A., Mazmanian, P.E., Fordis, M., Van Harrison, R., Thorpe, K.E. and Perrier, L. (2006). Accuracy of physician self-assessment compared with observed measures of competence a systematic review. *Journal of the American Medical Association* 296:1094–1102
38 Jeong, D., Presseau, J., ElChamaa, R. et al. (2018). Barriers and facilitators to self-directed jearning in continuing professional development for physicians in Canada: A scoping review. *Academic Medicine*. *April* 10: 2018. doi: 10.1097/ACM.0000000000002237.
39 Schumacher, D.J., Englander, R., and Carraccio, C. (2013). Developing the master learner. *Academic Medicine* 88: 1635–1645.
40 Bennett, E.E., Blanchard, R.D., and Hinchey, K.T. (2012). AM last page: applying Knowles' andragogy to resident teaching. *Academic Medicine* 87: 129.
41 Norman, G.R. (1999). The adult learner. *Academic Medicine* 74: 886–889.
42 Tough, A. (1979). *The Adult's Learning Projects: A Fresh Approach to Theory and Practice in Adult Learning*. Toronto: The Ontario Institute for Studies in Education.
43 Scott, K.W. (2006). Self-directed learners' concept of self as learner: Congrouous autonomy. *International Journal of Self-Directed Learning* 3: 1–13.
44 Davis, S.A. (2006). Enhancing self-directed learning in new professionals. *ERIC Education Resources Information Center* 245–249.
45 Mamary, E. and Charles, P. (2003). Promoting self-directed learning for continuing medical education. *Medical Teacher* 25: 188–190.
46 Candy, P.C. (1995). Physician teach thyself: the place of self-directed learning in continuing medical education. *Journal of Continuing Education in the Health Professions* 15: 80–90.
47 Davis, D., Andrew, M., and Oxman, A.H.B. (1995). Changing physician performance : a systematic review of the effect of continuing medical education strategies. *Journal of the American Medical Association* 274: 700–705.
48 Jennett, P., Jones, D., Mast, T. et al. (1994). The characteristics of self-directed learning. In: *The Physician as Learner: Linking Research to Practice* (D.A. Davis and R.D. Fox ed.), 47–66. Chicago, IL: American Medical Association.
49 Kaufman, D.M. and Mann, K.V. (2010). Teaching and learning in medical education: how theory can inform practice. *Understanding Medical Education* 16–36.
50 Kaufman, D.M. (2003). ABC of learning and teaching in medicine: applying educational theory in practice. *BMJ* 326: 213–216.
51 Brydges, R., Dubrowski, A., and Regehr, G. (2010). A new concept of unsupervised learning: directed self-guided learning in the health professions. *Academic Medicine* 85: S49–S55.
52 Violato, C. and Lockyer, J. (2006). Self and peer assessment of pediatricians, psychiatrists and medicine specialists: implications for self-directed learning. *Advances in Health Sciences Education* 11: 235–244.
53 Mann, K., van der Vleuten, C., Eva, K. et al. (2011). Tensions in informed self-assessment: how the desire for feedback and reticence to collect and use it can conflict. *Academic Medicine* 86: 1120–1127.
54 Eva, K.W. and Regehr, G. (2005). Self-assessment in the health professions: a reformulation and research agenda. *Academic Medicine* 80: S46–S54.
55 Ward, M., Gruppen, L., and Regehr, G. (2002). Measuring self-assessment: current state of the art. *Advances in Health Sciences Education* 7: 63–80.
56 Ainoda, N., Onishi, H., and Yasuda, Y. (2005). Definitions and goals of 'self-directed learning' in contemporary medical education literature. *Annals of the Academy of Medicine, Singapore* 34 (8): 515–519.
57 Mazmanian, P. and Feldman, M. (2011). Theory is needed to improve education, assessment and policy in self-directed learning. *Medical Education* 45: 324–326.
58 Hiemstra, R. and Brockett, R.G. (2012). Reframing the Meaning of Self-Directed Learning: An Updated Model. *Proceedings of the 54th Annual Adult Education Research Conference*, pp. 155–161.
59 Brocket, R.G. and Hiemstra, R. Self-direction in adult learning: perspectives on theory. *Research and Practice* 1991: 18–33.
60 Flexner, A. (1910). Medical Education in the United States and Canada: A Report to the Carnegie Foundation for the Advancement of Teaching. Carnegie Foundation Bulletin No. 4. New York.

61 The Scottish Doctor (2011). Learning Outcomes for The Role of the Doctor within the Health Service. http://www.scottishdoctor.org/node.asp?id=1100000000 (accessed 23 March 2017).

62 Laan, R.F.J.M., Leunissen, R.R.M., and van Herwaarden, C.L.A. (2010). The 2009 framework for undergraduate medical education in the Netherlands. *Tijdschr voor Med Onderwijs* 29: 1–6.

63 Frank, J.R., Jabbour, M., Fréchette, D., et al. (2005). The CanMEDS 2005 Physician Competency Framework. Better Standards. Better Physicians. Better Care. Royal College of Physicians and Surgeons of Canada.

64 Harris, P., Snell, L., Talbot, M., and Harden, R.M. (2010). Competency-based medical education: implications for undergraduate programs. *Medical Teacher* 32: 646–650.

65 McGaghie, W.C., Miller, G.E., Sajid, A.W., and Telder, T.V. (1978). Competency-based curriculum development on medical education: an introduction, Public Health Papers, vol. 68. Geneva: World Health Organization http://www.who.int/iris/handle/10665/39703 (date accessed 25 June 2018.

66 General Medical Council (1993). *Tomorrow's Doctors*. doi: 10.1016/S0140-6736(09)62088-2

67 Smith, S.R. (1999). AMEE guide no. 14: outcome-based education: part 2-planning, implementing and evaluating a competency-based curriculum. *Medical Teacher* 21: 15–22.

68 Frank, J.R., Snell, L.S., Ten, C.O. et al. ed. (2010). Competency-based medical education: theory to practice. *Medical Teacher* 32: 638–645.

69 Swing, S.R. (2007). The ACGME outcome project: retrospective and prospective. *Medical Teacher* 29: 648–654.

70 Boelen, C., Dharamsi, S., and Gibbs, T. (2012). The social accountability of medical schools and its indicators. *Education for Health (Abingdon)* 25: 180–194.

71 Royal College of Physicians and Surgeons of Canada (2017). Competence by Design: The Rationale for Change. http://www.royalcollege.ca/rcsite/cbd/rationale-why-cbd-e (accessed 24 April 2017).

72 Frank, J., Snell, L., and Sherbino, J. (2015). CanMEDS 2015 Physician Competency Framework, Ottawa.

73 Royal College of Physicians and Surgeons of Canada: Competence By Design (2017). Rationale for a Change to Competency Based Continuing Professional Development. In: Royal College of Physicians and Surgeons of Canada White Paper on Continuing Professional Development. http://www.royalcollege.ca/rcsite/cbd/cpd/competency-cpd-white-paper-e (accessed 24 April 2017).

74 Sargeant, J., Armson, H., Chesluk, B. et al. (2010). The processes and dimensions of informed self-assessment: a conceptual model. *Academic Medicine* 85: 1212–1220.

75 Bing-You, R.G. and Trowbridge, R.L. (2009). Why medical educators may be failing at feedback. *Journal of the American Medical Association* 302: 1330–1331.

76 Boud, D. and Molloy, E. (2013). *Feedback in Higher and Professional Education: Understanding it and Doing it Well*. doi: 10.4324/9780203074336

77 Southgate, L. and Pringle, M. (1999). Revalidation in the United Kingdom: general principles based on experience in general practice. *BMJ* 319: 1180–1183.

78 Federation of Medical Regulatory Authorities of Canada (2016). Physician Practice Improvement. http://fmrac.ca/physician-practice-improvement (accessed 24 April 2017).

79 Hawkins, R.E., Welcher, C.M., Holmboe, E.S. et al. (2015). Implementation of competency-based medical education: are we addressing the concerns and challenges? *Medical Education* 49: 1086–1102.

80 Jarvis-Selinger, S., Pratt, D.D., and Regehr, G. (2012). Competency is not enough: integrating identity formation into the medical education discourse. *Academic Medicine* 87: 1185–1190.

81 Harden, R.M. (1999). AMEE guide no. 14: outcome-based education: part 1-an introduction to outcome-based education. *Medical Teacher* 21: 7–14.

82 Morcke, A.M., Dornan, T., and Eika, B. (2013). Outcome (competency) based education: an exploration of its origins, theoretical basis, and empirical evidence. *Advances in Health Sciences Education* 18: 851–863.

83 Whitehead, C.R. and Kuper, A. (2017). Faith-based medical education. *Advances in Health Sciences Education* 22: 1–3.

84 Iobst, W.F., Sherbino, J., Ten, C.O. et al. (2010). Competency-based medical education in postgraduate medical education. *Medical Teacher* 32: 651–656.

85 Dath, D. and Iobst, W. (2010). The importance of faculty development in the transition to competency-based medical education. *Medical Teacher* 32: 683–686.

86 Hodges, B.D. and Lingard, L. (2012). *The Question of Competence : Reconsidering Medical Education in the Twenty-First Century*, 1e. Ithaca: Cornell University Press.

87 American Board of Medical Specialties Board Certification. Based on Core Competencies. http://www.abms.org/board-certification/a-trusted-credential/based-on-core-competencies (accessed 23 March 2017).

88 Good Medical Council. Good Medical Practice. http://www.gmc-uk.org/guidance/index.asp (accessed 3 March 2017).

89 Cruess, R.L. and Cruess, S.R. (1997). Teaching medicine as a profession in the service of healing. *Academic Medicine* 72: 941–952.

90 Cruess, S.R. and Cruess, R.L. (2005). The medical profession and self-regulation: a current challenge. *Virtual Mentor* doi: 10.1001/virtualmentor.2005.7.4.oped1-0504.

91 Institute of Medicine (US) (1994). America's Health in Transition: Protecting and Improving Quality. The National Academies Press.

92 Institute of Medicine (US) Committee on Quality of Health Care in America (2001). *Crossing the Quality Chasm. A New Health System for the 21st Century*. doi: 10.17226/10027

93 Kohn, L., Corrigan, J., and Donaldson, M. (2000). *To Err Is Human: Building a Safer Health System*. doi: 10.17226/9728

94 American Board of Medical Specialties (2014). Standards for the ABMS Program for Maintenance of Certification (MOC): For implementation in January 2015 https://www.abms.org/media/1109/standards-for-the-abms-program-for-moc-final.pdf (accessed 28 June 2018).

95 Campbell, C.M. and Parboosingh, J. (2013). The Royal College Experience and plans for the maintenance of certification program. *Journal of Continuing Education in the Health Professions* 33: S36–S47.

96 The Royal College of Physicians and Surgeons of Canada. Framework of Continuing Professional Development Activities. http://www.royalcollege.ca/rcsite/cpd/moc-program/moc-framework-e (accessed 23 March 2017).

97 Archer J, De Bere SR (2013) The United Kingdom's experience with and future plans for revalidation. *Journal of Continuing Education in the Health Professions* 33:S48–S53.

98 General Medical Council Working with Doctors, Working for Patients: Registration and licensing - An introduction to revalidation. http://www.gmc-uk.org/doctors/revalidation/9627.asp (accessed 23 March 2017).

99 Hawkins, R.E., Lipner, R.S., Ham, H.P. et al. (2013). American Board of Medical Specialties Maintenance of certification: theory and evidence regarding the current framework. *Journal of Continuing Education in the Health Professions* 33: S7–S19.

100 Price, D. (2017). Maintenance of certification, continuing professional development, and performance improvement. In: Rayburn W., Turco M., and Davis D.A. (eds.). Continuing Professional Development in Medicine and Health Care: Better Education, Better Patient Outcomes. Philadelphia, PA: Wolters Kluwer. p.191–202.

101 Moore, D.E., Green, J.S., and Gallis, H.A. (2009). Achieving desired results and improved outcomes: integrating planning and assessment

throughout learning activities. *Journal of Continuing Education in the Health Professions* 29: 1–15.
102 Bruning, R.H., Schraw, G.J., Norby, M.M., and Running, R.R. (2004). *Cognitive psychology and instruction, Long-term memory: Structures and models*, 4e. Chapter 3. Upper Saddle River, NJ: Pearson/Merrill/Prentice Hall.
103 Bruning, R.H., Schraw, G.J., Norby, M.M., and Ronning, R.R. (2004). *Cognitive psychology and instruction, Sensory, short-term, and working memory*, 4e. Chapter 2. Upper Saddle River, NJ: Pearson/Merrill/Prentice Hall.
104 American Board of Medical Specialties (ABMS) Multi-Specialty Portfolio Program. http://mocportfolioprogram.org (accessed 23 March 2017).
105 Clarke, A., Pawson, R., and Tilley, N. (1998). Realistic Evaluation. *British Journal of Sociology* doi: 10.2307/591330.
106 Price, D., Havens, C., and Bell, M. (2011). Continuing professional development and improvement to meet current and future continuing medical education needs of physicians. *Continuing medical education: Looking back, planning ahead*. Lebanon, NH: Dartmouth College Press. p. 1–14.
107 Reeves, S. (2009). An overview of continuing interprofessional education. *Journal of Continuing Education in the Health Professions* 29: 142–146.
108 The Joint Commission (2016). Sentinel Event Data: Root Causes by Event Type 2004–2015.
109 Reeves, S., Lewin, S., Espin, S., and Zwarenstein, M. (2010). *Interprofessional Teamwork for Health and Social Care*. Oxford: Wiley-Blackwell. doi: 10.1002/9781444325027
110 Kitto, S.C., Bell, M., Goldman, J. et al. (2013). (Mis)perceptions of continuing education: insights from knowledge translation, quality improvement, and patient safety leaders. *Journal of Continuing Education in the Health Professions* 33: 81–88.
111 Kitto, S., Goldman, J., Etchells, E. et al. (2015). Quality improvement, patient safety, and continuing education. *Academic Medicine* 90: 240–245.
112 Frenk, J., Chen, L., Bhutta, Z.A. et al. (2010). Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. *Lancet* 376: 1923–1958.
113 Institute of Medicine (2015). *Measuring the Impact of Interprofessional Education on Collaborative Practice and Patient Outcomes*. Washington, DC: The National Academies Press.
114 Barr, H., Koppel, I., and Reeves, S. (2005). *Effective Interprofessional Education: Argument, Assumption and Evidence*. Blackwell Pub [in association with] CAIPE.
115 Owen, J.A., Brashers, V.L., Littlewood, K.E. et al. (2014). Designing and evaluating an effective theory-based continuing interprofessional education program to improve sepsis care by enhancing healthcare team collaboration. *Journal of Interprofessional Care* 28: 212–217.
116 Luetsch, K. and Rowett, D. (2016). Developing interprofessional communication skills for pharmacists to improve their ability to collaborate with other professions. *Journal of Interprofessional Care* 30: 458–465.
117 Reeves, S., Pelone, F., Hendry, J. et al. (2016). Using a meta-ethnographic approach to explore the nature of facilitation and teaching approaches employed in interprofessional education. *Medical Teacher* 38: 1221–1228.
118 Reeves, S., Perrier, L., Goldman, J. et al. (2013). Interprofessional education: effects on professional practice and healthcare outcomes (update). *Cochrane Database of Systematic Reviews* (Art. No.: CD002213. doi: 10.1002/14651858.CD002213.pub3.
119 Reeves, S., Fletcher, S., Barr, H. et al. (2016). A BEME systematic review of the effects of interprofessional education: BEME guide no. 39. *Medical Teacher* 38: 656–668.
120 Reeves, S., Goldman, J., Burton, A., and Sawatzky-Girling, B. (2010). Synthesis of systematic review evidence of Interprofessional education. *Journal of Allied Health* 39: 198–203.
121 Reeves, S. and Zierler, B. (2017). An updated synthesis of review evidence of interprofessional education. *Journal of Allied Health* 44: 0–5.

Part 3

Assessment and Selection

20 How to Design a Useful Test: The Principles of Assessment

Lambert W. T. Schuwirth[1,2,3,4] *and Cees P. M. van der Vleuten*[5]

[1]Medical Education, Flinders University, Adelaide, Australia
[2]Innovative Assessment, Maastricht University, Maastricht, The Netherlands
[3]Medical Education, Chang Gung University, Taoyuan, Taiwan
[4]Medicine (Education), Uniformed Services University for the Health Professions, Bethesda, MD, USA
[5]School of Health Professions Education, University of Maastricht, Maastricht, The Netherlands

KEY MESSAGES

- Good assessment requires a variety of methods; no single method can test the whole of medical competence and performance.
- In a good assessment programme the total is more than the sum of the parts. To achieve this:
  - each instrument must be chosen on the basis of its strengths and weaknesses
  - all decisions are based on rational arguments and/or scientific underpinning
  - rational decisions are made about to how to combine results
  - the whole programme is part of a total quality management system.
- Modern assessment in medical education is a matter of good educational design and not exclusively a psychometric measurement problem.

Introduction

Designing a system of assessment is not easy. Many different instruments have been described in the medical education literature, each with its own advantages and disadvantages. The volume of published research on assessment is huge, and research findings often contradict long-established practices based on tradition and intuition. Furthermore, the discipline has its own jargon and set of statistical concepts. In amongst all this, the responsibility regularly falls on medical teachers to set up fair assessment systems. The aim of this chapter is to explain the basic concepts behind assessment and thus help reduce this 'burden'. But, it is not a cookbook providing assessment 'recipes'; instead, we hope that after reading this publication the medical educator will be better equipped to make informed choices in matters of assessment.

What is Assessment?

Although assessment is generally acknowledged to be fundamental to the educative process, there are many different constructions of the term, ranging from assessment as a certification procedure leading to a pass/fail decision, to assessment as an evaluative or feedback action in education. Here we will use 'any *purported* and *formal* action to obtain information about the competence and performance of a candidate' as our working definition of assessment.

Assessment is never undertaken without a specific purpose. This *purpose* can be summative and/or formative. Summative means that the assessment has been conducted for decision-making or certification purposes, such as deciding who is admitted, progresses, or qualifies. Formative relates to the feedback function of assessment or, more precisely, how the assessment informs the students about their performance. Sometimes it is argued that formative and summative assessment purposes should not be mixed in one assessment, but in educational settings we tend to advise otherwise. Purely summative assessment ('pass' or 'fail' only) does not help students to plan their study and to improve. If a student fails, s/he does not know what s/he needs to work on in order to pass at the next attempt. On the other hand, purely formative assessment with no consequences at all is often not taken seriously.

How to Choose the Best Approaches to Assessment

A first step in setting up an assessment system is to choose the most appropriate methods. This is not an easy task. Over the past few decades, a wide variety of instruments

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

have been developed and used in medical education, and often authors have claimed their instrument to be the best for all purposes. We may safely assume, however, that each method has its own strengths and weaknesses, and these must be weighed carefully against the desired purposes of the assessment. A testing authority whose purpose is licensing may demand different strengths from the methods from, say, a medical school. Several criteria that can be used in this evaluative process have been described in the literature [1]. The most popular ones are examined below.

Reliability

The reliability of a method pertains mostly to the *reproducibility* of its results, that is, how often the same result is obtained. Let us illustrate this with three examples.

Example 1: John scores 83% in a test. If John were given a different, but similar, test – a so-called parallel test – would he again score 83%?

Example 2: John is the best-scoring student in his class. Harry comes next, Peter's score is slightly lower, and Jim's score is the lowest. If this class of students were given a parallel test, would the rank ordering of scores be the same as in the first test?

Example 3: In another test, John passes. He scores 83% and the cut-off score for the test was 50%. If John were given a parallel test, would he pass again?

These examples illustrate three operational definitions of reliability. The first example is from a domain-referenced perspective. The test here aims to measure how much of a certain domain John knows. It may be intuitively clear that to be able to conclude that John's knowledge is *exactly* 83% you need a very fine – highly reliable – measurement. The second example – from a norm-referenced perspective – is somewhat less demanding. Here we are not interested in whether John's score was exactly 83%, but only whether his score was better than Harry's, Peter's, and Jim's. The measurement can be slightly coarser. In the third case, demands on the measurement are even lower.

Now let us turn the concept around. If we have the test results of a class of students, depending on which of the three conclusions we want to draw, the test is more or less reliable. This is all very well in theory, but we do not have a parallel test, so how can we establish the parallel test score? A typical approach is to retrospectively divide the test into two halves randomly and treat them as parallel tests. Most of the well-known reliability tests in classical test theory, for example, Kuder–Richardson and Cronbach's alpha, build on this approach (see Box 20.1).

Why is it so hard to design a perfectly reliable test? The largest source of unreliability is sampling error. For pragmatic reasons, a test usually comprises a sample of questions from the whole domain of possible questions. But because items can differ in difficulty, and because different students find different items difficult – John does better on the myocardial infarction items and Harry does better on the arrhythmia items – sampling error arises. The same applies equally to other assessment modalities, such as the examiner's choice of questions in oral or short-essay examinations. So if sampling error is such a major issue, it follows that brief assessments and assessments based on the judgement of only one examiner are very likely to be unreliable.

There is one final issue to discuss in relation to reliability, namely the relationship between reliability and objectivity. A common misconception is that subjective assessment is always unreliable and objective assessments are always reliable [2]. The following two illustrations demonstrate that this is not necessarily the case. The first example is a one-item multiple-choice examination. This is a so-called objective test, but as it has only one item, it is not reliable. On the other hand, suppose you were to write ten pieces of music and take, at random, ten pieces written by Mozart, and submit them all to a panel of experts who have to assess the musical artistry of the composer. The panel would reach the decision that Mozart is the better composer. It would not matter if we took another sample of your own compositions or another sample of Mozart's compositions or even another sample of experts, the decision would still be the same. Musical artistry, however, is not 'objective'; it is highly subjective. Yet the decision of the panel is highly reproducible and is thus reliable.

In conclusion, reliability is a matter of careful sampling. It relies on a sufficiently large sample through all possible sources of error, for example, items, examiners, and test occasions. But reliability is not the whole story.

Validity

Validity is defined as the extent to which the sort of competence the assessment claims to assess is actually being assessed. This is not always easy to demonstrate, and demonstrating the validity of a particular test method is a matter of collecting evidence from different sources and perspectives.

The first step is to define exactly for what purpose the method is valid. In the way that a thermometer is a valid instrument to measure temperature and only temperature (not weight, for example), an assessment method is valid only for a certain aspect of competence. Any claim by a test developer that their instrument is a valid instrument for medical competence is therefore to be regarded with suspicion. Validity has been defined in numerous different ways. For the purposes of this chapter, we will classify the different varieties of validity into two categories, as follows [3, 4].

Content Validity (also referred to as Direct Validity)

Content validity refers to the judgemental type of evidence collected in the validation process. This can be expert judgement about the specific item construction, the necessary faculty development for judges, etc. One specific element in content validation is blueprinting. This is done because an examination must be optimally representative of the whole testable domain. An examination on cardiology should not be composed only of items on myocardial infarction. To ensure adequate coverage, a test is typically constructed according to a blueprint. The blueprint is a table of specifications in which the test maker determines how many items per subject or category are to be asked. A blueprint may have multiple dimensions (e.g. organ systems by disciplines by cognitive level). A related issue in content validity is the relevance of the items. Only relevant items contribute to the content validity of the examination.

BOX 20.1 FOCUS ON: The measurement of reliability

Reliability is a central concept in test theory, as examiners and candidates want a test that gives a similar result on different occasions with different candidates. Test–retest analysis assesses the same candidate's performance in the same test on two separate occasions, whereas parallel testing assesses the same candidate using tests containing different questions that are thought to be equivalent. In each case, reliable tests should give very similar marks on both occasions. Mathematically, psychometricians *model* the responses of candidates in different ways, with three broad approaches.

Classical Test Theory (CTT)

This has been around for half a century or more. The main assumption is that a candidate has a true ability, or score, but because of measurement error, for whatever reason, the candidate does not obtain the exact same score, even if an identical test is used on two occasions. The similarity of marks on the two occasions is used to calculate the reliability, which becomes higher as measurement error becomes lower. A similar approach can be used for comparing parallel tests.

CTT works best with multiple-choice tests, where all candidates answer identical questions. Its major drawback is that reliability calculated from one group of candidates cannot be extrapolated easily to other groups. Apparent reliability of a test can be much inflated by including a few outstandingly bad candidates. CTT reliability is also very misleading as a description of the accuracy of a cut-off score (pass/fail boundary) in high-stakes examinations.

Generalisability Theory (GT)

CTT does not work well with typical clinical examinations, where not all candidates can see all patients or be seen by all examiners. There is therefore variability due to examiners (e.g. hawks and doves) and clinical scenarios, as well as case-specificity, with some candidates doing better with some types of case than others. GT generalises CTT to include such components. A measure equivalent to reliability ('generalisability') can be calculated; in effect, how similar would a candidate's mark be with different examiners and different scenarios?

GT allows sophisticated calculations of the effects of different types of exam – 'what if' questions – such as, 'Would the exam be more generalisable with more stations and fewer examiners per station'? The untested, perhaps unrealistic, assumption is that examiners and candidates will behave in precisely the same way in such new situations. As with CTT, GT estimates of generalisability cannot be easily extrapolated to new situations.

Rasch Modelling (RM) and Item–Response Theory (IRT)

IRT is a more complex variant of RM. RM starts with the assumption that every test item has a particular difficulty, with the probability of a candidate answering an item correctly depending on the item's difficulty and the ability of the candidate. Mathematical modelling then allows the item difficulties to be calculated, as well as candidate abilities. A new test comprised of previously used items from an item bank used in different combinations can have its reliability calculated before the test is used. The reliability can also be calculated for candidates of differing ability (e.g. fourth-year students rather than third-year). IRT is an extension of RM that not only calculates the difficulty of items, but also calculates discriminative value and chance-guessing rates.

IRT requires large data sets and is best used for large-scale examinations with over a thousand candidates. RM, however, is robust with small numbers. RM cannot handle clinical situations with different scenarios and examiners, but these can be handled with Facet theory. The major problem for RM and IRT is not the conceptual basis, which is rigorous, powerful, and realistic, but the mathematical and statistical skills of most biomedical examiners.

Construct Validity (also referred to as Indirect Validity)

Another category of validity arguments is based on the 'behaviour' of assessment scores; do they align with our expectations about the type of competence we want to assess? A competency can be seen as a construct or latent trait; a personal psychological characteristic that cannot be observed directly but which is assumed to exist. A typical example is 'intelligence'. We assume this construct to have certain characteristics: more intelligent people can learn faster, have better memory skills, and are better able to solve problems than less intelligent people. If we were to design a new test to measure intelligence, we would hope that people who learn faster outperform people who learn more slowly, and demonstrating this would contribute to the demonstration of validity of our new test *for the construct intelligence*. So the scores on the intelligence test would 'behave' according to our – theoretical – assumptions. Applying this principle to tests for medical problem solving, for example, means that for a test to have good construct validity, it would be necessary for people who solve problems more expertly to outperform those who are less good problem solvers.

There are many other types of validity support: weak correlations between two tests that are supposed to measure different constructs; strong correlations between two tests that are assumed to measure the same construct; and so on. Although they are sometimes labelled as different forms of validity, in effect they all contribute to the evidence for the validity of a test for a specific construct.

The concept of validity has engendered a wide variety of viewpoints. Some claim that only construct validity is worthwhile and dismiss content and face validity as just a collection of opinions [5]. Some take the usability of the test and the relationship between the test and its user as an important element of validity [6]. Others take a more holistic view, and we think that Kane (quoting Messick) strikes an intelligent balance in taking a definition of validity as 'the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment' [7]. 'Inferences' and 'actions' bring us neatly to considering the educational consequences of the test [8].

Educational Impact

The saying 'Students don't do what you expect, students do what you inspect' epitomises the educational impact of assessment. Although there is still a need for more empirical underpinning of this notion, the emerging research does show that assessment has a major impact on how students study and learn [9–13]. Often the standard response of teachers is to blame students for this strategic approach to learning, but a more rational response is for the assessment developer to capitalise on this behaviour. It is just normal human behaviour and it is not going to go away; indeed, we are all susceptible to these kinds of external motivators. More importantly, the driving influence of assessment is a powerful tool to ensure that students learn what, and how, teachers want them to learn. To maximise the effectiveness of this alignment, it is important to realise that assessment influences student learning in several ways: through its content, its format, the scheduling of examinations, and the regulatory structure of the assessment programme. It is more important to realise that this driving force is optimised by carefully combining behaviourist (stakes, rewards, and punishment) and constructivist (meaning making, translation of results to helpful learning activities, etc.) factors in the assessment and this is why purely summative assessment (ignoring the meaning making) and purely formative assessment (ignoring the behaviourist factors) do not work.

The influence of *content* is obvious. Topics that come up repeatedly in the examinations will be perceived by students as the most important. When it comes to *format*, the literature gives different views. Some studies seem to indicate that students prepare differently for different formats, especially multiple-choice versus open-ended examinations [14]. Others indicate that this is only the perception of students, whereas in fact they prepare similarly [15]. As a rule of thumb, we advocate using a variety of formats in an assessment programme. This way, students will not become used to one type of preparation.

A typical problem of *scheduling* concerns, for instance, annual examination periods. If several examinations are held during the same few weeks, students will not be able to prepare adequately for all of them. They will then have to select strategically the examinations for which they will prepare well and those for which they will not. From the faculty's point of view this is a waste of resources, as effort is put into constructing a high-quality examination that is not taken seriously. Better, therefore, to spread examinations rather than to cluster them. The assessment programme *as a whole* defines academic success. The credit points attributed to each part, the way scores are combined, the minimum number of points needed to progress, and so on, all define how students will study or prepare for the examinations [9–11]. Programmes with continuous assessment will promote continuous learning; pinpoint assessments will promote cramming behaviour. Because educational impact is such a major effect of assessment, it is often also seen as a part of validity, specifically termed 'consequential validity' [9–11].

Cost-effectiveness

This seems an easy and obvious issue, but the practice sometimes differs from the theory. For example, a prerequisite for a cost-effective assessment programme is an explicit description of its goals, both in terms of *what* is to be assessed and *how* it is to be assessed. Only then can an evaluation be conducted into whether the programme is optimally cost-effective. The most important factors hampering cost-efficiency are misconceptions based on tradition and intuition, poorly supportive infrastructure, and lack of collaboration.

Typical examples of the first are interviews, oral examinations, and open-ended questions. The literature on unstructured interviews for selection for medical schools converges on the conclusion that their predictive validity is unacceptably low [16–18]. Their use therefore constitutes a cost-ineffective assessment procedure. The recent development of multiple mini-interviews (MMIs) seem to address these concerns successfully and thus are not only an improvement in terms of reliability and validity but possibly have a huge impact on cost-effectiveness [19]. The same applies to many unstructured oral examinations [20]. Such orals have poor reliability and are often used to test simple factual knowledge, which can better be tested using more cost-effective methods. This is not to say that there is no place for oral examinations in an assessment programme, but they should be used only where they have an added value over other formats [21]. Similarly, there is an almost inextinguishable belief that open-ended questions test higher-order cognitive skills and multiple-choice questions do not. Again, the literature converges on the notion that in this matter the question format is quite unimportant, but the content of the question is [22–24]. Of course, in an assessment programme there may be indications to use open-ended questions – for example, if creativity or spontaneous generation of the answer is essential. In other cases there are more cost-effective alternatives.

Typical examples of poor infrastructure support are the lack of good item banking and the absence of a centralised management and administrative support for the logistics and administration of examinations. The consequences are that 'expensive' scientific staff members carry out work that could be done just as well – or even better – by administrative personnel.

The benefits of collaboration may perhaps be obvious. Many medical schools have comparable curricula with comparable end goals, and sharing test material would be a way to reduce costs, because it would mean not everybody has to reinvent the wheel [25]. This is not to say that collaboration is easy. Careful planning, commitment of all partners, and some sort of pre-investment are all needed to make collaboration successful [26].

A final argument about cost-effectiveness relates to assessment in an educational context. The primary goal of that context is to educate students to become competent, safe, and independent practitioners. So here, assessment programmes that are exclusively designed to provide summative pass/fail decisions instead of providing rich information to all stakeholders about how best to learn and become a competent, safe, and independent practitioner are inefficient; a huge amount of resources are spent on only a dichotomous decision. Programmatic assessment for learning seeks to address this issue and make the assessment more useful from a learning/education perspective without ignoring the purpose of ensuring competent, safe, and independent graduates [27–29].

See Box 20.2 for more on programmatic assessment.

Acceptability

Setting up an assessment system cannot be done in a vacuum. Sometimes a careful balance between what may be scientifically and educationally superior and what is acceptable to the stakeholders has to be struck. Even the best assessment method is useless if teachers and students will not subscribe to it. Some even go so far as to incorporate this issue into the argument for a test's validity [6]. This is not that strange a point of view because an instrument is valid only if it is used properly.

Popular Assessment Instruments

No single assessment instrument is perfect, and no single instrument can test all aspects of medical competence and performance. Each instrument has its strengths and weaknesses. A good assessment programme should therefore include a purposive and deliberate combination of various methods, each selected for a specific purpose, and each with its indications and contraindications [27, 30–32]. This section describes the major strengths and weaknesses of different groups of commonly used assessment methods, using the five criteria described above: reliability, validity, educational impact, cost-efficiency and acceptability.

Written Assessment Instruments

There are many different written assessment instruments currently in use. Some are case-based, some are context-free. Some use open-ended questions, some use multiple-choice questions.

BOX 20.2 FOCUS ON: Programmatic assessment

In clinical medicine we use certain principles to diagnose a disease. Firstly, we don't rely on a single 'test' result. Secondly, we don't combine information from different results within the same instrument, for example compensating for a high glucose level with a low sodium level, just because they are both laboratory values. And, finally, we don't make high-stakes decisions based on only one test. Instead, we use multiple instruments and their associated results to make a clinical decision. An example would be the diagnosis of diabetes where you would have a history of thirst and fatigue, physical examination revealing poor wound healing and weight loss, and a laboratory test of a high blood glucose.

The same principles are the basis for programmatic assessment.

In programmatic assessment various assessment instruments are used and information about a student's progress and achieved competence is combined across instruments. Poor performance on an abdominal examination would therefore not be compensated by good performance on a shoulder examination. But it could be combined with the abdominal anatomy portion of a written test to determine whether the poor performance was due to insufficient anatomy knowledge or whether it was just a lack of technique. Logically, the remediation advice to individual students would be different based on the outcome of this combination.

In programmatic assessment, the assessment takes place on a continuous basis and all information is collected and collated in a dossier or portfolio. Periodically, all the information is reviewed and evaluated in order to make a decision about a student's progress. If there is little information available, only low-stakes decisions can be made, but when high-stakes decisions have to be made they will be based on rich information collected over a longer period of time.

So, in short, the essential features of programmatic assessment are:

- The use of multiple instruments where each instrument informs about the progress in several competency domains and each competency domain is informed by multiple instruments.
- Information is combined based on the content of the information (poor abdominal examination with poor abdominal anatomy knowledge) rather than on the format of the assessment (poor abdominal examination combined with good shoulder examination).
- Decision moments about a student's progress are disconnected from assessment moments and only taken when sufficient and sufficiently rich information about a student's progress and competence has been obtained.
- Information is collected and collated in a dossier/portfolio and decisions about a student's progress are made based on human judgement, which is scaffolded by transparency, audit trails, and safety nets in the organisation. These judgements are made by faculty with specific training and expertise in assessment.

As a general rule, the amount of time it takes to answer a question has a negative impact on the reliability of the test. As you may recall, a test can only sample from the whole domain of possible items, and, as such, the sample must be large enough to be reliable. It follows then that the more items a test contains, the more reliable it is likely to be. This immediately places open-ended questions at a disadvantage because they require more answering time. So, essays are generally less reliable *per hour of testing time* than short-answer questions.

Where reliability is a fairly straightforward issue in written assessment, validity is much more complex. There are some popular beliefs about the validity of different types of question. For example, it is often thought that open-ended questions test higher-order cognitive skills and that multiple-choice questions can test only factual knowledge. This is a widespread misconception: the question *format* is quite unimportant with respect to validity, whereas the question *content* is very important [22–24]. So, *what* you ask is important, not how you record the answer. Of course, some contents do not fit certain formats. It is best not to ask items that require the spontaneous and creative generation of possibilities in multiple-choice format, whilst items requiring a selection from a finite list of realistic options are best not asked in an open-ended format.

Thus, careful consideration of content is essential. A further and important distinction relates to context. Context-rich items contain a case description and questions that ask for (essential) decisions or an evaluation of the problem. Typical examples of these are extended-matching items or key-feature approach items [33–35]. Context-free items do not have a case description and simply ask for general knowledge. Context-rich approaches test application of knowledge and problem solving, but context-free items do not [36]. Both, however, can be equally important aspects to assess; one is not superior to the other.

The idea that open-ended questions test superior cognitive skills over multiple-choice questions is also widespread among students. Although this may not actually be true, it will still influence students' perception and their learning. Using a variety of methods sends a clear message to students that they have to master the subject matter, irrespective of the assessment format.

Another aspect of educational impact is the influence that assessment has on test makers. If, for example, all tests have to be in multiple-choice format, examiners may construct only items that fit this format, and questions requiring spontaneous generation of the answer could be under-represented. On the other hand, if all items are of the open-ended format, the examiners may be burdened with the high workload of correcting tests and may start asking simple questions that are easy to score, which also means that important aspects may be neglected.

The use of multiple-choice-based assessments is highly cost-effective. They may be slightly more difficult to produce, but the use of Optical Mark Reading scanners certainly makes them easier to score. Nowadays, software is available to assess online. This is a particularly relevant consideration in medical schools with a large number of students per year class. Every pound, euro, or dollar can be spent only once; therefore, money spent on unnecessarily expensive assessment methods cannot be spent on improving education. So, from the viewpoint of cost-efficiency, it is best to use open-ended questions only if more efficient formats will not suffice.

There might be a wealth of scientific literature proving that test format is unimportant, but sometimes beliefs may be so strong that stakeholders cannot be convinced. In such cases arguments will be used claiming that multiple-choice assessments are too easy, make students lazy, and are not worthy of an academic environment, and that real life and real practice are not simply a question of selecting options from a list. Such arguments may seem incorrect from a sheer psychometric/rationalist view point, but they may be very strongly embedded in the teachers' and the institution's core values. Research into how so-called naïve beliefs are formed and maintained demonstrates that they are very hard to change and although this research originally related to physics concepts (gravity for instance) it does have pertinence to beliefs around education as well [37–39]. So, it is important then to consider whether it is useful to contradict them. Sometimes these values are very strong, and it may be better to aim for high levels of acceptability of an assessment system first, and to postpone the 'battle'. Energy may be better spent on good teaching and good-enough assessment, and, more importantly, any test can be valid only if it is used correctly. For this, it has to be acceptable to all stakeholders. You can read more about written assessment in Chapter 21.

Objective Structured Clinical Examinations and Simulated Patients

Objective Structured Clinical Examinations (OSCEs) and simulated patient (SP)-based examinations have become very popular for the assessment of (practical) skills [40, 41]. Both are based on a series of structured cases that must be addressed by the candidate. In an OSCE, a candidate enters a series of different rooms or stations in sequence. In each room there is a specified assignment (e.g. perform a resuscitation or take the blood pressure), a simulated patient or manikin, and an examiner with a checklist or rating scale. The candidate has to complete the assignment and his or her performance is scored against the checklist or rating scale. After a fixed period of time, a signal is given and the candidate proceeds to the next station.

OSCEs and SP-based examinations were developed in response to unstructured observations in practice. They are cleverly developed in that they address the inherent unreliability of observed practice in three ways. First, by adding some structure to the observations, they become more reliable. Second, by keeping each of the observations short (the original OSCEs had 5-minute stations), many different observations can be made per hour, thus enabling wider and more effective sampling. Third, by having the candidate move from station to station, such assessments also sample across different examiners. The 'hawks' will be compensated for by the 'doves', or better still all candidates will be examined by the same panel of examiners. The second of these issues – that of sampling across many cases – is the most important, because the biggest threat to reliability is having too small a sample. The many reliability studies on OSCEs have demonstrated this over and over again. One of the practical implications is that it is better to have more stations with one examiner per station than fewer stations with two examiners per station [42]. Despite the clever reliability approach, as a rule of thumb, OSCEs still require a minimum average of at least 2–3 hours of testing per candidate to achieve an acceptable reliability for summative decisions.

With respect to validity, two issues are of overriding importance: the length of the stations and the use of checklists versus global rating scales. One might be inclined to think that longer stations, that is, longer than 5 minutes, may be more (content) valid but less reliable, but this is not necessarily the case. Longer cases contain more information than shorter ones, and there seems to be an optimum balance between the length and the number of cases in an OSCE. Therefore, it is generally best to adapt the length of the stations to the content of the case, so durations of stations may be designed to vary from 5 to 20 minutes [43].

Checklists are detailed lists of behaviours, and they describe precisely the actions to be taken – for example, 'washes hands', 'puts left hand on the sternum of the patient'; whereas rating scales allow for more interpretation by the examiner, describing in broad terms only the skills to be performed – for example, 'explores patient's concerns', 'comes to the correct conclusions'. So should checklists or rating scales be used?

One would be inclined to think that as checklists are more structured they would be more reliable, but this is not always the case [44]. The choice of whether to use checklists or rating scales should be made mainly on the basis of the type of skill to be assessed. Technical skills, such as taking blood pressure or performing resuscitation, can easily be tested with checklists, whereas more complicated skills, such as short patient contacts, seem to be better tested with rating scales [45]. Many medical schools use short stations with checklists for technical skills in the more junior year groups, and integrated longer stations with rating scales in the more senior years.

In general, OSCEs are taken very seriously by students and have a high impact on student learning behaviour. This provides both a risk and an opportunity. The risk comes with detailed checklists. Even if they are not handed out officially, a 'black market' in old checklists may develop, and memorising these may be a successful study strategy for students. Memorising rating scales is less useful. The use of rating scales in OSCEs induces a study behaviour that is aimed more at practising the skill, and an opportunity here is to allow some time – about 2 minutes – at the end of each station for specific feedback. When the OSCE is solely for certification purposes, this is not desirable. In such cases, optimising the reliability (and thus the sampling) is more essential. However, most OSCEs are held within the educational environment of medical schools and can provide a wonderful opportunity for learning.

Unfortunately, OSCEs are very expensive to run. They require extensive resources and good logistics. It is therefore important to use OSCEs effectively, and using an extensive part of the OSCE time to explore general knowledge is not efficient. This does not mean that no knowledge should be tested in an OSCE, but that the knowledge tested should be background knowledge and should have a direct relationship to the case.

OSCEs are widely accepted and popular throughout the world. The only threat to their acceptability is when OSCEs are used to test highly technical skills with very detailed checklists. They then tend to become monkey tricks, and examiners may feel that their expertise is not being used or valued. A more detailed exploration of OSCEs and other structured assessments of clinical competence can be found in Chapter 23.

Oral Examinations

Oral examinations come in various forms, ranging from the completely unstructured to the highly structured, case-based examination. The oral examination has tended to be discarded, being considered too unreliable and too expensive. However, recently views on the oral have shifted in a more favourable direction [21, 46], and the prevailing view is that there is room for an oral examination in an assessment programme, as long as it is used in the correct way and for the correct purpose.

This does imply, though, that to be acceptable, the oral examination must be constructed in such a way that it achieves sufficient reliability. For this, some structure – but not too much – is needed; a situation analogous to the OSCEs, where detailed checklists do not lead to higher reliabilities than rating scales. Reliability can be further enhanced by asking about a good variety of topics rather than homing in on only one. If multiple examiners are used, it is also better to 'nest' cases within examiners instead of using panels [20].

There is a largely unsubstantiated view that orals are somehow more valid than written examinations. You may recall that in considering the validity of an assessment, the content is more important than the format. Often, the answers to oral questions require a good deal of (factual) knowledge, which can be assessed just as well by less expensive methods. If orals are to be used correctly, they have to be aimed at examining aspects that cannot be examined otherwise, such as hypothesis generation, explanations, and the transfer of principles through various contexts. Another misapprehension is the perceived advantage that orals offer in following through on a certain topic – 'to see if they really understand it'. In such cases, the law of diminishing returns rules: the first question on a topic may prove a rich source of information about the candidate's competence, but the tenth question will add virtually nothing new.

Of course, just as with any type of examination, students will prepare strategically for the oral. In doing this they often try to get the lenient examiners or find out what the examiners' hobby horses are. It is therefore best to use a rotational approach in which students rotate from examiner to examiner, each of whom addresses a different, but predetermined, case or topic.

Despite the high costs of oral examinations, they are widely accepted in assessment programmes, and it is the experience at many institutes that the expert judgement emanating from orals is less frequently the subject of appeals and litigation as are written assessments.

Workplace-based Assessment

Recent developments have placed the assessment in the authentic medical context once again. Where OSCEs were developed to test students in a simulated environment, instruments such as mini-CEX (Clinical Evaluation Exercise), DOPS (direct observation of procedural skills), OSATS (objective structured assessment of technical skill), and 360° feedback assess the candidate in their professional environment [47–49].

Mini-CEX uses a generic form with rating scales, which an examiner uses to score the student's performance during a patient encounter. Items include history-taking, physical examination, professionalism, clinical judgement, counselling, organisation, and efficiency, and an overall impression. The competence of the candidate is assessed by a series of direct observations. Another workplace-based approach, 360° feedback, uses standard lists of rating scale items, which are sent to various parties. So, not only colleagues, but also nursing staff and patients, are sent a form and asked to give their rating on the items. Examples of items include the following:

- ability to diagnose patients
- ability to use evidence-based medicine approaches in practice
- verbal communication with colleagues.

At first sight, both of these methods may seem like a step back to the old in-training judgements, but this is not the case. The mini-CEX draws on the lessons learnt from the OSCEs about structure and sampling. Observations in practice, such as mini-CEX, can become reliable, as long as the examiners have some criteria; a sufficient number (roughly seven to ten) of different cases are observed; and there is more than one observer [50, 51]. The added value is that what is being assessed is more authentic than in a simulated environment. In many simulated assessments, certain symptoms cannot be simulated, but in real contexts these symptoms are present. A relatively recent promising development in this field combines mini-CEX with entrustable professional activities by asking the assessors to mark their level of trust in the candidate when s/he is performing clinical activities rather than the more abstract construct of 'competence' [52, 53]. Such approaches acknowledge that for any observation-based assessment to be valid it has to rely on the assessor's assessment literacy [54] and use a jargon familiar to them.

The 360° feedback method is not based on direct observations, but on a judgement in retrospect. Normally, this is ill advised, since such judgements tend to be unreliable; but two aspects remedy this problem. First, many different people are asked, so a broad sample of observations is obtained. Second, judges are not asked for a global impression but to give a judgement about specific aspects of someone's strengths and weaknesses. In both cases judges need to be trained to use the instrument correctly and the instrument has to be designed to support them in making a judgement.

Apart from being 'measurements' of performance, these instruments are also intended to provide the candidate with extensive feedback. Recent research shows that narrative feedback is much more informative than ratings or scores [55]. This is essential to influence learning behaviour. So rather than being measurement-only instruments, they are also educational tools aimed at improving the performance of candidates. Furthermore, the supervisor cannot complete them if he or she has not observed the candidate directly. So, in those educational environments where direct observation is not part of the educational culture, the use of these instruments may help to change the educational routine.

Using workplace-based instruments well does not have to be time-consuming, especially in those situations where frequent observation and feedback are already part of the teaching culture. However, it is important not to make the forms too long, as this will make such an approach less acceptable to users. Workplace-based assessment is considered in more detail in Chapter 22.

Portfolios

The word 'portfolio' is a container term used to describe all kinds of educational tools. From the assessment point of view, there are two approaches that it may be useful to discuss here, as follows:

- portfolio as an instrument to measure the *reflective ability* of the candidate
- portfolio as an instrument to *collate assessment information* from various sources.

In both cases the portfolio contains a 'dossier' and an 'analysis'. The analysis contains a self-assessment of strengths and weaknesses, learning goals, and a learning plan.

The reflective portfolio focuses on self-assessment; it is used to assess the extent to which the candidate's self-assessment demonstrates a good reflective ability. The second portfolio approach collates all assessment information about the student. The analysis section is used to evaluate current performance and to plan future learning. This approach is best compared to a patient chart, in which information from various sources, such as laboratory data, imaging data, and results from history-taking and physical examination, is collected, but it also contains a regular evaluation about the well-being of the patient and a plan for further diagnostic and therapeutic actions. So, the portfolio becomes not only an assessment but also an educational instrument.

It is difficult to say anything definitive about the reliability and validity of portfolios. Studies calculating reliability in the traditional psychometric way, using either generalisability theory or inter-rater agreement measures such as Cohen's Kappa, report moderate reliabilities at best [56]. Other authors suggest an organisational approach to rendering the portfolios dependable, using concepts from qualitative research methodologies, such as benchmarking, peer evaluation, member checking, and stepwise replication [57]. In such cases, reliability cannot be expressed as a number but must be derived from the carefulness of the decision-making processes on learner progression.

The validity of the portfolio approach requires further study. It is apparent that conventional construct validity methods do not apply here, so other methodologies need to be developed. Moreover, since portfolios are used for so many purposes, content validity cannot easily be established.

It may also be obvious that portfolios are expensive. They are often *perceived* as time-consuming not only to produce, but also to assess, especially if more than one judge has to assess each portfolio. It is tempting then to try to produce a simple scoring list or rubric to increase efficiency, but this only serves to trivialise the assessment. Training of assessors and using global criteria for judging the portfolio is a better approach, and it may be more efficient to set up a procedure in which multiple judges are used only if there is doubt about the result, with only a very limited number of judges used in all clear cases [57].

You can read more about portfolios in Chapter 18.

Computers in Assessment

There are many different ways in which computers can play a role in assessment. The most obvious is computerised administration. But there are other, and more important, roles for computers in assessment, and these are discussed briefly below.

Administrative Support

Item banks can be very powerful in supporting quality control. Indeed, this aspect is often more important than their role in enabling the re-use of old items. Attempts to build a complete item bank from which items can be drawn at random often prove unsuccessful. There are two reasons for this. First, an examination is more than a randomly generated set of items, even if the individual items are of good quality. There is always an extra quality-control step needed to ensure that the combination of items is good. A second reason is concerned with the nature of medicine and other health sciences. In these disciplines, things evolve quite rapidly, quickly rendering items outdated. Also, ideas about what constitutes a good item may evolve. Item banks are, therefore, more useful in tracking an item in the quality-control process. They also provide the opportunity to scan the domain coverage quickly, so that production of redundant items is prevented and under-represented subjects can be completed with specific new items.

There are many good commercial products available for item banking. For simple purposes it is also possible to use standard database software with self-produced scripts, shells, or queries. Developing complete high-brow software systems is always more time-consuming than one would expect. When an item bank system is needed, the best approach is to determine carefully the needs of the organisation, or the functional and operational specifications, and then determine which of the available software can meet these needs sufficiently.

Test Analysis

In the quality control of tests, computers can be used to evaluate test results. The most well-known application is a standard item analysis with p-values, a-values, item–total or item–rest correlations, and the calculation of reliabilities. The p-value represents the proportion of students answering the question correctly. As such, it is an indication of how difficult this item was f*or this particular group of candidates*. A p-value of 1.00 means that every candidate answered the item correctly, whereas a p-value of 0.00 means that nobody gave the correct answer. The a-values indicate the proportion of candidates choosing each option in a multiple-choice question and, as such, are an indicator for the attractiveness of each distracter. The item–total and item–rest correlations indicate the extent to which the item was answered correctly by the high achievers on the test and answered incorrectly by the low scorers on the test. Standard statistical software, such as SPSS, often allows for such analyses, and such structured item analyses can be very valuable and may have a major impact on the quality of tests.

Computerised Testing

There are many obvious advantages of computerised testing. Hand-scoring is not needed, the results can be calculated immediately, and data files for further analysis are easily available. Also, there are no added costs for reproduction of test booklets and answer sheets. Audio and video clips can be added and can help to improve the content validity of the test.

But there are also downsides to computerised testing. First, open-ended questions are difficult to score automatically and may need hand-scoring or at least verification of the computer scoring procedures. Although hand-scoring of typed text may be easier than that of – sometimes illegible – handwriting, this still nullifies the advantages of immediate results and availability of data files. This may lead test developers to use multiple-choice type questions exclusively. Second, if there are many more candidates than there are computers available, equivalent test forms have to be developed. Although corralling of students is an option to prevent unwanted information exchanges, this is possible only to a limited extent. Producing extra equivalent tests is more expensive than reproduction costs. Third, the necessity of a systems administrator for the computer network adds to the costs. Finally, without sufficient back-up systems, computer or network problems may disturb the test administration. Although with ongoing technology these problems may become rarer, and despite the fact that things can also go wrong with paper-and-pencil administration, problems are still more likely to occur with computerised testing. So, before deciding to use computers to administer a test that could also be a pen-and-paper exercise, it is important to consider all the pros and cons very carefully.

Assessment Possibilities Unique to Computer Testing

There are some interesting and potential possibilities presented by computer-based assessment that are unique to the format and not merely logistical advantages. We discuss three of them here – namely, real-time simulations, sequential testing, and computer-adaptive testing (CAT).

Real-time simulations are useful to test the management of cases in which time is essential in real life, such as emergency medicine. As a formative tool it can also be helpful to demonstrate to the candidate how much time they took to solve the case and where a gain in efficiency could be made. For summative testing it tends to complicate things, as one cannot simply add up response time and proficiency. So, to come to a pass/fail decision, one has to find a valid way of combining these different qualities into one single score, and this is not easy.

Sequential testing is an approach that enables a more efficient use of time and resources in assessment. It basically comes down to administering a short screening test to all candidates. Based on the reliability of this test, a (95%) confidence interval can be calculated around the pass/fail score. For every candidate whose score is outside this interval, there is sufficient certainty ($p \leq 0.05$) to say that his or her score is either a pass or a fail. The remainder of the students must answer an additional set of questions. The scores of these are added to their scores in the first part of the test. This way, a longer test is presented only to those candidates where there is doubt about their passing or failing. Such an approach is feasible only if the scores and the confidence interval can be calculated quickly and an additional test is available on request [58].

A further development is CAT. This approach is based on a so-called calibrated item bank – an item bank in which the precise difficulty of all items is known beforehand. For such a calibration, classical test theory (CTT) is often not sufficient. A more complicated statistical approach – IRT – is used (see Box 20.1). Unlike CTT, the use of IRT allows the difficulty and discriminative power of items to be estimated, regardless of the specific group of candidates. In CAT, the computer selects an item of moderate difficulty for the candidate. When the candidate answers the item correctly, the computer selects a slightly more difficult item, and when the answer is incorrect, an easier item is selected. This process is repeated either until a specified number of items have been answered or until a certain level of preciseness of the test is reached [59]. In the former, the precision proficiency estimate varies across candidates (but in the majority of cases is better than a standard test); in the latter, the precision is fixed, but the number of items needed may vary from candidate to candidate (in the majority of cases this will be fewer than in a standard test). Although CAT is a wonderful concept, the statistical requirements for achieving a well-enough calibrated item bank are quite heavy, requiring considerable pre-testing of all items [59].

Combining Assessment Methods

It is currently generally accepted that in order to obtain a complete picture of someone's competence and performance, one assessment instrument is not enough; a variety of well-chosen instruments is needed [27, 30, 31]. How then should assessment methods be combined? Essentially, there are two approaches, one quantitative, the other qualitative.

Quantitative methods:

- compensatory
- partially compensatory
- conjunctive.

Qualitative methods:

- expert judgement
- explicit procedures.

In a quantitative combination, results are somehow translated into numerical values. These values are then combined in a compensatory, partially compensatory, or conjunctive manner.

Compensatory means that the results of the tests are averaged or summed and that the average or sum needs to be above the pass/fail score, regardless of the scores on the individual tests. For example, averaging the two sets of marks 4/10 and 8/10, and 2/10 and 10/10, gives a result of 6/10 in both cases. A model in which every test result contributes to a total score with a certain percentage – for example, test 1 accounts for 30% and test 2 for 20%, etc. – is also a compensatory model. Compensatory models often result in high reliabilities for the final decision, because such decisions are made on the basis of many items within multiple tests held on different occasions, so sampling across many sources of error. The major downside, however, is that compensatory models may induce a minimalist study strategy. Some students may have such high scores on previous tests that they tend not to take the later tests (and the related courses) seriously.

A *partially compensatory* model corrects for this. Here, the scores can be averaged, but for each test there is an absolute minimum score, and if this is not reached, the student has to repeat the test. This is a compromise in that the combined reliability is somewhat lower than in a completely compensatory model, but the negative educational impact is also diminished.

A (completely) *conjunctive* model requires that the student achieves a score above the pass/fail score in all the tests in order to gain an overall pass. Such an approach stimulates students to take all the individual tests (and courses) seriously and to study hard for all of them, but it is less reliable overall. In every test there is a probability of a false-negative result, that is, a student who fails who should in fact have passed. Such failure is then largely due to measurement error, rather than incompetence. And, as each failure leads to a consequence (which is the case in conjunctive but not in compensatory models), in fact the false-negative results of all the individual tests are combined, which results in a lower overall reliability.

The results of some assessments are simply not quantitative in nature. So, it is not possible to add them up to form a total score. Although it is common practice to convert the results of qualitative assessments into numerical scores, we want to warn against this as a methodologically incorrect practice [60]. This is perhaps best illustrated by an analogy to medical practice. You do not add your first impression of 'very sick patient' to a sodium level of $133\,\mathrm{mmol\,l^{-1}}$ – these two pieces of information need to be combined qualitatively. The same applies to different observations with results such as 'performed extremely well' and 'good bedside manner'. These cannot be combined in a quantitative way but need to be evaluated qualitatively. Such a combination requires – again similar to medical practice – expert judgement and careful procedures. Good examples of such approaches are the General Medical Council's practice performance procedures and some portfolio assessments [57, 61, 62].

Currently there seems to be a reappraisal for qualitative ways of combining information as it becomes clearer that a multi-faceted construct such as 'competence' cannot be assessed numerically only, much like a patient's health cannot be determined by lab values only [63].

Standard Setting

Perhaps the most heavily debated issue in assessment is the issue of standard setting. It is the cut-off score that determines the consequences of the assessment, that is, who passes and who fails. It is an important issue because often quite small changes in cut-off scores represent substantial changes in the numbers of students who pass and fail. The Holy Grail is therefore the *true* cut-off score. Unfortunately, like its mythical counterpart, there is no such thing. The literature describes a wide variation of methods [64], each of which has its own specific purposes, and a distinction is usually made between relative and absolute standards. There is no one single-best standard setting method for all tests, but there is probably a most appropriate method for each individual test in a specific context.

No matter which method of standard setting is used, it will always be arbitrary, as there will always be assumptions made about the required level. Relative methods are based on assumptions about the stability of the mean competence of large groups of students. Especially large year classes of medical students appear to be comparable across cohorts and universities in many, but not all, countries. Absolute methods are based on assumptions about the required level of competence, the teaching the candidates have received, and the end goals of the curriculum. In every case, therefore, there must at least be an explicit rationale for the decisions about the standard setting method. This is sometimes expressed in the aphorism 'standard setting may be arbitrary, it may never be capricious'. Any standard must therefore be:

- explicable, through the rationales behind the decisions made
- defensible, to the extent that it can assure the stakeholders about its validity (an issue in this may be 'due diligence', that is, demonstrating that good effort was put into setting the standard)
- stable, as it is not defensible that the standards vary from year to year [65].

It should be realised that items, and therefore tests as a whole, vary tremendously in difficulty. Unless one controls with advanced psychometric techniques, the variability in test difficulty is sizeable. This means that any standard setting method that does not reconcile this variability in difficulty is less credible. The most widely used fixed standard, e.g. performance should be more than 55%, is the least credible standard. In the choice of a standard setting method, cost should factor in. Some methods are much more expensive than other methods.

Chapter 24 examines the subject of standard setting in more detail and looks at the different approaches currently used by test developers.

Future Directions

A section on future directions is always a dangerous one to write since so-called 'future developments' may, with hindsight, appear to have been flukes. Still, we would like to make some predictions.

The change from defining educational outcomes in constructs – such as knowledge, skills, attitudes, and problem-solving skills – to actions, as described in Miller's pyramid (knows, knows how, shows how and does), and the further change from there into more or less complex tasks that require the timely availability of relevant knowledge, skills, attitudes, and problem-solving ability, the so-called competencies or entrustable professional activities [52], must have an influence on our way of thinking about assessment.

Also, the emergence of new assessment methods such as mini-CEX, 360° feedback, and portfolio, in which the main goal is *not* to add up the individual items to give a total score, must have an influence on our way of thinking about assessment. We see the following three main developments here.

1 Assessment will be less viewed as an external measurement of the results of the educational process but more as an integral part of the process. Currently, it is still fairly common to take students out of the authentic educational context to be tested on their competence. Assessments such as mini-CEX take the assessment back into the authentic educational context. Current approaches to assessing professional behaviour even acknowledge that it is impossible to evaluate this outside the authentic context. This leads, in our view, to a second development.

2 Assessment is no longer seen exclusively as a psychometric measurement problem, but more as an educational design problem. This implies that the purpose of assessment is not merely to determine whether a candidate is up to standard, but more how the information about the candidate's competence can best be used to tailor the teaching or the courses to individual needs. So, instead of striving for a standardised curriculum with standardised testing, it will entail a development of tailored assessment with flexible curricula. This, in turn, may lead to a third development.

3 Standard psychometric approaches to issues such as fairness and defensibility of examinations will have to be expanded with other measures. For example, basic assumptions underlying the standard psychometric approach, such as stable and generic constructs, homogeneity of the universe (e.g. the total universe of possible items), and assumption of local independency of the observations, cannot always be met. Moreover, some modern instruments aim precisely at being locally dependent observations (mini-CEX, longitudinal testing), acknowledge the heterogeneity of the universe (360° feedback), and acknowledge the non-existence of traits in competence (portfolio). Assessment becomes, then, more of a diagnostic process (much like health care), where multiple methods are used in a bespoke way to diagnose 'dyscompetence' rather than leading every student through the same standardised processes [63]. This does not make the issues of defensibility, fairness, and carefulness go away but will require different – statistical – models [66] or strategies to validate qualitative data [67].

Epilogue

Designing assessment programmes and selecting the best instruments for each purpose is not easy. To complicate matters further, medical education is a rapidly evolving discipline. This may easily lead to a perception that assessment is not scientific because the truths of yesterday are obsolete and will be replaced with new ones. We would argue au contraire; any evolving discipline questions truths critically and scientifically, and this is a strength rather than a weakness. Medical education does not differ from medicine. In this respect – what was true when we were students often no longer holds true today. The purpose of this chapter is to guide the reader through the field of assessment of medical competence and performance by providing background information and a few guidelines. The most important messages we have tried to convey are that in designing high-quality assessments, foundations are rational decisions based on the best available evidence and careful quality control.

Acknowledgement

With thanks to Professor Chris McManus of University College London for providing a lucid summary of approaches to the assessment of reliability in Box 20.1.

References

1 Van der Vleuten, C.P.M. (1996). The assessment of Professional Competence: developments, research and practical implications. *Advances in Health Science Education* 1 (1): 41–67.
2 Van der Vleuten, C.P.M., Norman, G.R. and De Graaf, E. (1991). Pitfalls in the pursuit of objectivity: issues of reliability. *Medical Education* 25: 110–8.
3 Cronbach, L.J. (1983). What price simplicity? *Educational Measurement: Issues and Practice* 2 (2): 11–12.
4 Ebel, R.L. (1983). The practical validation of tests of ability. *Educational Measurement: Issues and Practice* 2 (2): 7–10.
5 Downing, S.M. and Haladyna, T.M. (2004). Validity threats: overcoming interference with proposed interpretations of assessment data. *Medical Education* 38 (3): 327–333.
6 Evaluation Checklists. https://www.wmich.edu/evaluation/checklists (accessed 28 June 2018).
7 Kane, M.T. (2006). Validation. In: *Educational Measurement* (ed. R.L. Brennan), 17–64. Westport: ACE/Praeger.
8 Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher* 23 (2): 13–23.
9 Cilliers, F.J., Schuwirth, L.W.T., Herman, N. et al. (2012). A model of the pre-assessment learning effects of summative assessment in medical education. *Advances in Health Sciences Education* 17 (1): 39–53.
10 Cilliers, F.J., Schuwirth, L.W.T. and Van der Vleuten, C.P.M. (2012). A model of the pre-assessment learning effects of assessment is operational in an undergraduate clinical context. *BMC Medical Education* 12 (9).
11 Cilliers, F.J., Schuwirth, L.W.T., Adendorff, H.J. et al. (2010). The mechanisms of impact of summative assessment on medical students' learning. *Advances in Health Sciences Education* 15: 695–715.
12 Frederiksen, N. (1984). The real test bias: influences of testing on teaching and learning. *American Psychologist* 39 (3): 193–202.
13 Newble, D.I. and Jaeger, K. (1983). The effect of assessments and examinations on the learning of medical students. *Medical Education* 17: 165–171.
14 Stalenhoef-Halling, B.F., Van der Vleuten, C.P.M., Jaspers, T.A.M. et al. (1990). A new approach to assessing clinical problem-solving skills by written examination: conceptual basis and initial pilot test results. In: *Teaching and Assessing Clinical Competence, Proceedings of the fourth Ottawa Conference* (ed. W. Bender, R.J. Hiemstra, A. Scherpbier et al.), 552–7. Groningen: Boekwerk Publications.
15 Hakstian, R.A. (1971). The effects of type of examination anticipated on test preparation and performance. *Journal of Educational Research* 64 (7): 319–324.
16 Eva, K., Rosenfeld, J., Reiter, H. et al. (2004). An admissions OSCE: the multiple mini-interview. *Medical Education* 38 (3): 314–326.
17 McManus, I., Iqbal, S., Chandrarajan, A. et al. (2005). Unhappiness and dissatisfaction in doctors cannot be predicted by selectors for medical school application forms: a prospective longitudinal study. *BMC Medical Education* 5 (38).
18 Salvatori, P. (2001). Reliability and validity of admissions tools used to select students for the health professions. *Advances in Health Science Education* 6 (2): 159–175.
19 Eva, K., Rosenfeld, J., Reiter, H. et al. (2004). The ability of the multiple mini-interview to predict preclerkship performance in Medical School. *Academic Medicine* 79 (10): S40–S42.
20 Swanson, D.B. (1987). A measurement framework for performance-based tests. In: *Further Developments in Assessing Clinical Competence* (ed. I. Hart and R. Harden), 13–45. Montreal: Can-Heal publications.
21 Wass, V., Wakeford, R., Neighbour, R. et al. (2003). Achieving acceptable reliability in oral examinations: an analysis of the Royal College of General Practitioners membership examination's oral component. *Medical Education* 37: 126–131.
22 Maatsch, J. and Huang, R. (1986). An evaluation of the construct validity of four alternative theories of clinical competence. *Proceedings of the 25th Annual RIME Conference*, Chicago, pp. 69–74.
23 Norman, G.R., Smith, E.K.M., Powles, A.C. et al. (1987). Factors underlying performance on written tests of knowledge. *Medical Education* 21: 297–304.
24 Schuwirth, L.W.T., Van der Vleuten, C.P.M. and Donkers, H.H.L.M. (1996). A closer look at cueing effects in multiple-choice questions. *Medical Education* 30: 44–9.
25 Van der Vleuten C, Schuwirth L, Muijtjens A, et al.(2004). Cross institutional collaboration in assessment: a case on progress testing. *Medical Teacher* 26 (8): 719–25.
26 Schuwirth, L.W.T., Bosman, G., Henning, R.H. et al. (2010). Collaboration on progress testing in medical schools in the Netherlands. *Medical Teacher* 32 (6): 476–479.
27 Van der Vleuten, C.P.M. and Schuwirth, L.W.T. (2005). Assessing professional competence: from methods to programmes. *Medical Education* 39 (3): 309–17.
28 Schuwirth, L.W.T. and Van der Vleuten, C.P.M. (2011). Programmatic assessment: from assessment of learning to assessment for learning. *Medical Teacher* 33 (6): 478–85.
29 Schuwirth, L.W.T., Van der Vleuten, C.P.M., Driessen, E. et al. (2012). A model for programmatic assessment fit for purpose. *Medical Teacher* 34: 205–14.

30 Dijkstra, J., Galbraith, R., Hodges, B. et al. (2012). Expert validation of fit-for-purpose guidelines for designing programmes of assessment. *BMC Medical Education* 12 (20): 12–20.
31 Dijkstra, J., Van der Vleuten, C. and Schuwirth, L. (2010). A new framework for designing programmes of assessment. *Advances in Health Sciences Education* 15: 379–93.
32 Schuwirth, L.W.T., Southgate, L., P.G.G. et al. (2002). When enough is enough: a conceptual basis for fair and defensible practice performance assessment. *Medical Education* 36: 925–930.
33 Bordage, G. (1987). An alternative approach to PMP's: the "key-features" concept. In: *Further Developments in Assessing Clinical Competence, Proceedings of the Second Ottawa Conference* (ed. I.R. Hart and R. Harden), 59–75. Montreal: Can-Heal Publications Inc.
34 Case, S.M. and Swanson, D.B. (1993). Extended-matching items: a practical alternative to free response questions. *Teaching and Learning in Medicine* 5 (2): 107–115.
35 Page, G. and Bordage, G. (1995). The Medical Council of Canada's key features project: a more valid written examination of clinical decision-making skills. *Academic Medicine* 70 (2): 104–110.
36 Schuwirth, L.W.T., Verheggen, M.M., Van der Vleuten, C.P.M. et al. (2001). Do short cases elicit different thinking processes than factual knowledge questions do? *Medical Education* 35 (4): 348–56.
37 Vosniado, S. (2012). Reframing the classical approach to conceptual change: preconceptions, misconceptions and synthetic models. In: *Second International Handbook of Science Education* (ed. B.J. Fraser, K. Tobin and C.J. McRobbie), 119–130. Dordrecht: Springer.
38 Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction* 4: 45–69.
39 Jacobs, J.C., Muijtjens, A.M., Van Luijk, S.J. et al. (2015). Impact of institute and person variables on teachers' conceptions of learning and teaching. *Medical Teacher* 37 (8): 738–46.
40 Harden, R.M. and Gleeson, F.A. (1979). Assessment of clinical competence using an objective structured clinical examination (OSCE). *Medical Education* 13 (1): 41–54.
41 Stillman, P., Ruggill, J., Rutala, P. et al. (1980). Patient Instructors as teachers and evaluators. *Journal of Medical Education* 55: 186–193.
42 Swanson, D.B. and Norcini, J.J. (1989). Factors influencing reproducibility of tests using standardized patients. *Teaching and Learning in Medicine* 1 (3): 158–166.
43 Petrusa, E. (2002). Clinical Performance Assessments. In: *International Handbook of Research in Medical Education* (ed. G.R. Norman, C.P.M. Van der Vleuten, D.I. Newble, et al.), 673–709. Dordrecht: Springer.
44 Regehr, G., MacRae, H., Reznick, R. et al. (1998). Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Academic Medicine* 73 (9): 993–997.
45 Ilgen, J.S., Ma, I.W., Hatala, R. et al. (2015). A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment. *Medical Education* 49 (2): 161–173.
46 Wass, V., Jones, J. and Van der Vleuten, C. (2001). Standardized or real patients to test clinical competence? The long case revisited *Medical Education* 35 (4): 321–5.
47 Norcini, J., Blank, L.L., Arnold, G.K. et al. (1995). The Mini-CEX (Clinical Evaluation Exercise); a preliminary investigation. *Annals of Internal Medicine* 123 (10): 795–799.
48 Bodle, J.F., Kaufmann, S.J., Bisson, D. et al. (2008). Value and face validity of objective structured assessment of technical skills (OSATS) for work based assessment of surgical skills in obstetrics and gynaecology. *Medical Teacher* 30 (2): 212–216.
49. Garnan, A.N., Tyler, J.L., and Darnall, J.S. (2004). Development and validation of a 360-degree-feedback instrument for healthcare administrators. *Journal of Healthcare Management* 49 (5): 307.
50 Williams, M., Klamen, D., and McGaghie, W. (2003). Cognitive, social and environmental sources of bias in clinical performance ratings. *Teaching and Learning in Medicine* 15 (4): 270–292.
51 Norcini, J.J., Blank, L.L., Duffy, F.D. et al. (2003). The mini-CEX: a method for assessing clinical skills. *Annals of Internal Medicine* 138 (6): 476–481.
52 Ten Cate, O. (2005). Entrustability of professional activities and competency-based training. *Medical Education* 39: 1176–7.
53 Weller, J.M., Misur, M., Nicolson, S. et al. (2014). Can I leave the theatre? A key to more reliable workplace-based assessment. *British Journal of Anaesthesia* 112 (6): 1083–1091.
54 Popham, W.J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory Into Practice* 48: 4–11.
55 Ginsburg, S., Vleuten, C.P., Eva, K.W. et al. (2017). Cracking the code: residents' interpretations of written assessment comment. *Medical Education* 51: 401–410.
56 Roberts, C. (2002). Portfolio-based assessments in medical education; are they valid and reliable for summative purposes? *Medical Education* 36: 899–900.
57 Driessen, E., Van der Vleuten, C.P.M., Schuwirth, L.W.T. et al. (2005). The use of qualitative research criteria for portfolio assessment as an alternative to reliability evaluation: a case study. *Medical Education* 39 (2): 214–20.
58 Melnick, D. (ed.) (1998). Pioneering a sequenced OSCE for the Medical Council of Canada: an administrative overview. *The 8th International Ottawa Conference of Medical Education and Assessment*.
59 Clauser, B.E. and Schuwirth, L.W.T. (2002). *The use of computers in assessment. International handbook of research in medical education*. Dordrecht: Kluwer Academic Publishers.
60 Delandshere, G. and Petrosky, A.R. (1998). Assessment of complex performances: limitations of key measurement assumptions. *Educational Researcher* 27 (2): 14–24.
61 Southgate, L., Campbell, M., Cox, H. et al. (2001). The General Medical Council's Performance Procedures; the development and implementation of tests of competence with examples from general practice. *Medical Education* 35 (1): 20–28.
62 Southgate, L., Cox, J., David, T. et al. (2001). The General Medical Council's Performance Procedures; Peer review of performance in the workplace. *Medical Education* 35 (1): 9–19.
63 Schuwirth, L., Van der Vleuten, C., Durning, S.J. (2017). What programmatic assessment for learning in medical education can learn from healthcare. *Perspectives on Medical Education* 6 (4), 211–215 doi: 10.1007/s40037-017-0345-1
64 Cusimano, M.D. (1996). Standard setting in medical education. *Academic Medicine* 71 (10 Suppl): S112–S120.
65. Norcini, J. (2003). Setting standards on educational tests. *Medical Education* 37: 464–469.
66 Schuwirth, L.W.T. and Van der Vleuten, C.P.M. (2006). A plea for new psychometrical models in educational assessment. *Medical Education* 40 (4): 296–300.
67 Cook, D.A., Kuper, A., Hatala, R. et al. (2016). When assessment data are words: validity evidence for qualitative educational assessments. *Academic Medicine* 91 (10): 1359–1369.

21 Written Assessment

Brian Jolly[1,2] and Mary Jane Dalton[1]
[1]School of Medicine and Public Health, University of Newcastle, Callaghan, New South Wales, Australia
[2]School of Rural Medicine, University of New England, Armidale, Australia

KEY MESSAGES

- Start early; at least six months before the examination for intra-school assessments.
- Review or construct an assessment blueprint for the training programme, course, or unit that is being assessed.
- Decide on the objectives or domains in the blueprint that need to be assessed with written formats. Such objectives would typically require recall of information, recognition of patterns of presenting symptoms, clinical decision-making, choice of investigations, analysis of data and synthesis of ideas, and choice of strategies for management.
- Decide whether a constructed response is necessary to assess the objective or whether a selected response would do.
- Choose the appropriate format for each item of the blueprint.
- Collect all the items into their format groupings.
- Arrange a workshop to develop any new required items, in the appropriate format, along the lines identified above.
- Once the test is proofed and prepared, arrange a further workshop to set standards, methods for which may vary depending upon item type and the setting.
- Deliver the test.
- Prior to result determination, thoroughly review the items using item-analysis software. Make sure that key personnel know how to interpret the output.
- Determine the final scores of candidates and cut scores for the test with all the poorly performing items eliminated.

Introduction

Assessment is a critical component of educational and accreditation strategies used in health professions education. The long standing senior guardians of those professions have traditionally had high standards and developed refined means of ensuring them [1]. With increased international concern for patient safety [2, 3], the need for efficient, reliable, and valid assessment has become vital. In turn, this has led to a search for increasing levels of sophistication in assessment, mostly targeted at the development of high-fidelity simulations and work-based assessments [4, 5]. However, health professionals operate across a very broad spectrum of human activity, from listening and talking to patients, to intricate and physically demanding interventions. Knowledge and thinking play a crucial role in these undertakings, and there is considerable evidence to suggest that knowledge, and its storage in clinically useful frameworks, are the most important attributes that divide the novice from the expert [6, 7].

One of the most useful ways of finding out what people know is to ask them a question. When such enquiries are written down or the person is required to give a written response, we are in the territory of written assessment. This chapter introduces the reader to the use of written assessment in medical and health professions education and covers the following four key areas:

- the placement of assessment within the curriculum or training programme
- different formats of written assessments
  - constructed response
  - selected response
  - combined constructed and selected responses
- how well-written test items do their job
- how to set appropriate standards for these assessments.

Historical Background

Assessment in medicine historically involved oral activities: general discussion, case discussion, demonstration, answers at the bedside, etc. In this tradition the long case was adopted, in 1858, by the newly established General Medical Council (UK), as a means of assuring the competence of physician apprentices [8]. However, with the post-Flexnerian (c. 1910) emphasis on scientific endeavour in medicine, the measurement of knowledge through reliable and explicit means became increasingly emphasised [9]. In general, psychometric perspectives on assessment have reflected the

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

need for efficiency, reliability, and validity, whilst the constructivist perspective has stressed the need for the assessment processes to have appropriate influences on learning [10, 11]. In recent years these two heritages have converged considerably.

Research on assessment in medicine has been most productive in the last 40 years [12], and we do not intend to present all of this research here; there are books available that summarise the area and will help the interested reader [13–16]. We will, though, try to provide a balanced overview of the field and consider the theoretical determinants of assessment in as much detail as is necessary for a newcomer to the field.

Assessment in the Programme

Before designing any assessment event, it is useful to think through the purpose of that assessment. Although we use the term 'assessment' quite loosely, each assessment that a learner undertakes usually has a particular function, which reflects the complexity of professional training. These functions include the following:

- measuring competence
- measuring performance
- diagnosing trainees' problems
- measuring improvement
- self-evaluation
- selecting individuals
- identifying effective learning/teaching
- showing effectiveness of the programme
- measuring curriculum change.

Each of these purposes will have constraining influences on the content of the assessment, the strategy used, and the techniques employed. For example, assessments used to certify competence need to be closely aligned to the core programme objectives, that is, demonstrate content validity. They also need to be reliable and focused at pre-specified levels of competence. However, an assessment used to select trainees into a restricted-entry postgraduate training programme may be targeted at a level of excellence, and should ideally contain specific elements that predict success in that programme [17].

One process for assuring the content validity of assessment is called 'blueprinting' [17–19]. Essentially, this needs to be done for the assessment regime or strategy of the whole course, *before* individual components are designed, because it is one way of ensuring that the content is appropriately assigned to those components with the best fit to the mental processes being assessed. That is, there will be specific domains of activity that are better suited to some types of assessment, and this will become apparent when a thorough blueprinting exercise, that reflects the complete assessment strategy, is conducted for the whole programme. At undergraduate level this blueprint can be designed by mapping the objectives of the curriculum, and might reflect mostly knowledge and basic skills. At postgraduate level, where the outcomes might be much broader, following a framework such as CanMEDS for example [20], the blueprint might be complex and involve many more modalities. Trying to design the written segments of an assessment process without adequate blueprinting of the whole assessment of the programme is likely to result in an imbalanced assessment.

Knowledge, Reasoning, and Written Assessment

When the blueprinting is complete, it will identify a subset of attributes of a qualified health professional that reflect the cognitive domain of human endeavour – understanding, recalling, recognising, reasoning, inferring, deducing, and deciding. Such attributes can be assessed in a number of ways, and there is little doubt that an effective clinician needs to be able to do all these things with patients and colleagues in clinical situations. However, not only is it impracticable to assess these attributes comprehensively or effectively in clinical situations, there is some evidence that it leads to contamination of the measurement process by other factors – these factors are generally referred to as 'construct-irrelevant variance' [21]. For example, oral examinations purportedly aimed at examining clinical reasoning can be reduced to the investigation of factual recall, frequently focus on minutiae, and can reveal examiners' 'cultural incompetence' [22].

One way of avoiding these influences is to decontextualise the assessment. In most of the twentieth century this trend, and the recognition that, psychometrically, a minimum number of questions on a topic was needed to give a reliable estimate of a person's knowledge, gave rise to a number of developments in assessment. These included the multiple-choice question (MCQ), in its various guises, and the short-answer question (SAQ). However, over the last 20 years, researchers have rediscovered that context is a crucial element of memory and thinking. Generally, the milieu in which something is learned has an important function in the formation of memory [23] and in clinical reasoning in particular [24]. The way that this knowledge is obtained and organised is as important as its sheer volume; expertise is significantly dependent on well-structured knowledge [25]. For practising clinicians, the retrievability and utility of knowledge are critical, and depend both on the efficiency of learning and where and how that learning has taken place. In brief, clinical reasoning depends on integrated knowledge, preferably learnt or repeatedly accessed by the learner in complex clinical contexts, where appropriate scientific principles are articulated to address patient problems [23–26]. These findings mean that we now have new varieties of written assessments. In MCQ format, the extended matching item and the script concordance test both typically establish a clinical context for the questions posed to assessees. And new approaches to short-answer items include the key features item. At the same time, previously endemic item types such as the multiple true-false MCQ (X type MCQ) have come under cogent criticism [27].

There is another dimension to written assessment that needs to be considered: the nature of the response. This can either be a selected response or one constructed by the test taker. In general, although there has been some dispute over this, it is thought that constructed responses require candidates to operate at a higher level than selected responses; for example, recall and synthesis versus recognition and choice.

In the next section we look at each of the format types. There are exhaustive treatments of many of these available in the literature [15, 27]. Here, we will abbreviate much of this extensive debate in order to help the newcomer to make some practical decisions about assessment.

Formats of Written Assessment

Usefully, Epstein [28] has summarised assessment techniques and their general potential usage. We have expanded his summary table slightly for the written assessment components in Box 21.1. We have also removed Epstein's column referring to where each method might be best used to allow more flexibility. The assessment designer needs to think carefully about all of the following factors to make appropriate choices:

- domains of activity or objectives that need to be assessed
- the need to reflect the blueprint outcomes
- the need to encompass local requirements for assessment of certain attributes.

Constructed-response Formats

The Essay

An essay is a form of assessment 'which requires a response composed by the examinee, usually in the form of one or more sentences, of a nature that no single response or pattern of responses can be listed as correct, and the accuracy and quality of which can be judged subjectively only by one skilled or informed in the subject' [29, p. 495].

On the face of it, essays are one of the most effective ways of ascertaining how good a student is at constructing a complex response to a challenging question. The other approaches that might be used would be oral examinations, projects, observation of discussion, and many more. However, many of these are subject to variability due to extraneous serendipitous factors. For example, in orals there is no 'product', so judgements about performance are made 'on the run', unless the examination is recorded and analysed later, and this would significantly increase the burden of administration.

Essays can be delivered in two contexts 'unseen' or 'seen'. In the former a question, or usually a suite of questions, is prepared and delivered 'de novo' to students under examination conditions. In the latter a topic is provided to students, and they are given a time limit in which to address it. They may use any resources they can find. Occasionally, questions may be given to students in advance, but the essay is written under examination conditions.

The attributes of the essay and issues in its construction, delivery, and marking are far from straightforward. The key questions to consider when choosing any assessment are as follows:

- What type of response is required? (a content-validity issue)
- What cognitive processes are involved? (a content- and construct-validity issue)
- How well do the response and the cognitive processes invoked map to the expectations of the assessor about student performance, and to the objectives of the curriculum? (a content-validity issue)

Clearly, in the unseen examination, critical components of a successful essay will depend on memory (both short- and long-term). The essay's quality will also depend on the ability of the student to construct sentences of the appropriate length that are unambiguous and grammatically correct, and to organise his or her knowledge in a way that addresses the question. If these abilities are all critical elements of the programme or of the environment into which this examination might be the entry point (e.g. internship), then the essay may be a rational assessment. If they are only prerequisites and were assessed earlier, or if only knowledge and memory are important, would another assessment method be superior?

Constructing an Essay Question

Decide on the constellation of attributes that you need to assess and decide whether an unseen or a seen essay would be the most appropriate. For example, if most of the knowledge you require the student or trainee to have is basic, core, and extensive, and must enable them to solve, manage, and monitor a real clinical problem at some time in the future, and possess an in-depth knowledge of related or differential conditions, then probably the unseen essay would be appropriate. Then:

- Choose a problem or issue that can be addressed satisfactorily in the time allocated, or limit it in such a way as to make it answerable.
- Define the problem/issue and describe the task clearly.
- Describe the structure that the answer should take and its scope.
- Do not use complex language in the question such as double negatives, ambiguities, and abbreviations.
- Use terms that cover the cognitive processes that you are expecting to be used in marshalling the answer: for example, compare, predict, prioritise, rather than discuss, outline, examine, elucidate. (See Box 21.2.)
- Avoid questions where radically different answers will be acceptable for a given question. This is fine for a group discussion or debate, but it makes marking an essay difficult, especially where assessors might favour one or other of the answers.
- For every question, preferably construct a model answer, or list the essential features that should be contained in the answer.
- Trial-run the questions on a group who should know the answers.

BOX 21.1 Types of written assessments and their primary usages. Adapted from Epstein [27].

Method	Domain usage/response mode	Design Factors	Limitations	Strengths
Constructed-response formats				
1a Essay – Traditional. The typical university essay, either seen or unseen, where the writer is required to describe, discuss, and propose new perspectives on one or more issues. The answer may or may not be predetermined.	Any situation where lengthy explanation or detail is required. Detailed synthesis of information; interpretation of literature, evaluation of management options. Context frequently provided by the candidate.	Traditionally questions can vary from the blindingly obvious to the very obscure. Large number of dimensions to the constructed response. Getting questions right takes time. Model answers or protocols help marking. High marking workload.	Can be usurped into provision of lists, e.g. for treatments. Can become memory dumps. Can be misinterpreted. Long testing time per topic, so limited coverage possible. Reliability variable and susceptible to rater and candidate bias	Total flexibility in question setting. Can avoid cueing. Regarded as using higher-order cognitive processes.
1b Modified essay Specifically developed for medicine – mostly used in general practice. Highly structured case vignette followed by questions on any aspect. Focused on candidates' management of a case or cases. Answer(s) usually predetermined.	Clinical management issues. Some cue identification and reasoning required to link, e.g. signs and symptoms to investigations and management. Context provided by the question.	Can move from one stage of clinical management to another easily, by using slightly different cases to address issues, e.g. patient management in one case, and ethics in a similar one. More efficient sampling of a wide area of knowledge possible	Needs careful design to avoid cueing. As a result can be patchy in sampling knowledge across cases.	Can avoid cueing. Context is controllable by question setter. Can demand wide range of cognitive processes. May be machine scoreable in next 5–10 years.
2a Short answer – Traditional. A short question that asks for a constructed specific answer, usually requiring one word, a short phrase, or a line or two of text. Answers mostly predetermined.	Recall of specific facts or statements about biomedical or clinical processes. Context provided by the question.	Deceptively simple to construct. Can sample widely different domains of knowledge easily.	Very wide variety of formats and little research on their use and psychometric properties. Can lead to cueing across items. Context provided by question.	Scoring by machine becoming a reality. Can replace MCQs where recall is thought to be vital (e.g. decisions based on core knowledge and experience)
2b Short answer – Extended. A question that asks for an extended answer, usually requiring a paragraph or two, that may address different aspects, or an extension, of the issue. Answers may be predetermined.	Recall of related groups of concepts or relatively short explanations. Context provided by the question.	Deceptively simple to construct. Can sample widely different domains of knowledge but in more depth than short answer.	As above. Scoring more difficult as depends on multiple attributes of answers involved in essay construction. Machine scoring not possible. Context provided by question. Recent research on analysis of answers can give more insight into level of functioning of candidate.	Total flexibility in question setting.
3. The portfolio A means of collecting evidence of learning, sometimes in hardcopy format, but increasingly as an electronic repository.	Can include any of the formats discussed here, but also can retain photographic, audio, and video evidence of activities and/or scholarly outputs.	Has the potential to be a powerful mediator or vehicle for programmatic assessment.	Difficult to assess using psychometric approaches. Debate about whether psychometric approaches should or should not be used.	Total almost limitless flexibility in what the portfolio can contain.

Selected-response modes				
1a MCQ – True/False. Typically a short statement or brief paragraph followed by several (3–6) options. Candidates are asked to identify which options are true and which ones false in relation to the initial statement.	Recognition of consonance between two facts, attributes, or concepts. Can test recognition of clear-cut knowledge in many domains. Complex items requiring calculations or problem solving have been used.	Requires all options to be absolutely true or false. Can test knowledge of contraindications through the 'false' option.	Difficult to write. The number used in most assessments can lead to cross-cueing. Can involve silly or irrelevant options due to lack of absolute falsehoods. Getting statements into an absolute true/false mode sometimes requires convolutions such as double-negatives. Extreme controversy over 'correction for guessing' as random choice of options results in 50% score. Rapidly waning in popularity.	Can test range of knowledge in limited assessment time. Machine scoreable. True/False requirement restricts applicability and engenders artificiality.
1b MCQ – 1 from N. Typically a short statement or brief paragraph followed by several (3–6) options. Candidates are asked to identify the option that best fits with or is the best outcome for the initial statement.	Recognition of consonance between two facts, attributes, or concepts. Can reflect basic clinical decisions, basic science, or hypothesis generation.	Easier to write than MCQ T/F. Choosing one best answer is more salient to most areas of medicine.	The number used in most assessments can lead to cross cueing. Need not involve a correction for guessing.	Efficient sampling of knowledge. Allow more subtle distinctions than T/F types. Machine scoreable.
1c MCQ – Extended matching. Typically a topic area (e.g. headache), followed by many (6–26) options homogenous to a clinical grouping (e.g. diagnosis). There is a linked question asking candidates to choose the most likely diagnosis. Then one or more paragraphs each comprising a clinical case vignette, including, e.g. headache presentation at various stages of progression, each of which may indicate different 'best' diagnoses.	Recognition of consonance between (typically) clinical presentations and their underlying pathology, investigation, and outcome; diagnoses, prognoses, tests, pharmacology, etc. Items appear to involve basic clinical reasoning. Students report fidelity to 'real' medicine.	Relatively easy to generate first drafts. Salient to most areas of medicine that depend on a clinical context.	Not easy to write in some areas of medicine, especially non-clinical ones, e.g. epidemiology. Some argue that the 'extended' list of options is not as useful as first thought – many options are redundant.	Seem to be more reliable than one-best-answer MCQs and T/F MCQs, due to increased difficulty. No corrections for guessing needed. Good discriminators at higher levels of ability.
1d MCQ – Integrated item Typically a topic area (e.g. headache), and a brief patient-based vignette, followed by a series of one best answer MCQs that each approach the topic from a different perspective, anatomy, physiology epidemiology, stage of progression, etc.	Assessment of breadth of knowledge and understanding of the basic processes involved. Best format still under investigation.	Relatively easy to generate first drafts. Can cover a wider framework than a single item on the topic.	May need review by different specialty groups to authenticate the linkages between the topic and the disciplinary perspective. Using multiple questions on the same case may restrict the sampling across the programme blueprint.	Work in progress …

(*Continued*)

BOX 21.1 (*Continued*)

Method	Domain usage/response mode	Design Factors	Limitations	Strengths
2d MCQ – Script concordance Typically a case vignette followed, for example in items on diagnosis, by statements that give an additional sign or symptom and a question that asks whether a specific diagnosis would be more or less likely if such an attribute were present in the case (see example in text). For example, given a description of a 67-year-old man with chest pain, if pain radiating down the left arm were present, would the likelihood of myocardial infarction be 'strongly increased, …, strongly decreased'.	Recognition of relationship between, and agreement with an expert group on, attributes of case presentations that are predictive of diagnoses, prognoses, findings on investigation, etc. Appears to involve basic clinical reasoning and personal probabilities.	New type of item, limited experience available of construction. Scoring generated by expert group. May have more than one answer that scores marks. Appears to discriminate effectively between experts and non-experts in some specialties.	Probably limited to diagnostic and prognostic decisions.	More research needed, but does show high construct validity for clinical experience. Writing protocols and rules still in development.
Constructed and selected response				
1. Short answer – Key features. Usually a short case vignette followed by between one and three questions that investigate the taker's knowledge of the main aspects of the case. Answers may be constructed or selected, usually requiring words or short phrases.	Answers that attempt to focus *only* on the critical aspects of clinical cases, e.g. key decisions and the factors underpinning those. Developed (1990s) to counter arguments that short answers led to isolated recall of facts and trivialisation. Context provided by the question.	Strict rules for design, done usually by a small team. Items may involve some *selected* responses as well as *constructed* ones. Can explore wide variety of cases. Can match response mode to attributes of the context – e.g. selecting the most important features in clinical investigation results. Shares some properties of modified essay questions.	Scoring and standard setting can be complex. For example, although single word answers are common, there may be several answers to one question, each differentially weighted. There may be totally inappropriate or dangerous answers given by test takers. Can be challenging to avoid cueing between different parts of the item.	Partial scoring by machine is now becoming a reality. Can replace MCQ-style questions where recall is thought to be important (e.g. decisions based on core knowledge and experience). Has embedded quickly into assessment technology in medicine.

BOX 21.2 Words that can be used to drive learners towards certain cognitive processes when answering an essay question

analyse	diagnose	justify
apply	explain	match
classify	evaluate	plan
compare	generate	predict
compose	identify	propose
defend	infer	summarise
develop	interpret	synthesise

- Make sure that *all* the defined essential features appear in the trial group's answers as a whole, not necessarily in every member's answers.
- Proofread the paper three times using a different reader each time.

When marking, ideally a pair of assessors working independently should score each answer. If that is not possible, the same marker should score the same question for all candidates. This minimises extraneous variance stemming from the different ways each examiner marks different questions. To help with quality control of marking, all examiners of a question should preferably see all, or at least the complete range, of answers to that question. Scores for a question should be the mean of all the examiners marking that question.

Although assessor training is desirable for most written assessments, a model answer is often better than attempts at calibration. In fact a series of model answers specifically written to be at the boundaries between two grades is a very useful way of enabling accurate classification. For example, in an A–E grade system, model answers should be at A/B and B/C boundaries. This enables most essays to be rated using two anchor points, since most will fall between the boundaries. A model answer is also useful when there is a common core of content that needs to be covered in the essay. And a model answer can help minimise extraneous variance from different examiners' perceptions or biases. For example, marking can be influenced by better or worse answers coming adjacent to the answer you have scored. This is minimised by grade boundary model answers; you can check where the current essay stands in relation to them. Another trick, if there are no model answers available, is to scan briefly all the answers you have before you start and pick out what appear to be a good, a middle, and a poor answer as the first three you mark. If marking in a pair, sharing these same three essays can be useful as calibration.

Where students each write an individual assignment on a pre-assigned or selected topic, it may not be possible to have a model answer (there would have to be as many models as there are questions/students). In this case a process-based framework can be used (see Box 21.3). Papers should be marked anonymously. A procedure needs to be in place to address wide marking variations within one question.

BOX 21.3 Example: Guidelines for scoring of resident/medical student essays

A. Content (25%)

Reviews major and relevant articles for topic
Content is current; content is accurate
Thorough; sufficient detail to understand issues being discussed
Articles cited are salient to the discussion

B. Critical review (15%)

Critique of methodology used in studies cited
Assesses quality of studies cited and compares to population of interest
Presents differing views; compares and contrasts

C. Conclusion/Synthesis (25%)

Synthesises data presented
Clear recommendations with clinical and/or research
Implications; includes implications for clinical practice
Conclusions are based on critique

D. Organisation (20%)

Title reflects content of paper
Presence of an abstract and summary
Abstract is reflective of the paper; not simply a repeat of the introduction
Introduction includes statement of what will be covered in the essay
Sections follow each other in a logical order with use of headings and sub-headings

E. Style (15%)

Formal scientific writing style
Easy to read and follow line of thought
Uses plain English, good sentence length, and good use of paragraphs
Avoids unnecessary jargon

Adapted from Canadian Association for Physical Medicine, with permission

Assessors are frequently urged to use the complete scoring range and avoid centralising tendencies. However, health profession students tend to be very high achievers, so it is not uncommon to have skewed distributions on tests. If high scores seem warranted, they should be given.

There are constraints on the essay. Many clinical teachers do not regard them as relevant, but others suggest that they give training in marshalling arguments and practice in writing. However, the practice that essays give may be that of 'bad' writing – rushed, unedited, poorly planned (because of time constraints), and incompletely organised [30]. Many authors like to let the issues sit in their minds for some time before launching into print (like the authors of this chapter), but that is not possible in an essay delivered under examination conditions and may be difficult even in a seen paper.

An essay question does not necessarily assess higher-order cognitive skills. It often merely assesses recall, dressed up as something more profound. This happens in two ways, caused either by the students or the teacher. Students can memorise vast tracts. Asking students how they went about answering the questions in a debriefing session will assist you in deciding how your essays performed. However, beware the bright student who unwittingly uses memory and reports he/she is 'thinking'. If you design an essay question meant to assess higher-order processing but then arrange the scoring framework in a way that allots marks to recall, or biases towards recall processes, you are not assessing higher-order cognition.

Where's the Evidence for Essays?

There is a great deal of evidence about the impact of construct-irrelevant variance on essay and other qualitative assessments; however, little of this comes from tertiary sector medical and health sciences.

First, there are increasing problems in higher education generally with cheating on some types of assessments. Essays are probably prone to this more than other types. All 'seen' essays should be checked for multiply used content through a system such as TurnitIn [31]. It also helps, in courses taught regularly with similar assignments year on year, to check content against previous students' work, as 'borrowing' a previous student's essay is not uncommon.

Additionally, the impact of assessors' perspectives on the outcomes of the marking process cannot be underestimated. A recent study of marking across four disciplines (psychology, nursing, chemistry, and history) [32] attempted to examine and evaluate the constructs that the sample of assessors used to make judgements about the pieces submitted for assessment. It confirmed previous findings of other studies showing considerable variation in grading. However, this study isolated five potential reasons for this variance: assessors were actually using different criteria to those published as the marking criteria, assessors had different understandings of shared criteria, they had different perceptions of appropriate standards, the criteria they used contained various and heterogeneous sub criteria, and assessors valued and weighed criteria idiosyncratically [32, pp. 473–7]. So, in order to produce consistent ratings all five impediments would need to be addressed. These researchers conclude, at least as far as complex assessments are concerned:

> the real issue is not about artificial manipulation of marks without reference to evidence. It is more that we should recognise the impossibility of a 'right' mark in the case of complex assignments, and avoid over extensive, detailed, internal or external moderation. Perhaps, a better approach is to recognise that a profile made up of multiple assessors' judgements is a more accurate, and therefore fairer, way to determine the final degree outcome for an individual [p. 479].

In terms of the quality of expression, experiments that change the quality of writing but not the content show moderate influences of style or construction factors on assessments of language and writing, but not for other content areas such as science and mathematics [33]. In addition:

'Content scores do not seem to be appreciably affected by writing style when the scoring is done by teachers who have been trained by scoring professionals. Ratings in reading, social studies, science, and mathematics should be unaffected by writing style, and the results indicate that the scorers were reasonably successful at assessing the content of these responses without meaningful confounding with writing style' [33, p. 26].

In a study on law essays, agreement amongst law professors in how they graded a single typical essay was 0.58 (intra-class correlation) [34]. However, although the basis for this agreement was not investigated, and it was not clear in the paper whether these essays were marked anonymously, assessors' marks were higher for longer answers and for those written by brighter (higher grade point average) students. Notably, the measures of length of essay and intellectual ability were unrelated. However, a combination of these two factors yielded a very high correlation with the assessors' grades. In other words, longer essays were marked higher even if written by the poorer students, and long essays by more able students attracted even higher marks. Markers who had no law training generally assigned the same grades to the papers as did the law professors. It appeared, therefore, that both professors and lay markers could and did identify and reward those papers that presented a persuasive 'common sense' answer to the question.

A more refined study showed that the amount that a student wrote on major issues, the use of jargon, the use of transitional phrases, and quality of handwriting all had significant positive correlations with grade, and grammatical and construction errors both had significant negative ones [35]. Two further predictors of success were 'strength of argument for conclusions reached' and 'tendency to argue both sides of an issue'. In further studies, quality of handwriting has been confirmed to be a factor [36]. However, on the whole handwritten essays tend to score higher than those 'marked' by computer programs [37].

An experimental study of structuring essay questions in medicine showed that the reliability of the structured questions was higher, due to the reduced variance between examiners in this format, and there was better agreement between scores on individual questions in the structured format [38].

Finally, the major disadvantage of the traditional essay is that it samples a small area in depth and this restricts inferences that you can make about a person's competence to just those specific areas.

Modified Essay Questions

A modified essay question (MEQ) consists of a brief scenario or clinical vignette, followed by one or more short but searching questions. MEQ takers are required to construct answers, usually of a paragraph or two. Each question is designed to assess a wide range of issues and the ability to think rationally and laterally. By way of illustration (and with acknowledgement to the UK Royal College of General Practitioners) here are two examples.

- *Daisy Boyd, aged 68 years, arrives late for her routine appointment smelling of urine. How would you manage this situation?*

The question could contain issues as diverse as the management of incontinence through to the management of time and the doctor's own feelings.

- *Mike Hornby, aged 44 years, is in the terminal phase of motor neurone disease. He says, 'Will you help me when the time comes?' What factors influence your response?*

Such a question could raise clinical, ethical, legal, referral, and personal issues. It may require further qualification to limit the range of answers to the areas you want, or the length of response to one that could be reliably scored.

The MEQ was first developed in Australia and the UK to overcome the major restrictions of sampling and scoring pertaining to the traditional essay and to the now largely disused 'patient-management-problem' (PMP). The PMP had some potentially useful attributes; the technique required the candidate to fully explore a case from initial diagnosis to management and follow-up, through a series of largely selected responses. However, it was dogged by psychometric inadequacies [39]. The MEQ allows exploration of different aspects of a case, by using constructed responses. Initially, whole cases were followed through, but the impact of cueing in this format is high and, in general, the principle now is to explore candidates' knowledge of cases and management through a wider sampling of content. There is a broad literature on MEQs [40, 41], and they are ideally suited to computer administration and response collection, and can be marked by computers.

The major difficulty involves the same issues that apply to essay marking – there is variability between markers on most constructed-response types of question [42, 43], and score weighting can be problematic. Also, because cueing is so difficult to eradicate in an MEQ, many assessment developers limit the scope of each clinical problem to just one or two issues, so that dealing with a whole case across a time frame (one of the original arguments for using MEQs) becomes impossible. This can be circumvented by computer presentation of the segments of the MEQ.

Constructing an MEQ

Decide first on the type of objectives that you wish to assess – diagnosis, decision-making, patient management, or self-management. In the example MEQs, the designer was looking for affective and professional components of the encounters as well as cognitive ones. With some examinees this requirement may need signposting more clearly.

Decide how far it is possible to delve down into the specific case without risking cueing and avoid vaguely worded questions. To illustrate these points here is a deliberately flawed example.

- *Mrs. Brown, a 38-year old primary school teacher, complains about fatigue and tachycardia. She has been admitted to the general medical unit on which you work, for further investigation.*

Question 1: What are the three most likely diagnoses?

Question 2: List five specific questions which would help you distinguish between these possibilities.

- *A routine blood test reveals microcytic hypochromic anaemia with a haemoglobin level of 9.8 g dl^{-1}.*

Question 3: List two typical signs you would look for when you examine the patient. Question 4: Did this information affect your first diagnosis? If yes, how (explain briefly)?

In this example, computer delivery or physical removal of the answers, first, to Question 1, and then Question 2, before giving the information about anaemia and asking candidates Questions 3 and 4, would be required to avoid cueing. Questions 1 and 2 assess broad knowledge of such clinical presentations and initial diagnostic strategy. They require understanding of the clinical significance of the scenario. Question 3 tests linkage between data from investigation (that may not have been initially considered by a test taker) and subsequent questioning. Question 4 is vague and open to misinterpretation – for example, Question 1 asks for three likely diagnoses – which one does Question 4 refer to? Is the test taker supposed to assume certain positive or negative outcomes from their examination of signs in Question 3? What does 'information' mean in Question 4? What does 'affect' mean? What is the designer's rationale for asking Question 3 after the delivery of the information about the blood test? Would this information be better provided after Question 3?

Where's the Evidence for MEQs?

Psychometric studies done on the MEQ in the 1980s showed that reliabilities ranged between 0.43 and 0.90 (Cronbach's alpha) for a 60-item test, depending on the content area [41]. However, one study suggested that over 50% of MEQ items in a general assessment for undergraduates in medicine and surgery tested nothing more than factual recall [42]. This contrasts with the rationale for MEQs that emphasises their ability to reflect analysis, interpretation, and clinical decision-making. A more recent study in the same institution has resulted in the removal of the MEQ from the undergraduate assessment programme [43]. Published research on the MEQ has decreased significantly in the last decade. However, they are still used in some specialties [44], and delivered by computer in undergraduate settings [45].

Short-answer Question

Many educators use SAQs in some form. Frequently, in vivo as it were, SAQs are used as means of gauging students' factual knowledge or understanding – for example, during lectures and ward rounds. In this verbal form they tend to be quite short, asking for one word or a few alternative answers, within a specific context, as in the following example:

- What is the most common feature of diabetic retinopathy we are likely to see in this patient?

The other major use of SAQs is in assessments. Various forms exist, requiring the test taker to complete the sentence or supply a missing word (a 'cloze test'), give short descriptive or analytical answers, or annotate diagrams. Such questions can demand a wide range of responses, from one or several words, a paragraph, to more than a page. The different forms of SAQ provide for great versatility in usage, but make classification difficult. An individual question can be used to assess a specific objective and unlike MCQs, SAQs have the advantage of requiring students to construct an answer, rather than choosing (or guessing) from provided options, so avoiding cueing (at least when SAQs are used sparingly).

SAQs are easier to mark than essay questions and usually involve a structured marking sheet that indicates all possible answers, and ones that should or should not get credit. Marking sheets should also indicate whether spelling needs to be perfect or which common misspellings are acceptable. One-word answers are computer scoreable. Currently programs are being developed for scoring that involves longer answers [45].

Items should be marked with assessors blind to the identity of candidates and different markers allocated to different questions or sets of questions. In this way, examiner bias is diluted for each candidate. Some assessors report that having the marking done at one time in a large room with all examiners able to talk to each other as unexpected responses are discovered is beneficial to efficient and equitable scoring. Assessment designers need to be prepared to accept answers not on the score sheet, some of which may or may not have been predicted. There will need to be a system for referring these to the assessment convenor or committee – do not allow discretion at the marker level, some markers may be unable to make this judgement.

Marking poses the major difficulties with this form of assessment, although there is variability between markers on most constructed-response types of question. Increasing the number of markers and number of questions can ameliorate the problem, but is frequently impractical [46, 47]. Many educators allege that SAQs reduce the likelihood that students will look for the relations between objectives or sections of the subject whilst studying, and that complex issues cannot always be satisfactorily addressed in short answers. However, there is little empirical evidence for these assertions.

Constructing an SAQ

- Identify the specific learning objectives the item will cover. These are generally in the area of factual recall, comprehension, application, or analysis. Higher levels such as evaluation or synthesis will probably require a longer test format, such as a modified essay.
- Choose the most appropriate SAQ format for the objective – a cloze or completion item, an open one-word or phrase answer, a series of answers or a question that requires a short paragraph.
- State the item concisely in clear, unambiguous simple language. A good SAQ tests factual knowledge or capacity to analyse and clinically interpret a scenario. Introducing an element of the test taker's ability to make sense of the question introduces construct-irrelevant variance into the assessment.
- Look at the draft item from a number of different perspectives – mentally try out adequate and inadequate responses. Ideally, an item aimed at one fact should have just one answer, and one aimed at alternatives (e.g. differential diagnoses) should have as many as are appropriate. However, what you may think of as a clear, straightforward question may frequently be answered in multiple ways, depending on how the reader reads it.
- It is good practice to give the test taker an indication of the length of answer required and to indicate how many marks are available for the question.
- Some research suggests that items asking for positive perspectives (e.g. knowing the best method, describing good practice, or identifying the most relevant facts) have greater educational significance (e.g. in terms of capacity to measure objectives) than knowing the poorest method, unsatisfactory practice, or the least relevant issues. However, clinical science sometimes depends on the capacity to rule out rare or unlikely occurrences, so research done in general educational settings may not always apply in the health context. If you have to word an item negatively, it is advisable to use some form of emphasis for the negative words (e.g. 'What is *not* an appropriate management option in this situation?'), using italics, bold type, or underlining to stress this for the test taker.
- Try to avoid grammatical cues to the answer or providing answer spaces that are equal or proportional to the lengths of the required responses.
- Where a numerical answer has to be supplied, for example from a calculation based on clinical data, indicate both:
  a) the degree of precision expected (e.g. give your answer to one decimal place and answers within 5% of the correct value will be given credit) and
  b) that the appropriate units must be indicated.

  Not doing this will result in uncertainty for markers about whether the answers supplied are acceptable or not.

Where's the Evidence for SAQs?

There is very little research on SAQs, particularly in medicine. However, there is some evidence from secondary education that constructed-response SAQs measure exactly the same thing as MCQ items, as long as the stems are the same [48]. In other words, the cognitive task set to the test taker is more important than the response format. However, once the task diverges, even in the same content domain, the correlation between the two forms falls off. Also, SAQs are more reliably scored than essays [49, 50], largely because the pitfalls of scoring lengthy answers are avoided, and because SAQs can sample more widely in a given time. In addition, using SAQs may reduce the reported differences between men and women, and black and white racial groups on propensity to omit items in MCQ assessments [51]. In medicine SAQs have successfully been used as a reliable alternative to MCQ items in a progress test in the Netherlands [52]. One study showed that using SAQ assessments could result in better retention of information over time, as long as the delayed test was a short-answer test. There was no difference between groups if the assessment was an MCQ test [53].

Scoring of SAQs

One of the largest impediments to the use of SAQs seems to be the work involved in scoring them. They need to be marked by experts in a reliable fashion and without inducing examiner effects (stringency and theatricality) that obscure construct variance. However, there is much growth in intelligent computing such that within a few years it is possible that such items will be scoreable by computers, an idea first mooted in 1966. Perera and colleagues [54] have developed programs that use Vector Space Models and Natural Language Processing (NLP) techniques to match student responses with model answers. In order to handle variations in the students' answers, NLP techniques (lemmatisation – grouping together inflections of the same word, tokenisation – a means of encoding sensitive data, handling of spelling mistakes, identifying relation of objects, and identification of upper and lower case of words) were successfully used to accurately reflect human marking. In some studies the correspondence between machine and human scoring for SAQs is as high as 95–98% and the same techniques can also be used for essays, as long as a model answer exists [40, 55].

Using Multiple Constructed Responses in One Tool

There have been examples developed of computerised approaches to written assessment that take advantage of the manipulations that can occur in case presentations to challenge students in, arguably, a more authentic manner. The National Board of Medical Examiners developed a computer-based case simulation in which candidates are presented with a vignette that describes a patient [56]. As virtual time moves on, the candidate can manage the 'patient' in various ways including asking questions, ordering tests, and reviewing patients' notes. The 'patient' develops symptoms and patterns of behaviour related to the underlying medical condition and responds to the actions of the examinee. Scores are generated based on a statistical program that compares candidates' management intentions with those of experts. To develop these computations, expert clinicians rated the approaches to the case of a small number of examinees, and these judgements were then used to derive case-based regression formulas that could be applied to the whole group of test takers [57]. This study found that the scores from the simulation were only moderately correlated with scores from similar tests based solely on MCQs and this suggested that the simulated scenarios were capable of measuring somewhat different skills.

Some attempts to rate the written components of students' or trainees' actual work have been made. One UK study attempted to look at the quality of registrars' referral letters in paediatrics [58]. This study investigated whether a series of criteria could be used to reliably rate the quality of the registrars clinical thinking and communication of their conclusions to general practitioners who were looking after these patients in the community.

In a similar approach in the USA researchers developed a framework that could be used both to teach and to assess students' clinical decision-making [59]. Called the IDEA tool, it contained criteria addressing the Interpretive summary, Differential diagnosis, Explanation of reasoning, and available Alternatives. Medical student new patient admission notes were chosen as the source for rating because written documentation was convenient and made assessment away from the patient, but about the student–patient encounter, possible. The authors reasoned that creating validated documentation standards would contribute to the assessment of clinical skills and was a needed additional element in the assessment of history-taking and physical examination.

A similar surgical exam used mixed SAQ, essay, and algorithm formats, with a focus on key features, all connected to clinical scenarios, to assess clinical reasoning skills in surgery [60]. The scenarios used were linear rather than branching and the content did not vary on the basis of students' responses to earlier questions that, because of computer delivery, could not be revised by the test takers. As the items progress additional information is provided about the case. However, this format allowed students to modify their thinking about the case in later questions. In this way the initial errors in clinical reasoning did not determine performance on higher order questions. The authors claim that this examination is realistic and presents authentic opportunities to 'think clinically' and make decisions reflecting those commonly required of practising surgeons. It requires 1 hour of marking time per student.

The Portfolio

This potential means of assessment, although largely 'written', might contain a wide range of 'learning objects', either as evidence of competence or as records of achievement that have been marked by assessors, such as assignments and project reports. Videos of simulation activity, audio records of discussions, and so on could all be stored. The portfolio is a kind of academic, more controlled, version of 'Facebook' [61]. The use of the portfolio for assessment is discussed elsewhere in this book (see Chapter 18). It may be most useful in the assessment of professional and personal development, and 'self-regulation' of learners, through its longitudinal monitoring capability and the capacity it affords learners to engage in critical reflection on their successes and struggles in the academic and clinical environments.

Selected-response Formats

Multiple-choice Questions: Multiple True/False Formats

Multiple-choice testing was once seen as an enduring option for the reliable and valid measurement of knowledge in 'knowledge-rich' or knowledge-dependent environments such as medicine, bioscience, and engineering. Invented in 1914 by Frederick Kelly, head of a training school in Kansas, USA, by 1926 the multiple-choice test had become the rite of passage for entering post-secondary education in the USA. The MCQ was developed into several forms. One of these is the multiple true/false item, called an 'X type' item in North America, which has become a significant feature of assessment of knowledge in medicine and many other professions over the last 50 years.

In essence, an MCQ is a question that proffers several answers from which the correct one or ones must be chosen. In multiple true/false types a set of options, usually 4 to 6, is given of which each can be either true or false, and the candidate is required to indicate which is correct for each option. An example is shown in Box 21.4.

Over the last few years the multiple true/false item has received a good deal of critical attention. Many examining bodies (for example, the National Board of Medical Examiners [NBME] in the USA) have given up using it altogether. The main reasons have been elucidated with a good

BOX 21.4 Example of a multiple true/false item

Stem	Options	
The following present as chronic (>3 months) airspace disease on a chest radiograph.	**A.** Streptococcal pneumonia	T/F
	B. Adult respiratory distress syndrome	T/F
	C. Pulmonary oedema	T/F
	D. Asbestosis	T/F

deal of empirical evidence [27]. In brief, Case and Swanson [27] state that:

- the distinction between 'true' and 'false' is not always clear, and it is not uncommon for subsequent reviewers to alter the answer key
- reviewers rewrite or discard true/false items far more frequently than items written in other formats
- some ambiguities can be clarified, but others cannot
- to avoid ambiguity, item writers are pushed towards assessing recall of an isolated fact, which is not desirable in most assessment situations
- application of knowledge, its integration, and the synthesis of and judgement around clinical decisions can better be assessed by one-best-answer questions.

It is also the case that using true/false restricts the choice of answers, as discussed in the NBME guidance [27], to a sub-set that can best be classified as 'completely true all of the time' or 'completely false all of the time'. For this reason we will strongly recommend not using this type of item.

Multiple-choice Questions: Single Best Answer

In single-best-answer questions a stem question asks the test taker to choose the one best answer typically from a set of 4 or 5 options. An example is given in Box 21.5.

MCQ items are usually scored optically or directly by computer. There are standard programs for marking and analysing test data straight from a scanner. The answer 'key' – a line of data containing the correct option for each item – is used in this process and should be double-, or even triple-, checked before use. The most common reason for problems at the marking or item analysis stage is a key that contains wrong answers. This may be because the answer has been wrongly transcribed from the item writer's design or (this is not as rare as it should be) because he/she has not provided the best answer.

Most MCQs are scored 1 or 0 for correct or incorrect answers, respectively. Weighting is not necessary for best-answer items; it has very little impact on rankings of students and can reduce reliability. A so-called 'correction for guessing' need not be used [62].

BOX 21.5 Example of a single-best-answer item

Stem	Options
A 32-year-old woman describes pain in her calf when dorsiflexing her foot with her knee in full extension. She advises there is no pain when she performs this same action with her knee flexed. Which of the following muscles is most likely to have been injured?	**A.** Extensor digitorum longus **B.** Fibularis **C.** Gastrocnemius **D.** Soleus **E.** Tibialis anterior

How to Construct a Single-Best-Answer Question

Writing multiple-choice items involves following a series of basic rules that, for the most part, apply to all types. A sensible approach to item construction is to have item-writing workshops that force item writers to work in small (2–3 person) groups. The second best option, for busy people, is to ask individuals to write 4–5 items per lecture/problem-based learning (PBL) session. After either of these activities, the items must be reviewed by a larger group. During workshops the rules are as follows and *should* be applied as a test to each of the items that you construct. Each item should pass *all* the rules.

- Focus on an important (non-trivial) concept, typically a common or potentially serious clinical problem. Avoid trivial, 'tricky', or overly complex questions.
- Focus on how knowledge is applied to a clinical situation, not on recognising an isolated fact or association between concept and exemplar.
- The stem must state a clear question, and it should be possible to arrive at an answer with the options hidden/covered (the cover test). To determine if the question is clearly focused, cover up the options and read the stem to make sure it is lucid and that other item designers can supply an answer dependent only on reading the stem.
- All distractors and the correct answer should be homogeneous, that is, they should fall into the same category. For example, in an anatomy question all answers should be the same type of structure – bones, vessels, nerves, etc. In a clinical item they should all be diagnoses, tests, treatments, prognoses, and so on.
- All distractors should be salient and plausible. Order the options in numeric order or in alphabetical order (see Box 21.5). If you cannot find 4 distractors to accompany a correct option just use 3.
- Try to write questions of moderate difficulty – if any of the item developers have a problem with the item it is probably too difficult. Make sure the correct answer has a sufficiently different degree of correctness when compared to the distractors across all the conditions identified in the stem. For example, let's assume we are assessing knowledge of a condition that affects men, usually in later life. If an incorrect option (distractor) is a diagnosis that does sometimes occur in the age group that the question stem has identified, and the correct answer is one of the rarer diagnoses, the two options may not be far enough apart to make a distinction clear. However, this degree of difference in correctness

between the most correct and the next most correct may vary depending on the level of the examinees. At specialty certification level, for example, clinicians would be expected to be able to make finer distinctions between diagnostic options or management plans, and be more responsive to environmental or epidemiological variations in morbidity. But whatever stage of training the group is at, a moderate difficulty of the item should be sought, and that should be related to expected levels of training or curriculum outcomes, and not to the special interest areas of item developers.

- Avoid technical item flaws. For example, all items and options should be grammatically consistent, logically compatible, and of the same (relative) length as the correct answer.
- Writing questions of the form 'Which of the following is correct?' followed by a set of brief, possibly unrelated postulates, one of which is correct, is not advisable. Such items are basically true/false items masquerading as a one best answer. Furthermore, such items ensure that the questions do focus on trivia, or more likely contain silly or irrelevant distractors. These questions will not be directed at course objectives in a coherent fashion and will likely contain multiple heterogeneous options.

Where's the Evidence for MCQs?

There is far too much research on MCQs to summarise in this chapter. The interested reader should look at recent evidence-based guidelines by Wood [15], Downing [21, 63], and Haladyna et al. [64] for comprehensive treatments of many issues. One interesting fact to emerge is that the number of options to use in a one-best-answer item for maximal reliability is more likely to be 3 or 4 than 5 or 6. There is long-standing theoretical and empirical evidence to support this position [65]. This is because lower reliability or discriminability is generated when the additional distractors, usually put in to provide a standard number of options, are not performing adequately. In items where the 4 or 5 distractors are operating effectively, the item tends to have increased reliability, but this situation is unusual – it's often difficult to find 4 or 5 salient and feasible distractors. There is also contradictory evidence that extended matching question distractors (see Box 21.7), usually a naturally occurring fairly large set of 10–15, may operate more effectively than 3 or 4 pre-selected ones [27, 66]. We would hope that the processes that students use to answer one-best-answer MCQ items are at least analytical and frequently reasoning-rich. However, evidence suggests that 'the problem with multiple-choice items is not that they are mere exercises in recognition, but that we are unable to predict the processes that will be evoked' [67, p. S9].

Integrated Single-Best-Answer (ISBA) Items

Recently we have been experimenting with an item format that can give information about the capacity of students to integrate knowledge around topics. This is an attempt to capture the alleged benefits of student-centred learning strategies, such as PBL and case-based learning, and assess students in a way that mirrors their learning. The ISBA type of item sits somewhere between SBA and extended matching and key feature formats (see below). An example of an ISBA is given in Box 21.6. In the integrated item some questions from different disciplines or clinical specialties are asked about the case that is the focus of the stem or vignette. The stem may be changed slightly, or information added (as in Parts 2–4 in Box 21.6) to give a broader picture of the case, and to probe into the basic science or other important elements of the presentation.

In Box 21.6 the case starts with a woman who has increasing shortness of breath when exercising over a period of a few months which broadens to include spirometry results, necessary lab tests, and basic mechanisms. Here we show 4 of 8 linked MCQs for this topic. For summative assessment

BOX 21.6 Example of an integrated single-best-answer item

Sigrid, a 39 year old woman, presents to her GP complaining of increased dyspnoea on exertion for several months. She reports she has always believed she has had some form of asthma or chronic lung infection, but lately she has had a great deal of difficulty performing any activity without shortness of breath. She says she has no current cough, haemoptysis, chest pain, weight loss, night sweats, or fevers. Sigrid indicates she used to smoke a few cigarettes a day while she was young but quit a few years ago. On examination a mild expiratory wheeze is present. The chest X-rays are shown in Figure 21.1.

1 This patient presentation and CXR is most consistent with which of the following:
  - **A.** Chronic recurrent asthma
  - **B.** Diffuse bilateral bronchopneumonia
  - **C.** Emphysema
  - **D.** Pulmonary fibrosis

2 On spirometry, FEV1 is 40% of predicted with a 9% improvement after bronchodilator. The FEV1/FVC is 0.50. Which of the following is the best interpretation of these spirometry results?
  - **A.** Severe obstructive disease without bronchodilator reversibility
  - **B.** Moderate obstructive disease with bronchodilator reversibility
  - **C.** Severe reversible airway obstruction (asthma)
  - **D.** Moderate restrictive airway disease

3 Which of the following tests would be indicated to investigate the underlying cause for Sigrid's lung disease?
  - **A.** Cystic fibrosis screening
  - **B.** Alpha one antitrypsin levels
  - **C.** Sputum culture
  - **D.** WBC differential looking for eosinophilia

4 Sigrid is found to have a deficiency of alpha one antitrypsin (AAT). What is the consequence of this deficiency that is believed to accelerate the development of COPD?
  - **A.** Inability to inactivate neutrophil elastase
  - **B.** Inability to phagocytise pathogenic bacteria
  - **C.** Inability to repair alveolar structural proteins
  - **D.** Inability to synthesise serine proteases

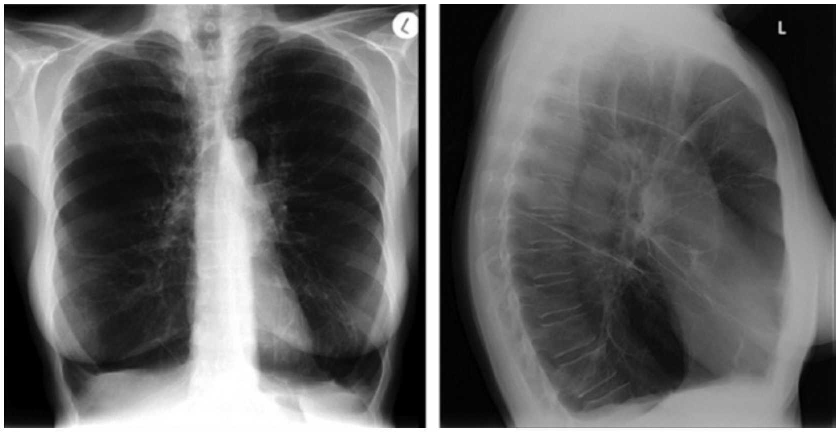


Figure 21.1 Case courtesy of Dr Andrew Dixon, http://Radiopaedia.org, rID: 9674

these items may need to be delivered by computer for avoidance of cueing, but used formatively they can test students' integration of basic science as well as clinical reasoning. The same guidance applies to ISBA as was discussed under the section on SAQs above. The major issue with exploring integration is that items are clustered around single cases. This may restrict the number of cases that can be covered in a test of set length. However, as one of the goals of many curricula is to integrate knowledge, if assessment designers want to test integration without resorting to SAQs, this is one option.

Extended Matching Questions

Extended matching questions (EMQs) were developed in the early 1990s [27]. However, the kernel of the idea was first conceived by Sue Case in her PhD thesis as early as 1983 [68]. She and David Swanson are credited with most of the development work on this format, whilst at the National Board of Medical Examiners. An EMQ is a selected-response item in which the item stem has been extended, usually, to a short clinical vignette or scenario and the choices have been extended to include all potentially acceptable ones for the clinical problem or issue that is being addressed by the item. This format was originally targeted towards the application of clinical knowledge to diagnostic and management problems, but has been extended to other areas such as basic science. In the example (Box 21.7) there are 20 options pertaining to the theme of diagnosis of abdominal pain. This is followed by one or more clinical vignettes. The options are all causes of abdominal pain. It is usual in such items to attempt to make all the questions and options homogenous in this way, so other issues concerning abdominal pain, such as initial management or investigations, are not included. An item should focus on a specific area of clinical cognitive activity that pertains to a specific phase of the clinical process – in this case diagnosis.

EMQs are usually scored 1 for a correct response and 0 for an incorrect one. It is sometimes possible to have more than one best answer – for example, when two (or more) diagnoses are equally likely, given the information in the vignette. However, the scoring of these requires more

BOX 21.7 Example of an extended matching item

Area: Abdominal pain – Diagnosis
Options

A. Abdominal aneurysm	K. Kidney stone
B. Appendicitis	L. Mesenteric adenitis
C. Bowel obstruction	M. Mesenteric artery thrombosis
D. Cholecystitis	N. Ovarian cyst – ruptured
E. Colon cancer	O. Pancreatitis
F. Constipation	P. Pelvic inflammatory disease
G. Diverticulitis	Q. Peptic ulcer disease
H. Ectopic pregnancy – ruptured	R. Perforated peptic ulcer
I. Endometriosis	S. Pyelonephritis
J. Hernia	T. Torsion

Lead In: For each patient with abdominal pain described below, select the most likely diagnosis.
Scenario/Stem: A 25-year-old woman has sudden onset of persistent right lower abdominal pain that is increasing in severity. She has nausea without vomiting. She had a normal bowel movement just before onset of pain. Examination shows exquisite deep tenderness to palpation in right lower abdomen with guarding but no rebound; bowel sounds are present. Pelvic examination shows a 7-cm, exquisitely tender right-sided mass. Haematocrit is 32%. Leukocyte count is 18 000/mm^3. Serum amylase activity is within normal limits. Test of the stool for occult blood is negative.
Answer: __
[Next scenario in the domain (diagnosis of abdominal pain) would appear here.]

Source: National Board of Medical Examiners [27].

attention during the scanning process, as for optical scoring two (or more) passes of the score sheets are necessary with each correct answer keyed on each pass. Unless it is clinically important to be able to recognise both potential diagnoses from the same vignette, such multiple responses are probably best avoided (for example, by removing one of the two diagnoses from the list).

Investigations of the reliability and construct validity of EMQs suggest that they have good measurement properties, and correlate well with other measures of recall, recognition, and cognitive functioning [69]. Recently, studies of experts and novices who talk aloud whilst trying to complete EMQ items have strongly suggested that EMQs have good construct validity for, and can be reliably used to assess, clinical reasoning [70], even though novices and experts approach the same item with different strategies (backwards versus forwards reasoning, respectively) [71]. Furthermore, when used in pathology EMQs are more reliable, better discriminate the competent from the borderline student, and can be written definitively to test core content [72].

EMQs seem to be easier to write than true/false or other types of one-best-answer items, because in that style of item the convolutions that writers go through to reduce the item set to a smaller number where each is true or false, or there is clearly one best answer, are not needed [73]. Clinicians from some disciplines such as public health, epidemiology, and statistics have suggested that EMQs are difficult to write for these content domains, but recent articles suggest they have been adopted or are being developed in some of these hitherto unexplored areas (e.g. psychiatry [74]).

How to Construct an EMQ

It is best initially to write these items by considering the area or domain of the assessment blueprint for which items need to be written (e.g. abdominal pain in Box 21.7). Then a general question is posed, followed by all the possible answers to that question (e.g. 'What are the causes of abdominal pain in adults?'). After these have been identified, scenarios that pertain to one or more of the answers are constructed. Ideally, create items (particularly the stems/scenarios) in pairs of writers, at workshops of about 8–12 people in total (4–6 pairs), with review every 2 hours or so in a larger group. This is an effective, and in most examiners' experience, an enjoyable way of generating items. The stages are as follows.

- *Identify the domain or subject for the set*. The domain is an area of cognitive activity (e.g. diagnosis, management planning). The subject can be a presenting complaint in a body system or systems (e.g. abdominal pain, so that diagnosis is the focus of the item), or a pre-diagnosed condition (e.g. community-acquired pneumonia, so that management is the focus of the item). Sometimes it might be appropriate to move directly from a non-diagnosed presenting complaint (e.g. abdominal pain) to an investigative option (e.g. ultrasound) or management plan (e.g. restricted diet). However, the more cognitive steps involved in moving between the presenting complaint and the focus of the item (e.g. asking about management), the less will be known about why an examinee might have answered the item incorrectly. For example, the examinee might have thought a patient with ulcerative colitis had appendicitis and ordered surgical intervention.
- *Write the lead-in for the set*, e.g. 'For each patient described below, select the most likely diagnosis' (Box 21.7) The lead-in indicates the relationship between the stems and options. It must be a clear question for examinees. It is an essential component of an extended-matching set. Sometimes two lead-ins can be written at the same time – for example, one based on diagnosis and one on indications for investigations or management. Subsequent scenarios can be used, usually with only minor modification, with either lead-in. In summary, the lead-in should consist of a single, clearly formulated task so that the examinee could, if necessary, create an answer without looking at the options.
- *Prepare the list of options*. The list of options should be single words or very short phrases. This list is best developed in a whole-group format. It will be generated in a fairly random order, but the options should be rearranged in alphabetical order for the final item presentation. For example, the initial list for Box 21.7 should contain all the likely causes of abdominal pain as options. Sometimes there are specific causes that occur only or predominantly in a particular subset of the population – for example, in women (e.g. ectopic pregnancy), in men (testicular torsion), in the elderly (dementia). Such options can sometimes become 'zebras' [27], which stand out as so obviously applying to one subgroup of patients that their inclusion is ill advised. In Box 21.7 there are some such options, but there are also sufficient important differential diagnoses in the list to warrant their inclusion for the given scenario.
- *Write the stems*. The stems (items) within a set should be similar in structure. Most often, patient vignettes are appropriate. The scenario should contain all the information that one would normally expect to be available from any conscious patient: the presenting problem, the history (including duration of signs and symptoms), the physical findings, and then the results of any immediate diagnostic tests carried out. Sometimes, for a complex case, further data pertaining to development of symptoms over several days might also be given – for example, initial management and subsequent clinical changes. Scenarios can include a smaller set of information, but it is unwise to exclude the information that would normally be collected by or available to the test taker in the real clinical context at the time they were seeing this patient. Specifying this information in a standardised order makes shorter reading time and hence allows more items to be delivered in a given time.
- *Review the items*. Make sure there is only one 'best' answer for each question. Having two right answers is possible, but entails more marking effort than it is usually worth. Also make sure that there are at least 4 reasonable distractors for each item to minimise guessing effects. Evaluate the extent to which the lead-in clearly identifies the task. See if the other examiners can create an answer without looking at the options. Satisfying this 'cover-the-options' rule is an important feature of a good question because, if the examinee cannot do that, it means the question is too vague, is not appropriately targeted to the skills the examination is testing,

or exhibits some other flaw of test writing. As a final check, review the items (without the correct answer indicated) across other pairs in the writing group. If the pair has difficulty determining the correct answer, modify the option list or the item to eliminate the ambiguity.

Where's the Evidence for EMQs?

There is evidence that some MCQs, of which the EMQ type seems to be the most suited to clinical tasks, can involve substantially more than recognition of knowledge learnt by rote memorisation [67]. These authors nevertheless suggest that unfocused items or those with negatively worded stems, as sometimes, of necessity, occur in the typical true/false or best of five types, do not appear to provoke problem-solving skills and forward-reasoning. Whilst the evidence for the link between item type and cognitive response is being developed, they suggest concentrating on items that are low-fidelity simulations of clinical situations with examinee tasks that are relevant for them (e.g. diagnosis and management). EMQs are ideally suited to this role. EMQs also substantially reduce the likelihood of obtaining a correct answer by chance alone.

Although this area is fraught with controversy, and not all of the options provided for any one EMQ stem will be active for that item, modelling suggests that EMQs with between 7 and 12 active distractors will provide good insulation against the need to be concerned about the so-called 'guessing' factor in multiple-choice tests [75]. Research on EMQ formats has shown that a reduction in the length of item option lists, from the 15 to 20 previously thought necessary, is possible without much, if any, deleterious effects on item quality [76, 77]. Eight options seem to be a reasonable minimum number. In general, items with more options are more difficult, require more time to complete, and nevertheless have similar discriminating properties to items with 8 options. Reducing the whole list to a 'shortlist' of 8 or so can be done by carefully constructing physician panels to select the most appropriate set. Moreover, providing the panel with item-response statistics from the long item does not seem to improve option selection. The use of a smaller number of options reduces time spent on each item by candidates and therefore increases the number of items that can be used in a set time [76, 77]. Swanson et al. advised, 'We plan to begin advising [examiners] to reduce the number of options included on option lists in order to make more efficient use of testing time' [76, p. S95]. That advice may now be prudent to implement, as student numbers in medical schools have increased dramatically in a number of countries in the last few years.

Script Concordance Items

Over the last 15 years interest has developed in constructing a multiple-choice test that can reflect clinicians' capacity to weigh evidence in a clinical encounter. This work has its foundations in a clearer understanding of how clinicians approach the diagnostic task and how this information is remembered [78, 79]. Recently, this work has expanded dramatically; in the last 5 years more papers on this issue have been published than in the previous 20 years. The reason for this is probably that the value of appropriately and efficiently stored knowledge in clinical decision-making, and the need to assess these cognitive processes, have been accentuated through literature on poor decision-making and its relationship to patient safety.

Memory for clinically important information is developed in stages. As a student, possibly because of the preclinical/clinical divide in most educational programmes, biomedical knowledge dominates. Additional insight into clinical problems develops, as students gain more experience with patients and elaborate their knowledge into explanatory frameworks and linkages between symptoms, causes, basic mechanisms, and management. With more experience of patients, these frameworks depend less on biomedical detail and more on 'illness scripts' that involve applied (functional) knowledge and relate to both common (easy) and uncommon (more difficult) patient presentations. Eventually 'owners' of these illness scripts use them to promote rapid recognition of patterns of patient presentation, and reconcile any unfamiliar presentations with previous presentations. However, when a challenging presentation occurs, expert clinicians muster their existing knowledge and strive to activate biomedical principles and knowledge (albeit usually tied to specific patient or contextual exemplars), but these occasions occur less frequently as expertise develops [80]. Previously, attempts to create tests that reflect this process through patient management problems and modified essay questions have foundered, primarily because clinicians disagree about these issues and this, coupled with the variety of constructed responses that such items can generate from the test takers, results in great difficulty in scoring them. Moreover, in a typical MCQ item when the test makers cannot agree on the best answer, either before or after it has been used, the item is usually removed from the test.

However, a group in Canada [79] have devoted a considerable amount of effort to the task of developing an item format that overcomes these problems – this is the script concordance item (SCI).

The SCI is a selected-response item that depends on respondents choosing how well pieces of information contribute to a diagnostic or management strategy for a particular clinical problem. As described above, the script is the internal rubric that experts use to classify data and generate or choose a hypothesis quickly. In the SCI, concordance is the congruence of the test takers' script with elements of rubrics deemed by 'experts' to be most plausible. The items estimate the relative likelihood of diagnoses, given a certain piece of clinical or biomedical information. The test items ask for choices to be made based on the subjective clinical probabilities of the test takers. However, they are scored in such a way that takes into account the degree of similarity (concordance) between the illness scripts of the test takers and the test makers. An example is shown in Box 21.8.

In its original format the marking scheme is based on the performance of an expert group, and there are various ways to derive the marks. In one method a consensus is reached between the experts on the best answer for each item. In others, each response attracts a mark that is proportional to the frequency of its choice by the expert group.

Recommendations for exactly how this is done vary according to source. Some authors suggest awarding 1 mark for the experts' model answer and then giving marks for alternatives based on the frequency ratios for the other answers. For example, if a group of 20 experts opted 12, 6, and 2 for the options +2, +1, and 0 respectively in any one item, these options would be awarded 1, 0.5, and 0.167 marks, and the others zero marks. Other sources suggest awarding marks on a percentage distribution basis, so this would work out at 0.6, 0.3, and 0.1 per option for the three 'correct' options.

Another interesting feature of the marking scheme is that it can have more than one reference panel, depending upon the context in which the item is to be used; for example, in a metro or rural environment, a general practice, the emergency department, or medical ward. This is because the probabilities of some decisions or hypotheses may change depending on the context.

A test made up of SCIs is therefore a very flexible one, and one capable of rewarding partial or incomplete knowledge to some extent. Data so far suggest that a test composed of SCIs is very efficient; good reliability can be achieved with relatively small numbers of items [81]. The way that SCIs are constructed clearly has implications for how best to use them. They are a specialised item suitable for investigating test takers' ability to formulate and progress specific diagnostic and management decisions.

A systematic review of approaches to SCI development has identified several useful strategies in writing and scoring SCIs [82], which have implications for reliability and validity. The number of experts needed on the concordance-generating panel has also been researched. Samples of at least 10 panellists provided satisfactory internal consistency, and there was little gain when panel sizes exceeded 20. Larger panels give rise to higher cohort mean scores, presumably because the larger the panel, the higher the chance of an option being awarded weight by panellists, especially when the cases intrinsically demonstrate high uncertainty.

In an SCI expert panel, some panellists may act in an idiosyncratic manner with their answer, choosing options that are deviant or clearly incorrect, especially in tests with high degrees of uncertainty. This may concern test designers or users of the test information, even though credit awarded for these answers is small. Removing some discordant responses or all the responses from deviant or low-scoring panellists (because a low score would suggest that those experts were not so expert after all) has been suggested. However, the psychometric impact of excluding such panellists' responses in score derivation is minimal, as long as the panel has 15 or more members. However, if educators wish to remove these answers, the available methods [82] to do so appear equivalent in terms of psychometric consequences.

BOX 21.8 Example of a script concordance item

A 25-year-old male patient is admitted to the emergency room after a fall from a motorcycle with a direct impact to the pubis. Vital signs are normal. The X-ray reveals a fracture of the pelvis with a disjunction of the pubic symphysis.

If you were thinking of	And then you find	This hypothesis can be rated: (circle best response)
Urethral rupture	Urethral bleeding	−2 −1 0 +1 +2
Retroperitoneal bladder rupture	Bladder distension	−2 −1 0 +1 +2
Urethral rupture	Perineal haematoma	−2 −1 0 +1 +2

Where:
−2 = the hypothesis is almost eliminated;
−1 = the hypothesis becomes less probable;
0 = the information has no effect on the hypothesis;
+1 = the hypothesis has become more probable;
+2 = the hypothesis is very likely to be correct.

Source: Charlin et al. [67].

How to Write a Script Concordance Item

An SCI is created in a similar fashion to an EMQ, except that the specific diagnostic decision is the key to the choice of vignette, rather than a list of potential diagnoses, and hence there may be fewer data in an SCI than an EMQ. There are two stages, as follows.

- A vignette is created containing data that present a challenging clinical situation. This is usually text, but other information, such as an X-ray or other pertinent data such as blood analysis, can be given, depending upon the test takers' clinical decisions that are being investigated. Not all the data needed are provided because these can be revealed as part of the item hypotheses–information links. Each of these clinically relevant pieces of information might help the clinician refine, improve, confirm, or eliminate the clinical decisions or hypotheses made about what is happening. The response required from the test taker is to appraise the effect of each piece of new information on their hypothesis.
- Responses are pre-prepared using a five-point scale running from −2 (the hypothesis or decision is much less likely) to +2 (the hypothesis or decision is much more likely). To get the best out of the item, and to sample knowledge broadly, there should be no links (cues) between the different provided responses. Once designed, a group of experienced clinicians complete the SCIs, and their answers are collated. These answers are then used to decide on the marking scheme for scoring items.

Although it would seem relatively easy to design such items, Charlin and Van der Vleuten [83] warn that the expert group writing the items must be familiar with the tool and be able to choose cases that are complex enough to fit with the level of training being assessed. Also the item is designed to assess decision-making that is grounded in evidence, as opposed to received wisdom, so there must be enough data available on the clinical solution to the case and the links provided in the case must be widely known by experts in the area.

Where's the evidence for script concordance items?

For reliability and validity purposes, attention should be paid to both the number of cases sampled in the test and the number of items used per case [82]. In general, sampling between 25 and 36 cases, with approximately three items per case should result in an SCI test with reliability of between 0.75 and 0.86. Adding items (e.g. more than one per case), rather than cases, is more effective in increasing test reliability and can reduce the workload of test designers, but there appears to be a ceiling effect after 3 or 4 items per case.

SCIs focus to a large extent on clinical judgement in areas of uncertainty. This is what makes them 'content'- and 'construct'-valid for the measurement of clinical judgement and reasoning. Good construct validity has been shown in a number of studies. SCIs discriminate between different levels of surgical expertise and have also been used in pharmacy education [84, 85].

Because SCIs are used on cases that are often far from clear-cut, this may mean that SCIs are unsuited to junior students with less experience of clinical conundrums than trainees with more clinical experience. Indeed a study comparing SCIs to traditional MCQs showed that interns rated them higher as a method of assessment and also these items' reliability was higher in an intern cohort, as opposed to a medical student cohort [86]. An SCT test has also been shown to be a useful adjunct to a suite of tests designed to examine the performance of poorly performing doctors [87]. There has been increasing controversy around the appropriate scoring methods for SCTs [88–90]. The discussion has centred on the extent to which the competing hypotheses, apart from the modal value of the expert panel, should attract marks, and the bias that might be introduced by test takers who do not use the extremes −2 or +2 because in the nature of clinical decision-making these will occur less commonly in items. However, a recent study [91] showed that a group of different scoring approaches correlated quite highly with the exception of one based on the single best answer (the mode) attracting 1 mark and the other options zero. This study also showed that an SCT focused on decision-making of 4th and 5th year medical students around presentations of the acute abdomen did not correlate with an MCQ covering the same content, indicating that perhaps the SCT was measuring a construct beyond factual knowledge. In addition, one of the critics of the current scoring systems, Kreiter [92], has recently proposed some modifications using Bayesian probability that could be used to make the SCT style of item scored more reliably. This style of item is still under development.

Formats using both Selected and Constructed Responses

Key-features Items

The key-features item (KFI) is an SAQ that can use both selected and constructed responses. The defining characteristic of a KFI is that it is aimed at assessing whether the test taker can recognise, deduce, or infer the most important features of a clinical problem and, if required, subsequently choose the most salient, urgent, and effective management strategies for that clinical problem [93].

Readers may think that it could be reasonably assumed that all SAQs would possess these properties. Unfortunately, in the 1970s and 1980s this was not the case [94]. Curricula and assessments were frequently designed from discipline or specialty perspectives, with each concentrating on unusual or esoteric aspects of their craft in order to discriminate between the truly well-grounded student and the rest. This led to a culture in which trivia, rare morbidity, atypical presentations, and specialty-specific issues dominated short-answer tests. It was also believed that decision-making skill was generic, so many short-answer tests focused on one or two problems, with the 'short' answers practically exhausting the available knowledge about that problem [95].

Two researchers, Georges Bordage and Gordon Page, coined the term 'key feature' after a critique of and further research on the nature and assessment of clinical decision-making skills [96]. Their central concept was that 'in any clinical case, there are a few unique, essential elements in decision-making which, alone or in combination, are the critical steps in the successful resolution of the clinical problem' [96, p. 1189]. With funding from the Medical Council of Canada, their concept led to the creation of a new format for assessing decision-making skills. Furthermore, by assessing only critical steps, candidates were tested both on important objectives and on a much larger number of clinical problems than was the case with previous formats. A typical key feature item with the scoring key is shown in Box 21.9.

The key features tested by such questions are: (i) the ability to synthesise presenting complaints for recognition of an important and life-threatening condition; (ii) the ability to identify essential areas to investigate in the history to confirm or rule out this hypothesis.

In Box 21.9, Questions 1 and 2 directly address each of these key features. Each item challenges the candidate to apply his or her knowledge in making clinical decisions. In this item there is a constructed response for Question 1, with only one answer required, and a selection of six options only from a list of 28 for Question 2.

Usually the test taker would write both answers to both Question 1 and Question 2 on an optical scoring sheet. There would be space for free text to Question 1 and buttons or ovals to fill in for Question 2.

KFIs require a scoring rubric that reflects the importance of the various key elements of the case. It helps to test out the scoring rubric with colleagues and a few 'dummy' test takers. Higher weighted scores for more important answers should be encouraged. In Box 21.9, the diagnosis was awarded 1 mark and the elements of the history a total of 6 marks. The scoring key shows that only 7 of the 28 options would be considered appropriate, so the candidate must select 6 of these to score maximum marks. The question also weights aspects of the history more highly than obtaining the correct diagnosis. This may be because in this situation knowing the best questions to ask would probably result in a change of diagnostic preference, if an inappropriate answer was initially given to Question 1. There are very

BOX 21.9 Example of key features item

Paul Young, a 56-year-old male, consults you in your surgery because of pain in his left leg which began two days ago and has been getting progressively worse. He states his leg is tender below the knee and swollen around the ankle. He has never had similar problems. His other leg is fine.

Question 1

What is your principal working diagnosis? List, in note form only, your single (1) diagnosis.

Answer 1

....

Question 2

To establish your diagnosis, what elements of his history would you particularly want to elicit? Choose up to six (6) from the following list.

1. Activity at onset of symptoms	15. Palpitations
2. Alcohol intake	16. Paraesthesia
3. Allergies	17. Paroxysmal nocturnal dyspnoea
4. Angina pectoris	18. Polydipsia
5. Anti-inflammatory therapy	19. Previous back problems
6. Cigarette smoking	20. Previous knee problems
7. Colour of stools	21. Previous neoplasia
8. Cough	22. Previous urinary tract infection
9. Headache	23. Recent dental procedure
10. Haematemesis	24. Recent immobilisation
11. Impotence	25. Recent sore throat
12. Intermittent claudication	26. Recent surgery
13. Low back pain	27. Wounds on foot
14. Nocturia	28. Wounds on hand

Answer 2

1. .... 2. .... 3. .... 4. .... 5. .... 6. ....

Scoring Key
Question 1

Deep venous thrombosis must be the differential diagnosis (Score 1)

Total requested = 1. Total accepted = 1. Total score = 1.
More than one answer scores 0.

Question 2

Any six (6) of the following items are required:
Activity at onset of symptoms
Cigarette smoking
Previous knee problems
Previous neoplasia
Recent immobilisation
Recent surgery
Wounds on foot
Each scores 1. Total requested = 6. Total accepted = 6. Total possible score = 6.
More than 6 answers scores 0.

Source: Royal Australian College of General Practitioners Key Features Practice Paper 2007.

few question formats that truly assess clinical decision-making, but KFIs can make this claim. Certainly, the foundation work done on the validity and reliability of the format [96] would suggest that KFIs robustly measure these facets of competence, and test takers seem to like their salience to clinical practice. However, KFIs are difficult to construct well, and although scoring seems simple, some questions can be very difficult to weight appropriately. Standard-setting with KFIs can also be complex, as although both the modified Angoff and Ebel methods have been used, some assessors report that they do not work very well for this format. This is because the items are generally multi-dimensional and can include many aspects of a case, some of which are interdependent. Making decisions about how borderline test takers would respond to such items is challenging.

How to Write a KFI

- Select a clinical problem from the assessment blueprint in which analysis of the context, identification of clinical conditions, and synthesis of the diagnosis and/or management (clinical decisions) are required objectives.
- Think of several real instances of the case in everyday practice. With respect to these cases, consider what the essential (necessary and sufficient) steps are to resolve this problem. This will allow you to focus solely on the most critical decisions for each case. Make an effort to distinguish between decisions or steps that are appropriate, but not critical, and those that are really vital. It can help to elucidate the case's key features by identifying the attributes of or inconsistencies in the presentation that are most likely to result in clinical errors in dealing with the case by the particular test takers.
- Any of the typical cognitive processes used or actions planned can be tested in KFIs – the scope depends on the objectives being tested and the case presentation. For example, framing initial hypotheses, looking for particular clinical findings, choosing tests, management options, or specific drugs are all possible.
- Select one of the real cases for development into a problem scenario and related questions. The attributes of the case can be written according to the same rules as for EMQ items (see the section 'How to construct an EMQ'), that is age, sex, setting of the encounter. When you have written the scenario, use the key features to frame the questions for the case. Usually only two-to-three questions are possible without invoking prompting or cueing. The more questions are asked about a particular case, the more opportunities arise for cueing. For example, in order to ask about management, further information may need to be given about test results that will strongly imply the diagnosis, so asking about differential diagnosis, history and investigations, management, and follow-up all in the same item becomes difficult. (This is the same issue that arises in MCQs, see the section 'How to construct a single-best-answer question'.)
- Select the response format, or formats – write-in (constructed) or option choice (selected). This needs to be done carefully. The guiding principle is that if the key feature

entails a cognitive process that is or should be generated by the professional (e.g. recognition of a myocardial infarction and decision on immediate action), then the response should be write-in. The number of required answers must be stated. This is to inhibit 'blunderbussing' – the tendency of test takers to write down everything they can think of in the hope that the correct answer(s) are included in their responses. Exceeding the number required should be penalised. Write-in items need to be marked, whereas options can be electronically scored.

Where's the Evidence for KFIs?

KFIs have been used for both summative and formative assessment in a number of different contexts [97, 98]. The reliability of the KFI paper (26 items) of the Royal Australian College of General Practitioners increased from 0.64 in 1999 to around 0.83 in 2005 [99]. KFIs were rated adequate on tests of clinical decision-making skills by approximately 90% of candidates in the 2004 Fellowship examination. Colo-rectal surgeons (N = 256) found them useful as a means of self-assessment and nine KFIs were almost as reliable as 50 MCQs (0.95 versus 0.97, respectively) [81].

Although research on KFIs has not kept pace with that on EMQs and SCIs, it is notable that the generation of key features for a wide range of common priority topics in family medicine has been the focus of work by the College of Family Physicians of Canada over a 7-year period [100]. The purpose of this work was to generate key features for use on other approaches to learning and assessment beyond the specific KFI methodology

Computer-based Written Assessment

With the advent of accessible computers with high processing power to individual teachers and assessors, a number of the assessments discussed in this chapter can now be designed, delivered, and marked on a computer. Even the SAQs and perhaps essays will soon be marked by computers. Another feature open to computer-delivered written items is the possibility of individually adaptive tests. Both the NBME and the Australian Medical Council are already using these. Computers deliver items to examinees that are targeted towards candidates' ability level. Further items are titrated by the computer against that individual's difficulty level based on their answers to the previous question(s). This continues until the computer is satisfied (using predesigned algorithms) that it has reliably identified the candidate's ability level, and it then stops the assessment process. This has enormous potential for delivering tailored assessment, protecting large item data banks from piracy or examinee recall, and being time efficient because very few examinees would get the same test.

The computer also has the capacity to deliver immediate feedback. This has given rise to a new item format – the F-format – that is currently under development. F-type items [101] have been trialled to fill the gap left by the demise of patient management problems [94]. F-types are essentially a hybrid of extended matching and PMP styles. F-type tests present authentic patient management scenarios that develop over time. Typically, two-to-three test items sit within one case. Each item addresses an objective relating to one or more clinical decisions that could be made during the case. The test items occur in a fixed sequence within the case; items start with patient presentation and end with management or resolution. Because the items within each case reflect the unfolding nature of the case over time, sometimes examinees learn about how successful their previous answers were from subsequent items. Because such cues might affect examinees' decisions, they are not permitted to revise their responses to earlier within-case items; the computer bars their access to going back to change responses. However, examinees can be given feedback or not (depending on the item) after completing each item. Post-answer feedback on items allows all examinees to be kept on an equal footing, as the case develops. For example, when a clinical scenario is presented, examinees would be asked to identify the most appropriate next step in investigation. On the next test item, the scenario might be presented as if the correct next step had been taken, and the patient is being sent for a further definitive investigation. Presenting this updated information automatically 'gives away' the correct answer to the previous question but may enable someone who has previously answered incorrectly to get back on track with the new context. In this way examinees may improve their understanding of the case and/or be able to answer subsequent items without being disadvantaged by their previous inaccurate perceptions. The items types are still in development but look quite promising for areas in which cases can be quite complex or for which it is difficult to write items without cross-item cueing, such as key-feature tests. These enormously exciting developments herald the possibility that there could be a whole new approach to item-writing that uses elements of different formats all in the same item.

Item Analysis

After items have been used in a test, we wish to know how well each item has performed. In essence we want to know the following.

a) Did the item do its job in the understanding of an individual on that particular element of the content of the test? This would include identifying any problems with its construction that led to misinterpretation by test takers. For selected-response items, distractors might have been included that did not work effectively, or were too close to the correct response to allow even good candidates to make an appropriate choice.

b) Did the item have demonstrable relationships with the content of other items on the test and with the content of the test as a whole? This would include looking at inter-item correlation, looking at item-test correlation, looking at relationships with other parts of an assessment regime – for example, other assessments of the same domain, and other assessments in different domains, such as procedural skills.

One way to approach problem (a) is to read every item on the test again. Of course, most items will have been read many, many times already, and after reading hundreds of items, test designers will have become immune to the potential inadequacies of the items.

Another way is to use basic item analysis. However, all item-analysis data must be interpreted with extreme caution and with reference to the overall purpose of the test. Rules of thumb about deleting or modifying items (for example, based on poor discrimination indexes for items) should probably be avoided.

Many programs are available to process item data, and most operate in similar ways. The core of item analysis is to look at a set of parameters for each item that describe its performance as an item, its relationship to the test as a whole, and its contribution to the performance of the current cohort on that test. In a selected-response test such parameters include the following.

- The difficulty of the item – how many test takers got the item correct?
- Which of the options were chosen most frequently?
- Which options were chosen by high, medium, and low performers on the test as a whole?
- How did success on each item correlate with performance on the test as whole?
- What were the inter-item correlations? (This might be particularly important for short answer and key features questions.)

This section demonstrates how to use a typical set of item-analysis data to make inferences about these questions. The particular one shown here is from the IDEAL Consortium item analysis software, but many others provide analogous output.

Consider the item in Box 21.10 which is item 87 from a 100-item MCQ Test on Emergency Care for Year 3 Medical Students. A cohort of 148 students (see the N for TOTAL in Box 21.11) took this item as part of a 100-item test on emergency management. The item was chosen because it was thought that students needed to be aware of the repeated changes to guidelines over the years and that recognition of the correct guideline was a vital decision. Some previously recommended ratios were included as distractors. The output from the analysis software for this item is shown in Box 21.11.

Box 21.11 shows a number of item parameters in the top row, then below this row it can be read as a table comparing the proportions of whole cohort (row TOTAL) and tertile groups (rows HIGH, MID, and LOW) who selected each of the options in the test. The final three rows describe parameters of the various groups who selected the different distractors.

Returning to the areas, we need to investigate the following.

- Did the item do its job in capturing the understanding of an individual on that particular element of the content of the test?

Well, the item was completed by all test takers. (The values for INV, the number of examinees not providing a valid response to this item, and for OMIT, the number of examinees omitting this item, was zero.) Furthermore, the number of examinees not finishing the test from this item onwards (NF) was also zero. Takers did not get 'stuck' answering this item and not have time to finish as a result. As this item was number 87 of 100, we can probably conclude that all candidates finished the test, and we already know that all did this item. Most test takers got it right – 90% in fact, expressed in this table as 'DIF' (the proportion of candidates who got the item right, 0.90. In some data analysis packages the difficulty index is often more appropriately called the 'facility' index). This number also appears in the box for the proportion of the TOTAL test takers who identified B as the correct answer – 0.90 in Column B. So if this item covers core knowledge the indications are that this was achieved in the vast majority of candidates.

How did the item work? Well, as we have seen, there were very few responders who got the item wrong (only 15 of 148; 0.10 as a proportion). The proportions choosing each of the distractors varied between 0.03 and 0.01 (see the other proportions in the TOTAL row of the table). Because of the low numbers choosing distracters, it is not possible to confidently say that any one of the distractors was more or less attractive than the others. The TEST SCORE MEAN % is the mean score on the total test of those candidates giving the indicated response to this item. Variation in this row would indicate that perhaps there was one, or more, distractors that was attractive to a high- or low-scoring group. If a particular distractor was chosen by a group that collectively did well on the test overall, there may have been some teaching that impacted on this issue for those students. This may have been the case for the 5% of the cohort choosing C, the old guideline, who collectively scored 79% on this test. However, the very small numbers involved would make this interpretation risky.

Now we will consider the second issue:

- Did the item have demonstrable relationships with the content of other items on the test and with the content of the test as a whole?

This would include looking at inter-item correlation, and looking at item-test correlation.

RPB is the point bi-serial correlation between item 87 and the score on the whole test, including the present item. Point bi-serial correlations are used to determine the

BOX 21.10 Example item for analysis (i)

Item 87

In Melbourne, Australia, a man collapses in the presence of a single health care worker who, after careful assessment of the patient, and whilst waiting for the paramedic team, decides to deliver cardio-pulmonary resuscitation (CPR).

Choose the most appropriate ratio for breaths : compressions with which the health worker should commence CPR.

A. 1 : 30
B. 2 : 30
C. 2 : 15
D. 3 : 15
E. 3 : 30

* The correct answer (Australian Resuscitation Council Guidelines) is **B**.

BOX 21.11 Item analysis data for item 87

DIF = 0.90, RPB = −0.039, CRPB = −0.075, (95% CON = −0.298, 0.157)
RBIS = −0.068, CRBIS = −0.130, IRI = −0.012

GROUP	N	INV	NF	OMIT	A	B	C	D	E
TOTAL	148	0	0	0	0.03	0.90	0.01	0.03	0.03
HIGH	38	0			0.05	0.84	0.05	0.05	0.00
MID	64	0			0.00	0.97	0.00	0.00	0.03
LOW	46	0			0.04	0.87	0.00	0.04	0.04
TEST SCORE MEAN %					68	68	79	69	67
DISCRIMINATING POWER (D.P.)					0.01	−0.03	0.05	0.01	−0.04
STANDARD ERROR OF D.P.					0.07	0.11	0.05	0.07	0.05

co-variation of a continuous variable (the total score on the test) with one that has a truly dichotomous distribution (in this case item 87). The point bi-serial is mathematically equivalent to a Pearson correlation coefficient in which one variable has only two values. The RPB for this item is almost zero. If the purpose of this test was to make discriminations between individual test takers, there would be a case for excluding this item, because it does not separate good from less good candidates. However, the focus of this item is to assess a core piece of knowledge that everyone should know – which is, in fact, what the item analysis shows the item does.

CRPB is the corrected point bi-serial. In this calculation the present item is removed from the test total, whilst the correlation is calculated. This is done because in some test situations the inclusion of the analysed item (item 87) in the test total can inflate this value and introduce additional co-variation. This problem is more troublesome when SAQs are being analysed. Typically each short answer has a weight greater than 1 mark and so variation in the item can make a large difference to the total score and hence to differences between the corrected and uncorrected versions of the statistic, with the uncorrected score being somewhat inflated (more positive). Even in the data for item 87, we can see considerable variation between the RPB and the CRPB.

RBIS and CRBIS are the 'non-point' versions of the previously discussed correlations. Strictly speaking the bi-serial coefficient is used where the dichotomous variable in fact has an underlying continuity and has been separated or recoded into a dichotomous one. The RBIS and CRBIS are always larger than their respective 'point' values, and may give an inflated value for variables where the continuity does not in fact exist.

IRI is the item's overall reliability index, which is virtually zero. However, whether it is retained in future tests depends upon the purpose of the test. As we discussed above, if the test is designed to be a test of competence in decisions around emergency medicine, then excluding all the items on which students scored highly might defeat the object of the test. If the test has some selective purpose – for example, identifying students who might be competing for an elective in emergency medicine in later years – then excluding item 87 might be appropriate.

In summary, the item is correctly answered by almost all the cohort, has little discriminating power as a predictor of total test score, appears to reflect the possibility that the previous CPR guideline (Option C) was being used by a few individuals, whilst the majority were aware of the new recommendation to use 2 : 30 in single bystander CPR.

Item analysis is a relatively quick way to identify problems with items, especially when such problems reside with their distractors. It is surprising how often items are identified as having distractors that are chosen by a significant proportion of students and then issues are discovered with the validity of the 'correct' option or with ambiguous wording of the question.

It is crucial that one or more of the group of academics/clinicians who designed the items understand how to interpret item analysis software, and that they appreciate the need to tailor decisions about items to the purpose of the test. Indiscriminate use of the item parameters as a criterion for exclusion is highly undesirable, even though this may appear more 'objective' it will introduce more problems than it solves.

In Box 21.12, we have another item in the test; item 46, and in Box 21.13, the item analysis data for that item. See if you agree with our interpretation.

In brief this shows that there is competition in the cohort's mind as to which is the correct answer with sport and alcohol being the two main contenders, with a minor one, hereditary. The item is not worded particularly well, as the

BOX 21.12 Example item for analysis (ii)

Item 46

Predisposing factors for osteoarthritis include all of the following, EXCEPT?

A. Sport

B. Alcohol abuse

C. Manual occupation

D. Obesity

E. Hereditary

* The correct answer is **B.**

BOX 21.13 Item analysis data for item 46

DIF = 0.50, RPB = −0.438, CRPB = 0.378, (95% CON = 0.164, 0.559) RBIS = 0.549, CRBIS = 0.474, IRI = 0.219

GROUP	N	INV	NF	OMIT	A	B*	C	D	E
TOTAL	148	0	0	0	0.23	0.50	0.07	0.08	0.12
HIGH	38	0			0.11	0.68	0.05	0.05	0.11
MID	64	0			0.12	0.63	0.06	0.09	0.09
LOW	46	0			0.48	0.17	0.09	0.09	0.17
TEST SCORE MEAN %					64	72	66	65	67
DISCRIMINAT ING POWER (D.P.)					−0.37	0.51	−0.03	−0.03	−0.07
STANDARD ERROR OF D.P.					0.14	0.15	0.08	0.08	0.11

options do not all follow grammatically from the stem. This may introduce some bias in the option choices. It is also double negatively worded. It might have been better to ask, 'which one of the following is *not* a predisposing factor for osteoarthritis?'

The difficulty is 0.50 – only half the cohort answered the item correctly. As it happens, items with difficulties within the 0.45–0.55 range are usually the most discriminating (this is a mathematical function), although of course they can be negatively discriminating (this is rare in well-chosen content, and is usually the result of an error in keying the correct answer). Here the discrimination is good: nearly half (0.48) of the third tertile chose 'A' as the correct answer, but roughly a tenth (0.11) of the top tertile did too. Perhaps this is because sport is featured so strongly in health promotional literature that some students believe it has few side-effects. Although the item is discriminating, and has good item-test correlations in both corrected and uncorrected versions, one might argue that in an emergency setting this knowledge might be less relevant; perhaps it would be much more appropriate in a musculoskeletal clinic or in general practice assessment. The overall item reliability is good – a test of 100 similar items would provide a highly discriminating test, though with a pass mark of around 50% it may not be defensible from a blueprinting perspective. If all such items were of 50% difficulty, it would be difficult to argue that the knowledge they contained had been treated as core by test takers.

Standard Setting for Written Assessment

In the previous section we noted that the purpose of the test is a fundamental consideration in how to interpret item-analysis data. We noted that our main example item was answered correctly by 90% of the cohort. So what should be an appropriate passing score on the test which this item comes from? The mean score on the test was 69% and the standard deviation was 5.7, the highest score was 91% and the lowest 52%, so even the one student scoring 3 SDs below the mean obtained more than 50%.

There has been much recent interest in trying to define standards for written tests, because 'standards are always arbitrary but should not be capricious', a quote often attributed to Ronald Berk, but which is borrowed from the USA legal system for appeals [102]. There seems nothing quite as arbitrary in a test of high-quality health professionals as a pass mark of 50% – is this doctor half-full or half-empty of the knowledge needed to perform in practice?

Consequently, effort has gone into describing what the appropriate level of performance on any test should be. Clearly, this might vary from test to test and year to year; as knowledge changes, items have different inherent difficulty, and expectations change. Standard setting is discussed in detail in Chapter 24, and there are some good, practically focused, treatments elsewhere [103]. However, there are some important principles as they relate to written assessment, as follows.

1 Production of all written high-stakes tests should entail effort to set 'a priori' standards. This is perfectly possible in both 'objective' and other types of written items such as SAQs.
2 Usually, these standards should be based on careful consideration of test content and its appropriateness to meet course/professional requirements, by a group of judges familiar with the cohort, the curriculum, the content, and the context. Research has shown that including a discussion phase on the standard-setting panel reduces the minimum number of judges required and can improve the reliability of test standards. The minimum number of judges to obtain a reasonable standard error of measurement on a test is 10 or more judges without discussion, or six or more judges after discussion. This is a significant efficiency on some previous estimates (10–15 judges required) [104].
3 Standard setting methods that best fit the purpose of a test should be chosen. Setting a minimum standard of competence may be a different purpose to selecting students to enter an advanced course, or certifying to a specialty. Giving feedback to students is a different purpose to accrediting them to lead the advanced life support team.
4 Norm referencing is unacceptable for written items used in competency licensing tests. These tests are focused on candidates' safety to practise, amongst other things. A clear and defensible standard needs to be identified.

5 Usually, for written assessments, the minimum acceptable standard should be decided before the test. In this setting (high-stakes examinations and patient safety concerns) variation in test, and hence item, difficulty is a critical issue. Standards should be set for each test, item by item, using item-focused methods such as Angoff, modified Angoff, or Ebel [105, 106]. See Chapter 22.
6 The final choice of method will be governed by the available resources and the consequences of misclassifying examinees as having passed or failed.
7 In general, setting standards for selected-response single-best-answer questions is less complex than setting standards for script concordance and, particularly, for key features items. This is because, especially in the latter, decisions in one part of a key feature may have impact on responses to other parts, and each may also be weighted quite differently. This problem is still being discussed [97].
8 The most robust methods are invariably also the most time consuming, but frequently result in excellent insights into test construction flaws because each item or response is carefully examined.

Summary

Written assessment is the most commonly used form of testing in tertiary and professional education. Huge amounts of research have been carried out on multiple-choice item formats and essay assessments. There are many variations of these methods used for assessment in the health professions. Newer formats, such as F-type and script concordance items, are still very much in development. Here we have presented the basic principles behind the most commonly used types of written assessment. Whilst there are many other topics that could be discussed in relation to written assessment, these are addressed by other chapters in this book.

Assessments of the types we have discussed here usually happen at the end of a term or year. However, they require a significant amount of time to prepare well, as each item needs to be written, reviewed, verified against the blueprint, and compiled into the whole test. Item flaws can have a significant impact on quality [107]. Then the test needs to be reviewed to check for cross-cueing or repetition between items, and optical score sheets prepared to reflect the different categories of responses in the test (e.g. for key features items) and then printed for the assessment. This can require between two and six months for a complex and/or comprehensive assessment.

Faculty development is a key component of improvements in assessment strategies. Some written assessments are deceptively simple, when given a cursory glance, but their quality is dependent on expertise in the discipline, as an educationalist, and in item writing. Good items also require detailed preparation and extensive trialling and analysis of items.

Finally, it is clear that there are different uses for different items. It is unlikely that a good assessment of cognitive skills in health disciplines can be undertaken with an MCQ. It will require a carefully chosen group of items, matched to the curriculum, using an array of different response modes to reflect the complexity of health care in the twenty-first century.

References

1 Dauphinee, D. (2002). Licensure and certification. In: *International Handbook of Research on Medical Education*, vol. 2 (ed. G. Norman, C. van der Vleuten and D. Newble), 835–882. Dordrecht: Kluwer.
2 Smith, J. (2004). The Shipman Inquiry – 5th Report. Safeguarding patients: lessons from the past – proposals for the future. HMSO. (http://webarchive.nationalarchives.gov.uk/20090808160144/http://www.the-shipman-inquiry.org.uk/fifthreport.asp (accessed 26 June 2018).
3 Walton, M., Shaw, T.J., Barnet, S., and Ross, J. (2006). Developing a national patient safety education framework for Australia. *Quality and Safety in Healthcare* 15: 437–442.
4 Scalese, R.J., Obeso, V.T., and Issenberg, S.B. (2008). Simulation technology for skills training and competency assessment in medical education. *Journal of General Internal Medicine* 23 (Suppl 1): 46–49.
5 Norcini, J.J. (2003). ABC of learning and teaching in medicine: work based assessment. *British Medical Journal (Clinical Research Ed.)* 326: 753–755.
6 Glaser, R. (1984). Education and thinking: the role of knowledge. *American Psychologist* 39: 93–104.
7 Boshuizen, H.P.A. and Schmidt, H.G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognitive Science* 16: 153–184.
8 Newble, D.J. (1991). The observed long case in clinical assessment. *Medical Education* 25: 369–373.
9 Hubbard, J.P. and Clemens, W.V. (1961). *Multiple-Choice Examinations in Medicine*. Philadelphia, PA: Lea and Febiger.
10 Newble, D.I. and Swanson, D.B. (1988). Psychometric characteristics of the objective structured clinical examination. *Medical Education* 22: 325–334.
11 Rust, C., O'Donovan, B., and Price, M. (2005). A social constructivist assessment process model: how the research literature shows us this could be best practice. *Assessment & Evaluation in Higher Education* 30 (3): 231–240.
12 Norman, G. (2002). Research in medical education: three decades of progress. *British Medical Journal (Clinical Research Ed.)* 324: 1560–1562.
13 Newble, D.I., Jolly, B.C., and Wakeford, R.E. ed. (1994). *Certification and Recertification in Medicine: Issues in the Assessment of Clinical Competence*. Cambridge: Cambridge University Press.
14 Norman, G., van der Vleuten, C., and Newble, D. ed. (2002). *International Handbook of Research on Medical Education*, vol. 2. Dordrecht: Kluwer.
15 Wood, R. (1991). *Assessment and Testing*. Cambridge: Cambridge University Press.
16 Newble, D.I. and Cannon, R. (2001). *A Handbook for Medical Teachers*, 4e. New York, NY: Springer.
17 Dauphinee, D. (1994). The content of certifying examinations. In: *Certification and Recertification in Medicine: Issues in the Assessment of Clinical Competence* (ed. D.I. Newble, B.C. Jolly and R.E. Wakeford), 92–104. Cambridge: Cambridge University Press.
18 Hamdy, H. (2006). Blueprinting for the assessment of health care professionals. *Clinical Teacher* 3: 175–179.
19 Kane, M.T. (1992). The assessment of professional competence. *Evaluation and the Health Professions* 15: 163–182.
20 Frank, J.R., Snell, L., and Sherbino, J. (2015). *The Draft CanMEDS 2015 Physician Competency Framework–Series IV*, 11–12. Ottawa: The Royal College of Physicians and Surgeons of Canada.

21 Downing, S.M. (2002). Threats to the validity of locally developed multiple-choice tests in medical education: construct irrelevant variance and construct under-representation. *Advances in Health Sciences Education* 7: 235–241.
22 Roberts, C., Sarangi, S., Southgate, L. et al. (2000). Oral examinations – equal opportunities, ethnicity and fairness in the MRCGP. *British Medical Journal (Clinical Research Ed.)* 320: 370–375.
23 Hobus, P.P., Schmidt, H.G., Boshuizen, H.P., and Patel, V.L. (1987). Contextual factors in the activation of first diagnostic hypotheses: expert-novice differences. *Medical Education* 21: 471–476.
24 Bowen, J.L. (2006). Educational strategies to promote clinical diagnostic reasoning. *New England Journal of Medicine* 355: 2217–2225.
25 Mylopoulos, M. and Regehr, G. (2007). Cognitive metaphors of expertise and knowledge: prospects and limitations for medical education. *Medical Education* 41: 1159–1165.
26 Croskerry, P. and Norman, G. (2008). Overconfidence in clinical decision making. *American Journal of Medicine* 121: 24–29.
27 Case, S. and Swanson, D. (1998). *Constructing Written Test Questions for the Basic and Clinical Sciences*, 3e. Philadelphia, PA: National Board of Medical Examiners http://www.nbme.org/publications/index.html#iwman (accessed 1 December 2012).
28 Epstein, R. (2007). Assessment in medical education. *New England Journal of Medicine* 356: 387–396.
29 Stalnaker, J.M. (1951). The essay type of examination. In: *Educational Measurement* (ed. E.F. Lindquist), 495–530. Menasha, WI: George Banta.
30 Ebel, R.L. and Frisbie, D.A. (1991). *Essentials of Educational Measurement*, 5e. Englewood Cliffs, NJ: Prentice-Hall.
31 Heckler, N.C., Rice, M., and Hobson Bryan, C. (2013). Turnitin systems: a deterrent to plagiarism in college classrooms. *Journal of Research on Technology in Education* 45 (3): 229–248.
32 Bloxham, S., den-Outer, B., Hudson, J., and Price, M. (2016). Let's stop the pretence of consistent marking: exploring the multiple limitations of assessment criteria. *Assessment & Evaluation in Higher Education* 41 (3): 466–481.
33 Schafer, W.D., Gagné, P., and Lissitz, R.W. (2005). Resistance to confounding style and content in scoring constructed-response items. *Educational Measurement: Issues and Practice* 24: 22–28.
34 Klein, S.P. and Hart, F.M. (1968). Chance and systematic factors affecting essay grades. *Journal of Educational Measurement* 5: 197–206.
35 Linn, R.L., Klein, S.P., and Hart, F.M. (1972). The nature and correlates of law school essay grades. *Educational and Psychological Measurement* 32: 267–279.
36 Markham, L.R. (1976). Influences of handwriting quality on teacher evaluation of written work. *American Educational Research Journal* 13: 277–283.
37 Powers, D., Fowles, M., Farnum, M., and Ramsey, P. (1994). Will they think less of my handwritten essay if others word process theirs? Effects on essay scores of intermingling handwritten and word-processed essays. *Journal of Educational Measurement* 31: 220–233.
38 Verma, M., Chatwal, J., and Singh, T. (1997). Reliability of essay type questions – effect of structuring. *Assessment in Education* 4: 265–270.
39 Grosse, M.E. and Wright, B.D. (1988). Psychometric characteristics of scores on a patient-management problem test. *Educational and Psychological Measurement* 48: 297–305.
40 Lim, E., Hian, C., Seet, H. et al. (2007). Computer-based testing of the modified essay question: the Singapore experience. *Medical Teacher* 9: e261–e268.
41 Feletti, G.I. (1980). Reliability and validity studies on modified essay questions. *Journal of Medical Education* 55: 933–941.
42 Palmer, E. and Devitt, P.G. (2007). Assessment of higher order cognitive skills in undergraduate education: modified essay or multiple choice questions? Research paper. *BMC Medical Education* 7: 49–53.
43 Palmer, E., Duggan, P., Devitt, P., and Russell, R. (2010). The modified essay question: its exit from the exit examination? *Medical Teacher* 32 (7): 300–307.
44 Lampe, L. and Ferguson, J. (2015). RANZCP 2015 Examination Information Session. Approaching the critical essay question and the modified essay questions in the RANZCP written examination. *Australian and New Zealand Journal of Psychiatry* 49: 5–6.
45 Jordan, S. and Mitchell, T. (2009). E-assessment for learning? The potential of short free-text questions with tailored feedback. *British Journal of Educational Technology* 40 (2): 371–385.
46 Smith, B., Sinclair, H., Simpson, J. et al. (2002). What is the role of double marking? Evidence from an undergraduate medical course. *Education for Primary Care* 13: 497–503.
47 Cannings, R., Hawthorne, K., Hood, K., and Houston, H. (2005). Putting double marking to the test: a framework to assess if it is worth the trouble. *Medical Education* 39: 299–308.
48 Edwards, B.D. and Arthur, W. Jr. (2007). An examination of factors contributing to a reduction in subgroup differences on a constructed-response paper-and-pencil test of scholastic achievement. *Journal of Applied Psychology* 92: 794–801.
49 Rodriquez, M.C. (2003). Construct equivalence of multiple-choice and constructed-response items: a random effects synthesis of correlations. *Journal of Educational Measurement* 40: 163–184.
50 Grant, D.L. (1957). Studies in the reliability of the short-answer essay examination. *The Journal of Educational Research* 51: 109.
51 Matters, G. and Paul, C. (2003). Psychological predictors of the propensity to omit short-response items on a high-stakes achievement test. *Educational and Psychological Measurement* 63: 239–256.
52 Rademakers, J., Ten Cate, T.H.J. and Bar, P.R. (2005). Progress testing with short answer questions. *Medical Teacher* 27: 578–82.
53 Gay, L.R. (1980). The comparative effects of multiple-choice versus short-answer tests on retention. *Journal of Educational Measurement* 17: 45–50.
54 Perera, G.R., Perera, D.N., and Weerasinghe, A.R. (2015). A dynamic semantic space modelling approach for short essay grading. *Advances in ICT for Emerging Regions (ICTer), 2015 Fifteenth International Conference*, pp. 43–49.
55 Roy, S., Dandapat, S., Nagesh, A., and Narahari, Y. (2016). Wisdom of Students: A Consistent Automatic Short Answer Grading Technique. *13th International Conference on Natural Language Processing*, p. 178.
56 Dillon, G.F., Boulet, J.R., Hawkins, R.E., and Swanson, D.B. (2004). Simulations in the United States Medical Licensing Examination™ (USMLE™). *Quality and Safety in Health Care* 13: 41–45.
57 Dillon, G.F., Clyman, S.G., Clauser, B.E., and Margolis, M.J. (2002). The introduction of computer-based case simulation into the United States Medical Licensing Examination. *Academic Medicine* 77: S94–S96.
58 Sterling, M., Leung, P., Wright, D., and Bishop, T.F. (2017). The use of social media in graduate medical education: a systematic review. *Academic Medicine* 92 (7): 1043–1056.
59 Crossley, J.G., Howe, A., Newble, D. et al. (2001). Sheffield Assessment Instrument for Letters (SAIL): performance assessment using outpatient letters. *Medical Education* 35 (12): 1115–1124.
60 Baker, E.A., Ledford, C.H., Fogg, L. et al. (2015). The IDEA Assessment Tool: assessing the reporting, diagnostic reasoning, and decision-making skills demonstrated in medical students' hospital admission notes. *Teaching and Learning in Medicine* 27 (2): 163–173.
61 Reinert, A., Berlin, A., Swan-Sein, A. et al. (2014). Validity and reliability of a novel written examination to assess knowledge and clinical decision making skills of medical students on the surgery clerkship. *The American Journal of Surgery* 207 (2): 236–242.
62 Downing, S.M. (2003). Guessing on selected-response examinations. *Medical Education* 37: 670–671.

63 Downing, S.M. (2002). Assessment of knowledge in written test forms. In: *International Handbook of Research on Medical Education*, vol. 2 (ed. G. Norman, C. van der Vleuten and D. Newble), 647–672. Dordrecht: Kluwer.
64 Haladyna, T.M., Downing, S.M., and Rodriquez, M.C. (2002). A review of multiple-choice item writing guidelines for classroom assessment. *Applied Measurement in Education* 15: 309–334.
65 Haladyna, T.M. and Downing, S.M. (1993). How many options is enough for a multiple-choice test item? *Educational and Psychological Measurement* 53: 999–1009.
66 Case, S.M., Swanson, D.B., and Ripkey, D.R. (1994). Comparison of items in five-option and extended-matching format for assessment of diagnostic skills. *Academic Medicine* 69 (Suppl): S1–S3.
67 Skakun, E., Maguire, T., and Cook, D. (1994). Strategy choices in multiple choice items. *Academic Medicine* 69 (Suppl): S7–S9.
68 Case, S.M. (1983). The development and evaluation of a new instrument to assess medical problem solving. *Dissertation Abstracts International* 44: 1764.
69 Bhakta, B., Tennant, A., Horton, M. et al. (2005). Using item response theory to explore the psychometric properties of extended matching questions examination in undergraduate medical education. *BMC Medical Education* 5: 9.
70 Beullens, J., Struyf, E. and Van Damme, B. (2006). Diagnostic ability in relation to clinical seminars and extended-matching questions examinations. *Medical Education* 40: 1173–9.
71 Beullens, J., Struyf, E. and Van Damme, B. (2005). Do extended matching multiple-choice questions measure clinical reasoning? *Medical Education*. 39: 410–7.
72 Fenderson, B.A., Damjanov, I., Robeson, M.R. et al. (1997). The virtues of extended matching and uncued tests as alternatives to multiple choice questions. *Human Pathology* 28: 526–532.
73 McCoubrie, P. (2004). Improving the fairness of multiple-choice questions: a literature review. *Medical Teacher* 8: 709–712.
74 George, S. (2003). Extended matching items (EMIs): solving the conundrum. *Psychiatric Bulletin of the Royal College of Psychiatrists* 27: 230–232.
75 Zimmerman, D.W. and Williams, R.H.A. (2003). New look at the influence of guessing on the reliability of multiple-choice tests. *Applied Psychological Measurement* 27: 357–371.
76 Swanson, D.W., Holtzman, K.Z., Clauser, B.E., and Sawhill, A.J. (2005). Psychometric characteristics and response times for one-best-answer questions in relation to number and source of options. *Academic Medicine* 80 (Suppl): S93–S96.
77 Swanson, D.B., Holtzman, K.Z., and Allbee, K. (2008). Measurement characteristics of content-parallel single-best-answer and extended-matching questions in relation to number and source of options. *Academic Medicine* 83 (Suppl): S21–S24.
78 Charlin, B., Brailovsky, C., Leduc, C., and Blouin, D. (1998). The diagnosis script questionnaire: a new tool to assess a specific dimension of clinical competence. *Advances in Health Sciences Education* 3: 51–58.
79 Charlin, B., Brailovsky, C., Roy, L. et al. (2000). The script concordance test: a tool to assess the reflective clinician. *Teaching and Learning in Medicine* 12 (4): 189–195.
80 Schmidt, H., Norman, G., and Boshuizen, H. (1990). A cognitive perspective on medical expertise: theory and implications. *Academic Medicine* 65: 611–621.
81 Gagnon, R., Charlin, B., Roy, L. et al. (2006). The cognitive validity of the script concordance test: a processing time study. *Teaching and Learning in Medicine* 18: 22–27.
82 Dory, V., Gagnon, R., Vanpee, D., and Charlin, B. (2012). How to construct and implement script concordance tests: insights from a systematic review. *Medical Education* 46: 552–563.
83 Charlin, B. and van der Vleuten, C. (2004). Standardised assessment of clinical reasoning in contexts of uncertainty: the script concordance approach. *Evaluation and the Health Professions* 27: 304–319.
84 Sibert, L., Darmoni, S.J., Dahamna, B. et al. (2006). On line clinical reasoning assessment with script concordance test in urology: results of a French pilot study. *BMC Medical Education* 6: 45–47.
85 Meterissian, S., Zabolotny, B., Gagnon, R., and Charlin, B. (2007). Is the script concordance test a valid instrument for assessment of intra-operative decision-making skills? *American Journal of Surgery* 193: 248–251.
86 Kelly, W., Durning, S., and Denton, G. (2012). Comparing a script concordance examination to a multiple-choice examination on a core internal medicine clerkship. *Teaching and Learning in Medicine* 24: 187–193.
87 Goulet, F., Jacques, A., Gagnon, R. et al. (2010). Poorly performing physicians: does the script concordance test detect bad clinical reasoning? *Journal of Continuing Education in the Health Professions* 30: 161–166.
88 Lineberry, M., Kreiter, C.D., and Bordage, G. (2013). Threats to validity in the use and interpretation of script concordance test scores. *Medical Education* 47: 1175–1183.
89 Bland, A.C., Kreiter, C.D., and Gordon, J.A. (2005). The psychometric properties of five scoring methods applied to the script concordance test. *Academic Medicine* 80: 395–399.
90 Wilson, A.B., Pike, G.R., and Humbert, A.J. (2014). Analyzing script concordance test scoring methods and items by difficulty and type. *Teaching and Learning in Medicine* 26: 135–145.
91 Goos, M., Schubach, F., Seifert, G., and Boeker, M. (2016). Validation of undergraduate medical student script concordance test (SCT) scores on the clinical assessment of the acute abdomen. *BMC Surgery* 16 (1): 57–68.
92 Kreiter, C.D. (2017). A Bayesian perspective on constructing a written assessment of probabilistic clinical reasoning in experienced clinicians. *Journal of Evaluation in Clinical Practice* 23 (1): 44–48.
93 Farmer, E.A. and Page, G. (2005). A practical guide to assessing clinical decision-making skills using the key features approach. *Medical Education* 39: 1188–1194.
94 Norman, G., Bordage, G., Curry, L. et al. (1985). Review of recent innovations in assessment. In: *Directions in Clinical Assessment. Report of the Cambridge Conference on the Assessment of Clinical Competence* (ed. R. Wakeford), 8–27. Cambridge: Office of the Regius Professor of Physic, Cambridge University School of Clinical Medicine, Addenbrooke's Hospital.
95 McGuire, C.H., Solomon, L.M., and Bashook, P.G. (1976). *Construction and Use of Written Simulations*. New York, NY: Psychological Corporation of Harcourt, Brace, Jovanovich.
96 Page, G. and Bordage, G. (1995). The Medical Council of Canada's key features project: a more valid written examination of clinical decision-making skills. *Academic Medicine* 70: 104–110.
97 Fischer, M., Kopp, V., Holzer, M. et al. (2005). A modified electronic key feature examination for undergraduate medical students: validation threats and opportunities. *Medical Teacher* 27: 450–455.
98 Trudel, J.L., Bordage, G., and Downing, S.M. (2008). Reliability and validity of key feature cases for the self-assessment of colon and rectal surgeons. *Annals of Surgery* 248: 252–258.
99 Farmer, E.A. and Hinchy, J. (2005). Assessing clinical decision making skills in general practice: the key features approach. *Australian Family Physician* 34: 1059–1061.
100 Lawrence, K., Allen, T., Brailovsky, C. et al. (2011). Defining competency-based evaluation objectives in family medicine key feature approach. *Canadian Family Physician* 57: E373–E380.
101 Baldwin, P., Baldwin, S.G., and Haist, S.A. (2011). F-type testlets and the effects of feedback and case-specificity. *Academic Medicine* 86: S56–S59.
102 Cizek, G.J. ed. (2001). *Setting Performance Standards. Concepts, Methods and Perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates.

103 Norcini, J.J. (2003). Setting standards on educational tests. *Medical Education* 37: 464–469.

104 Fowell, S.L., Fewtrell, R., and Mclaughlin, P.J. (2008). Estimating the minimum number of judges required for test-centred standard setting on written assessments. Do discussion and iteration have an influence? *Advances in Health Sciences Education* 13: 11–24.

105 Cizek, G.J. (1996). Standard setting guidelines. *Educational Measurement: Issues and Practice* 15: 13–21.

106 Ebel, R.L. (1972). *Essentials of Educational Measurement*, 2e. Englewood Cliffs, NJ: Prentice-Hall.

107 Rush, B.R., Rankin, D.C., and White, B.J. (2016). The impact of item-writing flaws and item complexity on examination item difficulty and discrimination value. *BMC Medical Education* 16 (1): 250.

22 Workplace Assessment

John J. Norcini[1] *and Zareen Zaidi*[2]

[1]Foundation for Advancement of International Medical Education and Research (FAIMER), Philadelphia, PA, USA
[2]Department of Medicine, University of Florida College of Medicine, Gainesville, FL, USA

KEY MESSAGES

- Assessment in the setting of clinical training is not well developed, but methods based on observation of routine encounters offer a rich and feasible alternative.
- A framework for assessment based on observation has two dimensions: the grounds on which judgements are made (single encounter or routine performance) and the nature of the judgements themselves (occurrence, quality, or suitability).
- Tools such as the mini-CEX, DOPS, CbD, and mini-PAT offer valid assessment.
- Faculty development is a key to the successful use of the methods.
- The opportunity for educational feedback as part of these methods is as important as their contribution to the assessment process.

Introduction

Assessment in the context of medical education has changed dramatically over the past 50 years [1]. From near exclusive reliance on the essay question and the clinical viva, assessment methods have proliferated, particularly those appropriate to use in the setting of undergraduate medical education. These newer methods cover more of the competencies of a doctor, and there is a much better understanding of how to deploy them to produce scores that are valid and reliable.

Assessment in the setting of clinical training, particularly postgraduate training, is not as well developed as that in the undergraduate arena. The curriculum in most clinical training settings is less structured, and the trainees often have more responsibility. Consequently, assessment needs to pose a broader range of patient problems and include more complex and acute care, multi-system disease, and procedural skill. Moreover, the focus is more often on the assessment of integrated skills rather than on specific aspects of competence. Compared with practising doctors, trainees in clinical settings are not yet completely responsible for patients, and they have not differentiated within specialty. As a result, assessment needs to focus on the potential to practise, not actual practice, so assessments of work products are likely to be less useful [2]. Moreover, the results of an assessment programme need to support the educational enterprise.

It is difficult to develop high-quality written or performance-based assessments locally, especially in the context of postgraduate training. There are relatively small numbers of trainees and staff, resources are fragmented across the specialties, and assessment expertise is rare. There is also a need to address sophisticated content and skills, which are difficult to simulate at the level of advanced trainees.

Despite these challenges, there are two aspects of training in a clinical setting that offer significant advantages for assessment. First, there are routine interactions among members of the health care team and between trainees and patients, so the clinical material that can serve as the basis for assessment is readily available. Second, there are skilled clinician-educators in the setting who, with some training, can act as judges. Consequently, assessment methods that are based on observation of routine encounters are most feasible in the setting of clinical training. In addition, the use of these methods supports the educational process because they offer the opportunity for formative feedback and the development of a plan for remediation when it is needed.

In 1990, Miller proposed a structure, in the form of a pyramid, for categorising methods of assessment [3]. Knowledge ('knows') is at the lowest level of the pyramid followed by competence ('knows how'), performance ('shows how'), and action ('does') (see Figure 22.1). Assessments based on observation in the setting of work are representative of the top two levels of the pyramid. Miller distinguished between these two levels depending on whether trainees were in an artificial testing situation (i.e. they were aware that they were being assessed). Underlying this distinction is the reasonable but unproven assumption that the assessment methods that come closest to capturing a doctor's unobserved, routine functioning

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

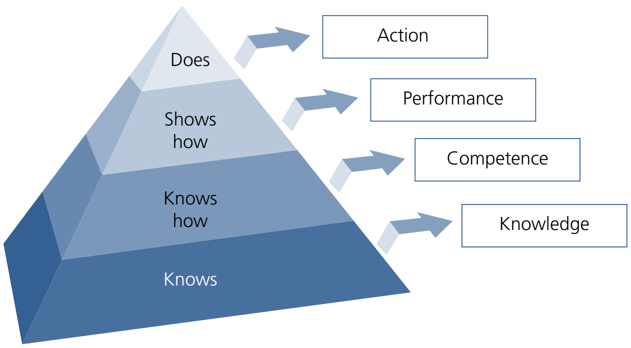


Figure 22.1 Miller's pyramid for assessing clinical competence.

will yield the most valid results. Of course, capturing authentic performance has problems as well since the clinical context poses a number of threats to validity (e.g. differences between doctors in case mix and the severity of illness of their patients) [2].

A Framework for Assessment based on Observation

In the work context, assessors are asked to make a number of different kinds of judgement based on their observations. For the purpose of this chapter, a two-dimensional framework will be used to describe these judgements [4, 5]. The first dimension relates to the grounds on which the judgements are made – a single encounter or routine performance. The second dimension relates to the nature of the judgement – whether it is occurrence, quality, or suitability.

Grounds for Judgement

Single Encounter

In this instance, the assessors base their judgements on observation of a single event. For example, a faculty member might observe the trainee in interaction with a particular patient, patient record, or procedure, and then provide an evaluation of it. The traditional clinical viva illustrates this. A trainee examines and interviews a patient, draws conclusions, and then presents all of this to one or more assessors. Judgements about the trainee are made on the basis of that one event.

The advantage of basing judgements on single events is the reassurance that the assessor has actually observed the performance and is focused on reaching a conclusion about it. Theoretically, this reduces the biasing effects of previous contact and clarifies what is to be evaluated. The disadvantage is that the performance of doctors is case or task specific [6]. This means that performance on one event does not predict with high accuracy performance on another. Therefore, several different events need to be sampled to obtain a generalisable estimate of performance (Box 22.1).

Routine Performance

In this instance, the assessors base their judgements on observations they have made over a period of time. This is one of the most common types of assessment, and most training programmes in the US and UK ask faculty members to periodically complete rating forms that attest to the competence of their trainees.

The major advantage of this basis for judgement is that it should include observations of the trainee on a number of different occasions. In this way, it reduces, to some degree, the problems of case specificity of performance. However, assessors sometimes offer evaluations of aspects of performance they do not observe. For example, Pulito et al. found that faculty members primarily observe cognitive skills and professionalism and have little basis for assessing other aspects of competence [7]. Moreover, Silber et al. found that faculty members tend to assess competence along the two dimensions of medical knowledge and interpersonal skills, and do not make distinctions among other aspects of competence [8].

BOX 22.1 FOCUS ON: How many encounters are needed?

In most assessment programmes, the same number of encounters is required of all trainees.

For example, the UK Foundation Programme calls for each trainee to undertake a minimum of nine direct observations of practice each year – of which six must be the mini clinical evaluation exercise (mini CEX) – and six case-based discussions. These estimates balance traditional estimates of reliability against feasibility. Depending on the purpose of the assessment, however, this may be too few or too many encounters for certain trainees. Use of the standard error of measurement (SEM) permits the application of a more refined strategy for making this decision.

The SEM is an alternative to traditional measures of reliability, and it can be used to construct a 95% confidence interval around scores [9]. For example, data from a study of the mini-CEX indicate that the 95% confidence interval for the overall rating of clinical competence is ±1.2 after two encounters and ± 0.8 after four encounters, and it continues to decrease more slowly with additional encounters [10]. These data are based on a nine-point scale where 1–3 is unsatisfactory, 4–6 is satisfactory, and 7–9 is superior. Thus, we can be 95% confident that the true score of a trainee student with an average rating of 4.0 on two encounters lies between 2.8 and 5.2.

If the purpose of assessment is simply to identify which trainees are unsatisfactory, two encounters are probably sufficient for many trainees, certainly those with average ratings of 6 or better. On the other hand, for trainees with averages between 2.8 and 5.2, additional encounters are needed. As encounters are added, the width of the confidence interval shrinks and the number of good decisions increases. Thus, limited assessment resources can be focused where they do the most good.

The assessment advantages of this strategy go together with significant educational advantages. Borderline trainees will have more encounters, and since each is accompanied by feedback, those who need it most will receive more intense educational interventions.

Nature of the Judgement

Based on their observations of single encounters or routine performance, the nature of judgements assessors are asked to make falls into three categories:
- occurrence – whether particular behaviours were demonstrated
- quality – the 'goodness' of the performance
- fitness or suitability – whether the performance was good enough for a particular purpose.

Occurrence

Assessors are sometimes asked to indicate whether they have observed a particular behaviour, and they are often given a checklist on which to note the occurrence. For instance, Martin et al. developed procedure-specific checklists containing 22–32 task steps [11]. They were applied by assessors to their observation of procedures done by trainees. Each time one of the steps on the checklist was completed, the assessor noted it, and after the procedure, the marks were tallied.

Simply noting the occurrence of behaviours is often viewed as objective, leaving less to the judgement of the observer. It structures the task of the assessor and ensures that it is more focused and consistent over observations and assessors. If they are clinician-educators, however, simply asking the assessor to note the occurrence of particular behaviours does not make best use of their ability to discriminate among performances.

Quality

More often, assessors are asked to make a judgement about the quality of the performance they observe. They typically record these judgements on a rating scale. For example, in the study by Martin et al., observers were also asked to complete a seven-item global rating form that captured the quality of the trainees' procedural skills on a five-point scale [11]. The same form was used for all procedures, and after the examination the ratings were tallied.

In the Martin et al. study, the global assessments of quality were strongly correlated with the checklists, which simply noted the occurrence of aspects of performance [11]. This is not an unusual finding, and it suggests that they are capturing the same aspects of competence [12]. However, the global ratings tend to be a bit more valid. For instance, in the Martin et al. study, they discriminated among levels of training, while the checklists did not.

Fitness

In some instances, assessors are asked to determine whether a performance is satisfactory or fit for purpose. For example, in the Martin et al. study, the assessors were also asked to make a pass/fail judgement [11]. In essence, this required them to make two judgements in sequence. They needed to first establish the quality of a performance, and then they had to decide whether it was good enough for a particular purpose. In the Martin et al. study, pass/fail results were not correlated with other measures (e.g. year of training) because so many of the participants passed [11].

The simultaneous decision-making underlying a single judgement about fitness is very efficient in academic settings where it is important to identify individuals for advancement and remediation. However, it combines two somewhat unreliable judgements into one and renders the meaning of the results unclear to a degree. For example, a failing judgement can be rendered because the performance was poor or the assessor had high standards. Since it is not possible to disentangle these, it is best to ask for judgements about quality and fitness separately, as they did in the study by Martin et al. [11].

Common Methods

A number of different methods of assessment based on observation, each known by an increasing bewildering array of acronyms, have been used in the setting of clinical training. The methods cluster into one of three common varieties:
- direct observation of practice (e.g. mini-CEX, DOPS, OCAT)
- chart-stimulated recall (e.g. CbD)
- multi-source or 360° feedback (MSF) [13].

Characteristically, such assessments are gathered into a portfolio that trainees complete and this collection forms the basis for making judgements about educational progress. The methods listed above will form the focus of the following section.

Mini-CEX

In the mini-clinical evaluation exercise (mini-CEX) a faculty member observes a trainee interact with a patient in a clinical setting [13, 14]. The trainee engages in a clinical activity (e.g. taking a focused history and performing relevant aspects of the physical examination) and afterwards summarises the encounter (e.g. provides a diagnosis and/or treatment plan). The faculty member scores the performance and then provides educational feedback. The encounters are intended to take about 15 minutes, and trainees are expected to be evaluated several times and by different faculty members.

The method was originally devised for use in internal medicine postgraduate training programmes in the US. Individual faculty members were responsible for deciding when trainees were to be assessed and for identifying appropriate patients. Ratings were gathered on a 9-point scale, where 1–3 was unsatisfactory, 4–6 was satisfactory, and 7–9 was superior. The dimensions of performance observed and evaluated were interviewing skill, physical examination, professionalism, clinical judgement, counselling, organisation and efficiency, and overall competence. Not every encounter permitted assessment of all of these dimensions. Depending on the purpose of the assessment, the ratings were aggregated across dimensions and encounters for each trainee [13, 14].

Given its structure, the ground for judgements underpinning the mini-CEX is always the single encounter. Depending on how it has been deployed, however, the nature of the judgements has varied depending on the purpose of the assessment. For instance, in the US, assessors were asked to judge both the quality (ranking from 1 to 9) of the performance and its fitness (unsatisfactory versus satisfactory/superior) for a first-year postgraduate trainee.

In the US, the mini-CEX has been used in a number of different inpatient, outpatient, and emergency department settings. In these settings, the mini-CEX has been applied to a broad range of patient problems. For example, in a study by Norcini et al., the presenting complaints included abdominal pain, chest pain, cough, dizziness, fever, headache, low back pain, shortness of breath, and weight gain. Common internal medicine problems, such as arthritis, asthma, chronic obstructive pulmonary disorder, congestive heart failure (CHF), coronary artery disease, diabetes, and hypertension, formed the basis for assessment, as well as other common problems, such as seizure, substance abuse, depression, dementia, and rash. The mini-CEX was also applied to trainees assessing patients with multiple problems, such as CHF, hypertension, and diabetes, and with acute problems such as sepsis and myocardial infarction [10].

The mini-CEX is analogous to a classroom test for the clinical setting. It is intended to identify the few trainees whose performance is wholly unsatisfactory and to provide documentation of their shortcomings. This documentation serves as the evidence in support of a later educational decision about the trainee. More importantly however, for the vast majority of trainees it provides an opportunity for ongoing formative assessment and feedback. It is also designed to ensure that the clinical skills of trainees have been observed and evaluated by faculty members. Unfortunately, observation and feedback occur far too rarely in the context of many busy clinical placements [15].

The mini-CEX is not intended for use in a high-stakes examination setting, nor should it be used to compare or rank trainees across different programmes.

Although more work remains, a recent systematic review finds a number of studies that provide evidence of the validity of the mini-CEX [16]. For example, in the undergraduate setting, Kogan et al. found that mini-CEX scores had modest correlations with examination scores, inpatient clerkship ratings, outpatient clerkship ratings, and final course grades [17]. In a postgraduate setting, Durning et al. found correlations between the individual components of the mini-CEX and the corresponding monthly evaluations by faculty members, as well as the results of an in-training examination [18]. In a study by Boulet et al., videotapes of standardised patient (SP)–student encounters were evaluated by faculty using the mini-CEX form [19]. The SP checklists predicted faculty global ratings, and SP ratings of doctor–patient communication correlated with faculty ratings of communications. Finally, Holmboe et al. scripted videotapes of trainees whose performance was unsatisfactory, satisfactory, or superior. Using the mini-CEX form, faculty successfully discriminated among the three levels of performance [20].

Since its initial development, the mini-CEX has been modified for use in a number of different settings and the original forms have been translated into several different languages. The competencies assessed have been tailored to particular needs (e.g. the Professionalism Mini-Evaluation Exercise) and to a variety of undergraduate and postgraduate disciplines and clinical settings [21–23]. The rating scales have been modified as well, with changes in the number of points on the scale, the definitions of those points, and, in some instances, the replacement of the ratings with written comments. All of these modifications are appropriate, and they ensure the relevance of the method to the setting in which it is being used.

The assessment programme of the UK Foundation Programme provides an example in which the mini-CEX has been appropriately modified to make it more relevant to a particular setting [13]. Trainees are assessed several times throughout the year with different faculty members for each encounter. Both the assessor and the patient are selected by the trainee, but the assessor must agree that the encounter is appropriate. Assessors include consultants, experienced specialist registrars, staff grade and associate specialists, and general practitioners. There is a list of core problems as part of the curriculum, and trainees are expected to sample from them. The immediate feedback given after the encounter includes strengths, weaknesses, and an action plan for further effort. Initially, ratings of performance were gathered but more recently these have been replaced by the provision of text comments.

Figure 22.2 shows a typical assessment form. The descriptors require the assessors to judge both the quality of the trainee's performance and whether it meets expectations for completion of the year of training.

DOPS

DOPS is a variation on the mini-CEX, which was originally designed by the Royal College of Physicians to assess and provide feedback on procedural skills [24]. Just as with the mini-CEX, trainees are observed with real patients, but in DOPS they are conducting procedures. After the encounter, the faculty member rates the trainee's performance and provides educational feedback. The encounters are necessarily brief (usually less than 15 minutes, with 5 minutes for feedback), and trainees are expected to be evaluated several times and by different faculty members.

An example of DOPS is its use in the UK Foundation Programme. Trainees must select from an approved list that contains many of the procedures used routinely in practice [13]. For example, it includes various injections, intubation, electrocardiogram, nasogastric tube insertion, venepuncture, cannulation, and arterial blood sampling. Trainees are assessed on their understanding of indications, anatomy, technique, aseptic technique, proper analgesia, communication, and other important aspects of procedural skill. They are also asked how often they have performed the procedure. More recently, the ratings have been removed and the trainee is offered free text assessment and feedback.

Trainees are to be assessed several times throughout the year with different procedures and faculty members for each encounter [13]. As with all the tools used as part of this programme, the timing, procedure, and assessor are selected by the trainees, but the assessor must agree that the procedure is appropriate. Like the mini-CEX, assessors include consultants, experienced specialist registrars, staff

Please refer to www.hcat.nhs.uk for guidance on this form and details of expected competencies for F1

Mini-Clinical Evaluation Exercise (CEX) – F1 Version

Please complete the question using a cross: ☒ Please use black ink and CAPITAL LETTERS

Doctor's Surname

Forename

GMC Number: **GMC NUMBER MUST BE COMPLETED**

Clinical setting: A&E ☐ OPD ☐ In-patient ☐ Acute Admission ☐ GP Surgery ☐

Clinical problem category: Airway/ Breathing ☐ CVS/ Circulation ☐ Gastro ☐ Neuro ☐ Pain ☐ Psych/ Behav ☐ Other

New or FU: New ☐ FU ☐ | Focus of clinical encounter: History ☐ Diagnosis ☐ Management ☐ Explanation ☐

Number of times patient seen before by trainee: 0 ☐ 1–4 ☐ 5–9 ☐ >10 ☐ | Complexity of case: Low ☐ Average ☐ High ☐

Assessor's position: Consultant ☐ GP ☐ SpR ☐ SASG ☐ SHO ☐ Other

Number of previous mini-CEXs observed by assessor with any trainee: 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5–9 ☐ >9 ☐

Please grade the following areas using the scale below:	Below expectations for F1 completion		Borderline for F1 completion	Meets expectations for F1 completion	Above expectations for F1 completion		U/C*
1. History Taking	☐	☐	☐	☐	☐	☐	☐
2. Physical Examination Skills	☐	☐	☐	☐	☐	☐	☐
3. Communication Skills	☐	☐	☐	☐	☐	☐	☐
4. Clinical Judgement	☐	☐	☐	☐	☐	☐	☐
5. Professionalism	☐	☐	☐	☐	☐	☐	☐
6. Organisation/Efficiency	☐	☐	☐	☐	☐	☐	☐
7. Overall clinical care	☐	☐	☐	☐	☐	☐	☐

*U/C Please mark this if you have not observed the behaviour and therefore feel unable to comment.

Anything especially good? | **Suggestions for development**

Agreed action:

Have you had training in the use of this assessment tool?: ☐ Face-to-Face ☐ Have Read Guidelines ☐ Web/CDrom

Assessor's Signature:

Date (mm/yy): M M / Y Y

Time taken for observation: (in minutes)

Time taken for feedback: (in minutes)

Assessor's Surname

Assessor's registration number:

Please note: Failure of return of all completed forms to your administrator is a probity issue

Acknowledgements: Adapted with permission from American Board of Internal Medicine

Figure 22.2 Typical mini-CEX assessment form.

grade and associate specialists, and general practitioners. However, nurses and other appropriate allied health professionals can also act as assessors. The immediate feedback given after the encounter includes areas of strengths and suggestions for development.

Since it is a variation, studies indicating the validity of the mini-CEX apply to DOPS as well. In addition, there is considerable research showing that global ratings of procedural skills can produce valid results [25]. For example, Goff et al. demonstrated that in an objective structured assessment of technical skills, judgements of occurrence, quality, and fitness by two assessors were able to distinguish among levels of training [26]. Similarly, a study of actual performance by Winckel et al. found that both checklists and global ratings distinguished among levels of training for 41 operations (see Box 22.2) [27]. More recently, Marriott et al. found good validity, reliability, and acceptability when using DOPS to assess trainees' skills in the operating theatre [28].

OCAT

The Ottawa Clinic Assessment Tool (OCAT) is also a variation on the mini-CEX, which was designed by The University of Ottawa, Canada [29]. OCAT is an example of a growing number of tools designed to observe a composite set of clinical activities over a defined timeframe. It is used for the assessment of a surgical trainee in an everyday clinic environment, specifically to assess their ability to safely and independently run a surgical clinic at the level of a generalist. It uses a modified 'entrustability scale' [30]. As described in the section below, Entrustable Professional Activities (EPAs) are descriptors of work, which often require proficiency in several different competencies [31]. Entrustability scales are designed to reflect the way clinical educators think in the workplace. For example, does a surgical educator feel comfortable leaving a resident alone with a patient during a procedure? [32–34]. The narrative entrustment scale used in OCAT has the following anchors which revolve around the feelings of the assessor:

1 I had to do.
2 I had to talk them through.
3 I had to direct them from time to time.
4 I needed to be available just in case.
5 I did not need to be there.

The OCAT is a nine-item tool, with a global assessment item and two short-answer questions regarding technical skills. At the end of a clinic, trainees request their supervising surgeon to fill out an OCAT form (Figure 22.3). Initial evidence demonstrates that at least three OCAT forms per trainee are required to produce a generalisability coefficient of 0.88 [29].

BOX 22.2 WHERE'S THE EVIDENCE: Checklists versus global ratings

There is a sizeable body of research on the use of checklists, which capture the occurrence of particular behaviours, and global rating scales, which capture the quality of a performance [35–40].

Scores based on checklists are strongly correlated with scores based on global rating scales.

Checklists are: (i) perceived to be more objective and (ii) can produce slightly more reliable scores, but (iii) they may not be as good at capturing advanced levels of expertise.

Global ratings are: (i) perceived to be more subjective, but (ii) tend to be slightly more valid.

Doctors or trained patients can use checklists since they require only that behaviour to be noted. Only experts can use global ratings. Overall differences between checklists and global ratings are relatively small.

CSR

Chart-stimulated recall (CSR) was developed by Maatsch for use by the American Board of Emergency Medicine [41]. A variation of it, called the Case-based Discussion, has been used in UK postgraduate training, including in the Foundation Programme. In this setting, the trainee must select two case records from patients they have seen recently and in which they have made entries [13]. The assessor selects one of the two cases and explores one aspect of it with the trainee. For example, they might choose to focus on which investigations the trainee ordered or on the ethical issues raised by a particular patient. In all instances, the assessor is interested in understanding the reasoning behind the trainee's choices.

CSR is designed to offer an assessment of medical record keeping and to stimulate trainees to discuss why they acted as they did. In this way it offers the opportunity for assessment of the application of knowledge, decision-making, and ethical issues. CSR uses single encounters as the grounds for measurement, and assessors are asked to make judgements about both the quality of the clinical assessment, investigation and referrals, treatment, professionalism, medical record keeping, and overall clinical care, and about whether they meet or exceed expectations (i.e. fitness). An assessment form for the Case-based Discussion version of CSR is shown in Figure 22.4.

When used in the Foundation Programme, each encounter is intended to take 15 minutes, followed by 5 minutes of feedback, and there should be four to six encounters during the year [13]. Assessors include consultants, experienced specialist registrars, staff grade and associate specialists, and general practitioners. There is a list of core problems as part of the curriculum, and trainees are expected to sample from them. The feedback given after the encounter should include strengths, suggestions for development, and an action plan that outlines a response to these suggestions.

The original work on the validity of the CSR was done in conjunction with the certification and recertification programmes of the American Board of Emergency Medicine [41]. When given to a sample of practising doctors, CSR score distributions and pass–fail results were consistent with those for initial certification. They were also correlated with scores from a variety of other methods, including an oral examination and an audit of practice records. Of all

The Ottawa Clinic Assessment Tool (OCAT)

Study ID:	Level: 1 2 3 4 5	Staff Name:
Clinic:		Date:

The purpose of this scale is to assess the trainee's ability to safely and independently run a CLINIC IN YOUR SPECIALTY (i.e. Urology, General Surgery, etc.) at the level of a GENERALIST (i.e. certified graduate of residency program).

With that in mind please use the scale below to rate each item, irrespective of the resident's level of training. Base your rating on the trainee's performance across the ENTIRE CLINIC (i.e. do not base your rating on only one specific patient encounter).

Please complete the assessment *IMMEDIATELY* following completion of the clinic.

Scale

1—**"I had to do"—**i.e., Requires complete guidance, unprepared to do, or had to do for them
2—"**I had to talk them through"—**i.e., Able to perform some tasks but requires repeated directions
3—"**I had to direct them from time to time**"—i.e., Demonstrates some independence, but requires intermittent prompting
4—**"I needed to be available just in case"—**i.e., Independence but needs assistance with nuances of certain patients and/or situations, unable to manage all patients, still requires supervision for safe practice
5—**"I did not need to be there"—**i.e., Complete independence, can safely manage a general clinic in your specialty

1. History Efficient data gathering	**1**	**2**	**3**	**4**	**5**
2. Physical Exam Efficient and accurate examination	**1**	**2**	**3**	**4**	**5**
3. Case Presentation Synthesis of history and physical, clear presentation	**1**	**2**	**3**	**4**	**5**
4. Differential Diagnosis Able to make a diagnosis and appropriately consider alternatives	**1**	**2**	**3**	**4**	**5**
5. Management Plan Able to develop relevant plan dependent on context and be decisive (i.e. appropriate investigations, procedures, etc)	**1**	**2**	**3**	**4**	**5**
6. Patient/Family Communication Effective, sensitive, and respectful communication skills (verbal & non-verbal), language appropriate to patient understanding, able to build rapport and trust	**1**	**2**	**3**	**4**	**5**
7. Documentation within Clinic Orders, prescriptions, forms, etc (may not include consultation report)	**1**	**2**	**3**	**4**	**5**
8. Collaboration Works well with and/or teaches other team members as appropriate (i.e. staff, student, other healthcare professional)	**1**	**2**	**3**	**4**	**5**
9. Time Management of Entire Clinic Able to economize time, manage interruptions, and modify time spent with individual patients appropriately	**1**	**2**	**3**	**4**	**5**
10. If Procedures Were Performed in Clinic: **Not Applicable** ☐					
a. Technical Skills Safely and effectively performs appropriate clinical procedures	**1**	**2**	**3**	**4**	**5**
b. Situational Awareness Non-technical aspects of procedure (i.e. insight into patient experience, respects patient comfort)	**1**	**2**	**3**	**4**	**5**
11. Concerns with Attitude or Professionalism **(If yes please describe in suggestions for improvement below)**		**No**		**Yes**	

Resident is safe to independently manage/run this clinic at a generalist level **No** **Yes**

13. Give at least 1 specific aspect of clinic done well

14. Give at least 1 specific suggestion for improvement

Figure 22.3 Typical OCAT form.

Please refer to curriculum at www.mmc.nhs.uk for details of expected competencies for F1 and F2

Case-based Discussion (CbD) – F2 Version

Please complete the question using a cross: ☒ Please use black ink and CAPITAL LETTERS

Doctor's Surname

Forename

GMC Number: **GMC NUMBER MUST BE COMPLETED**

Clinical setting: A&E ☐ OPD ☐ In-patient ☐ Acute Admission ☐ GP Surgery ☐

Clinical problem category: Pain ☐ Airway/ Breathing ☐ CVS/ Circulation ☐ Psych/ Behav ☐ Neuro ☐ Gastro ☐ Other

Focus of clinical encounter: Medical Record Keeping ☐ Clinical Assessment ☐ Management ☐ Professionalism ☐

Complexity of case: Low ☐ Average ☐ High ☐

Assessor's position: Consultant ☐ SpR ☐ GP ☐

Please grade the following areas using the scale below:	Below expectations for F2 completion		Borderline for F2 completion	Meets expectations for F2 completion	Above expectations for F2 completion		U/C*
	1	2	3	4	5	6	
1 Medical record keeping	☐	☐	☐	☐	☐	☐	☐
2 Clinical assessment	☐	☐	☐	☐	☐	☐	☐
3 Investigation and referrals	☐	☐	☐	☐	☐	☐	☐
4 Treatment	☐	☐	☐	☐	☐	☐	☐
5 Follow-up and future planning	☐	☐	☐	☐	☐	☐	☐
6 Professionalism	☐	☐	☐	☐	☐	☐	☐
7 Overall clinical judgement	☐	☐	☐	☐	☐	☐	☐

*U/C Please mark this if you have not observed the behaviour and therefore feel unable to comment.

Anything especially good? **Suggestions for development**

Agreed action:

	Not at all									Highly
Trainee satisfaction with CbD	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐	8 ☐	9 ☐	10 ☐
Assessor satisfaction with CbD	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐	8 ☐	9 ☐	10 ☐

What training have you had in the use of this assessment tool?: ☐ Have Read Guidelines ☐ Face-to-Face ☐ Web/CD rom

Time taken for discussion: (in minutes)

Assessor's Signature:

Date: / /

Time taken for feedback: (in minutes)

Assessor's Surname

Assessor's GMC Number:

Please note:
Failure of return of all completed forms to your administrator is a probity issue

2466400642

Figure 22.4 Typical CbD assessment form.

the methods used, CSR was considered most valid by practising doctors.

In a later study, Norman et al. applied several different assessment methods to a group of doctors, who were referred because of practice problems, and a group of volunteers [42]. CSR was correlated with an SP examination (0.74) and an oral examination (0.51) given to these same groups. In addition, CSR was able to distinguish the 'referred' from the volunteer group.

Finally, Solomon et al. gave CSR to a group of doctors eligible for recertification [43]. It was correlated with an oral exam (0.49), and when it was combined with the oral exam, CSR had correlations with written and oral exams administered 10 years earlier (0.45, 0.37).

MSF

Making judgements about the performance of colleagues has formed the basis of the referral process in medicine and other professions for centuries [44]. In recent years, these judgements have been collected in a systematic fashion and aggregated to provide an assessment of performance. There are several variations of this process (known as multisource, or 360-degree, feedback – MSF): the assessors (e.g. peers, seniors, patients), and the forms that are used, but as an example, this chapter focuses on the mini-Peer Assessment Tool or mini-PAT.

Trainees nominate assessors who are consultants, experienced specialist registrars, staff grade and associate specialists, general practitioners, nurses, or allied health personnel. Each is sent a questionnaire, which, after completion, is returned to a central location for processing. This ensures that the trainee does not know the views of their assessors. The trainee self-assesses and submits the questionnaire for processing as well. Figure 22.5 shows the mini-PAT questionnaire. It contains 16 questions addressing the categories of:

- good clinical care
- maintaining good medical practice
- teaching and training – appraising and assessing
- relationships with patients
- working with colleagues
- an overall assessment.

Unlike the other methods that are described in this chapter, the judgements underpinning the mini-PAT are grounded in routine performance rather than performance on a specific encounter. As with the other methods, however, assessors are asked to make judgements about both quality and fitness.

It is most efficient if the questionnaires are collated electronically in a central location, and feedback is prepared for the trainee. The reports to trainees typically provide self-ratings, the mean ratings of the assessors, and the national mean ratings. All comments are included verbatim, but they are anonymous. The trainee and their educational supervisor review the results together and agree on strengths, areas for development, and an action plan. This process can be repeated as often as makes sense in the context of training.

Multisource feedback (MSF) programmes of this type have been used at a number of different institutions [45–49]. For 30 years, medical students in the obstetrics/gynaecology and internal medicine clerkships at the University of Missouri-Kansas have been asked to evaluate the professionalism of their peers [50]. The programme has evolved over the years, but recently most of the negative reports about professional behaviours received by the promotions committee emanated from peers [51].

Similarly, the University of Florida uses an MSF assessment system to identify those medical students whose professionalism is outstanding [52]. The information is included in the Dean's letter of recommendation for postgraduate training, and there are reports that it has enabled some students to acquire a more desirable post [51].

There is a growing body of evidence supporting the validity of these assessments. A study of certification in the US by Ramsey et al. compared certified internists and non-certified internists who were 5–10 years past certification or training [53]. Several assessments were collected, and the certified doctors had higher peer ratings, even though their peers did not know their certification status. The results were also correlated with written examination performance.

In a follow-up study, Ramsey et al. focused only on practising internists who were 5–15 years past certification [54]. Two lists of peers were solicited, one from the participants and one from their medical supervisors. In addition, the questions on the assessment form were divided into two scales, one for cognitive/technical skills and one for professionalism. The source of the peers (participant versus medical supervisor) did not affect the ratings. Further, a written examination had a statistically significant correlation with the cognitive/technical scale but not the professionalism scale.

The mini-PAT (Figure 22.5) is a shortened form of the Sheffield Peer Review Assessment Tool, which was studied with paediatricians [55]. Results of that work indicate that it was feasible, produced scores with reasonable reliability, and was not significantly influenced by extraneous factors, such as occupation of assessor, length of working relationship, and the clinical setting in which the relationship took place. Moreover, it was able to distinguish between doctors of different grades. It was used in the Foundation Programme at its inception but has since been replaced with the Team Assessment of Behaviour (TAB), which is shorter and so more practical [56].

In order to develop a system for MSF it is important to nurture an institutional climate that motivates change in behaviour. It is also essential to train those giving feedback to deliver feedback in a non-threatening manner [57]. Performance change is more likely to occur when feedback was credible and accurate or when coaching is provided [58].

Portfolios

A portfolio is simply a collection of information intended to demonstrate achievement. However, there are many variations on this simple theme. Portfolios can include a number of different types of information, both single-encounter and routine performance. In some instances, the content of the portfolio is left wholly to the trainee, while in others they must all include exactly the same information. Typically, the portfolio is reviewed by assessors, who are asked to make judgements about it.

Please refer to curriculum at www.mmc.nhs.uk for details of expected competencies for F1 and F2

mini-PAT (Peer Assessment Tool) – F1 Version

Please complete the question using a cross: ☒ Please use black ink and CAPITA L LETTERS

Doctor's Surname

Forename

GMC Number:

How do you rate this Doctor in their:	Below expectations for F1 completion		Borderline for F1 completion	Meets expectations for F1 completion	Above expectations for F1 completion		U/C*
	1	2	3	4	5	6	
Good Clinical Care							
1 Ability to diagnose patient problems	☐	☐	☐	☐	☐	☐	☐
2 Ability to formulate appropriate management plans	☐	☐	☐	☐	☐	☐	☐
3 Awareness of their own limitations	☐	☐	☐	☐	☐	☐	☐
4 Ability to respond to psychosocial aspects of illness	☐	☐	☐	☐	☐	☐	☐
5 Appropriate utilisation of resources e.g. ordering investigations	☐	☐	☐	☐	☐	☐	☐
Maintaining good medical practice							
6 Ability to manage time effectively/ prioritise	☐	☐	☐	☐	☐	☐	☐
7 Technical skills (appropriate to current practice)	☐	☐	☐	☐	☐	☐	☐
Teaching and Training, Appraising and Assessing							
8 Willingness and effectiveness when teaching/training colleagues	☐	☐	☐	☐	☐	☐	☐
Relationship with Patients							
9 Communication with patients	☐	☐	☐	☐	☐	☐	☐
10 Communication with carers and/or family	☐	☐	☐	☐	☐	☐	☐
11 Respect for patients and their right to confidentiality	☐	☐	☐	☐	☐	☐	☐
Working with colleagues							
12 Verbal communication with colleagues	☐	☐	☐	☐	☐	☐	☐
13 Written communication with colleagues	☐	☐	☐	☐	☐	☐	☐
14 Ability to recognise and value the contribution of others	☐	☐	☐	☐	☐	☐	☐
15 Accessibility/reliability	☐	☐	☐	☐	☐	☐	☐
16 Overall, how do you rate this doctor compared to a doctor ready to complete F1 training?	☐	☐	☐	☐	☐	☐	☐

Do you have any concerns about this doctor's probity or health? ☐ Yes ☐ No
If yes please state your concerns:

*U/C Please mark this if you have not observed the behaviour and therefore feel unable to comment. 6927534062

Figure 22.5 The mini-PAT questionnaire.

A portfolio for use in postgraduate training might include several components or sections. For instance, it may contain a section devoted to the educational experiences of the trainee, such as procedure and patient logs, participation in didactic sessions, clinical rotations, research papers, and critical incidents. A second section might contain the results of workplace-based assessments plus scores on written knowledge examinations. A third section could be devoted to the trainee's reflections on these educational experiences and ongoing self-appraisal. A final section might contain all of the signoffs necessary to support a decision about the trainee's promotion.

Because portfolios used for different purposes might contain different forms of assessment, it is not reasonable to extrapolate the validity of one portfolio from another. However, it is sensible to extrapolate the validity of the portfolio from the validity of the contents (i.e. mini-CEX, DOPS, CSR, MSF) and the quality of the process used to make decisions based on those contents.

Regardless of the particular application, the use of portfolios in assessment requires that one or more experts make judgements about their contents. Applying the framework presented above, these judgements can be made about occurrence (e.g. whether a trainee has had the required number of mini-CEX encounters), quality (e.g. whether the mini-CEX results indicated good performance), and fitness (e.g. whether the mini-CEX results indicated satisfactory completion of a year of training). There are several factors that will contribute to the quality of these judgements, and they are similar to those for all assessments based on observation.

First, the purpose of the portfolio must be clear (e.g. it is intended for summative use). Second, it is important to be specific about what each portfolio must contain and to have as much commonality across them as possible; this enhances the ability to make comparisons both among trainees and against standards. Third, the portfolios should be based on as many independent assessments of performance as is feasible. This is equivalent to having a number of different encounters in the mini-CEX or DOPS. Fourth, several examiners should be involved in making judgements about each portfolio; this reduces the effects of examiners who differ significantly in stringency. Portfolio-based learning and assessment is covered in more detail in Chapter 18 of this book.

Influence on Learning

The workplace offers a rich environment for assessment, and the observational methods described above can be of considerable use in this regard. It is perhaps more important, however, that several of them offer the opportunity for formative assessment and feedback at the same time. Unfortunately, there are data suggesting that a sizeable majority of postgraduate trainees are never observed in a patient encounter [15]. Not only do the methods described above require that observation, but in addition to summative assessment, the observations are intended to serve as a basis for formative assessment and educational feedback (see Chapter 25).

Holmboe et al. have developed a programme for providing feedback [59]. At the completion of the assessment process, the observers need to provide an evaluation of the trainees' strengths and weaknesses while enabling trainees to react to these. Faculty members then need to encourage self-assessment and develop action plans that will enable the trainees to address any deficiencies. These components are sometimes documented on the mini-CEX, DOPS, and CSD forms that are used as part of workplace assessment. Feedback between the educational supervisor and trainee is also an important part of MSF.

Faculty Development

Faculty development is one key to the success of workplace assessment based on observation. Holmboe et al. have developed an excellent workshop, which has applicability to a variety of observational methods [60]. It consists of three major pieces. First, there is training in behavioural observation, including knowing what to look for, preparing the trainee and patient, and minimising intrusiveness and interference. Second, there is performance dimension training, in which faculty members decide on the dimensions of performance that are important. Finally, there is frame of reference training, where faculty members practise to improve their accuracy and discrimination and reduce differences in stringency.

The workshop consists of didactic mini-lectures, small group, and videotape evaluation exercises and practice with standardised trainees and patients. A randomised control trial of this model showed that faculty members who underwent training thought it was excellent, felt more comfortable performing direct observation, and were more stringent than the control group faculty members.

In addition to the focus on assessment, faculty development efforts should also be provided for the provision of educational feedback. This goes well beyond the methods presented here, but given the few times trainees are actually observed in their work, workplace-based assessment offers an opportunity that should not be missed.

Challenges in Workplace-based Assessment

For the methods described above, as well as others that are based on observation, there is a series of challenges to the assessment process. Among them are reliability, equivalence, integrating the summative and formative, stakes, relationships, the need for other forms of assessment, feasibility, and future research.

Reliability

If the same trainee were examined on different occasions with different patients by different assessors, we would want their results to be the same. This is called reliability or reproducibility. In the observation of clinical performance, three major factors affect reliability: the number of encounters observed (both single encounters and routine performance), the number of assessors, and the aspects of performance being evaluated [61].

Seminal work by Elstein et al. in the mid-1970s indicated that doctors' performances were case specific; performance on one case only weakly predicted performance on others [6]. This finding has been replicated several times. Consequently, it is important to observe trainees with a number of different patients before having confidence in the results of their assessment. All of the methods used in the Foundation Programme require multiple encounters with patients.

Likewise, there is research showing that experienced assessors, even when observing exactly the same encounter, differ in their opinions about its difficulty and quality [62]. More importantly, they interact with each encounter along the lines of their own strengths, weaknesses, and experiences. Consequently, it is important to include assessments from different faculty members to achieve reliable results. Again, all of the methods used in the Foundation Programme require multiple observers and a different assessor for each different encounter.

Finally, there is research indicating that more reliable results are obtained when assessors are asked to judge a number of aspects of a performance rather than making a single overall judgement about it [54]. At the same time, asking for too many judgements does little to improve reliability and adversely affects feasibility. The exact number will, of course, vary with the characteristics of the performance and the nature of the judgements being made (more for quality and fitness, less for occurrence), but generally 5–10 questions should be sufficient.

Some factors, such as the exact wording of the questions and the number of points on the scale, have only a very modest influence on reliability. But because these things are obvious and easy to change, some users spend considerable effort on them. Assuming that reasonable care has been taken, this effort would better be spent recruiting and training assessors or observing additional encounters.

Equivalence

An important issue in the use of observational methods is whether the assessments of trainees are equivalent to one another. In most of the methods described above, different patients or their records serve as the basis for assessment, and these may differ in complexity. Likewise, different faculty members and different peers act as assessors for the trainees, and they are not equally stringent. In portfolios that do not have strict requirements for what is included, there is considerable variability in the basis for assessment. As a consequence, it is not clear whether trainee differences are due to their ability or to the difficulty of the encounters and assessors that trainees face.

This is less of a problem within each training programme, where presumably the assessors and clinical settings are similar. It is also less troublesome when the stakes are lower and there is an emphasis on feedback and formative assessment. However, the observational assessments of the type described above do not yield comparable scores at the regional or national level because faculties can differ appreciably, as do the clinical settings for training. Using a common problem list, involving a number of assessors for each trainee, and providing good faculty development will lessen the impact of these problems, but they remain significant.

Integrating Formative and Summative Assessment

The aim of formative assessment is to support learning, in contrast to summative assessment, which acts as a measure of it. To prevent formative assessment from being a 'series of punch biopsies performed by independent physicians who do not communicate with each other', an assessment system that enables teachers to assist students in developing and achieving their learning goals is essential. The educational plan–do–study–act (PDSA) cycle is a tool adapted from quality improvement processes in other disciplines and it can be used by teachers and students to guide learning in a formatively focused assessment system. In the presence of longitudinal academic coaching, the PDSA cycle should generate information that: (i) is systematically translated into detailed feedback that informs students about their performance, (ii) leads to the development of specific plans for improvement supported and guided by faculty, (iii) is subject to follow-through whereby students present evidence of progress, and (iv) is part of a continuous cycle over time. Such a process would encourage the learner to engage in a personal, educational PDSA cycle [63, 64].

While summative assessment is essential to entrustment and competency decisions, integrating it with formative assessment in a systematic fashion will support the entire educational enterprise. In a move in this direction, UK training programmes have rebadged workplace-based assessments as 'supervised learning events' (SLEs) designed to engage doctors in training in reflective practice and self-directed learning from patients, clinical opportunities, books, journals, and electronic learning materials. The same WPBA tools are used (e.g. mini-CEX, DOPS, MSF) but their formative nature is made more explicit with overarching summative judgements of progress whilst informed by WPBAs, being made elsewhere [65, 66].

Stakes

Although it has not been studied extensively, there is some evidence to indicate that assessments based on observation are influenced by what is at stake. Certainly, authors have found that when used for an important purpose (e.g. promotion, continued certification), grades tend to be very high and very few trainees are considered unsatisfactory [67–69]. Although it is possible that doctors in these studies are well above average, it is more likely that peers and faculty members are reluctant to provide negative assessments when the stakes are moderate or high.

There is no way to avoid this problem completely, but there are some things that will lessen its effect. The use of external examiners reduces the amount of prior information available and lessens the assessors' personal stake in the trainee. For both external assessors and internal assessors, faculty development is important and has been shown to increase the stringency of the grades they provide [59]. For any sort of MSF, anonymity is crucial. Where appropriate, it is also helpful to restrict the nature of the judgements to occurrence or quality rather than fitness. The latter requires a high-stakes decision, and these are often better made using other faculty members and other methods [70].

Relationships

The relationship between trainees and their observers can adversely affect the validity of the assessments. In the Foundation Programme, these effects might be exacerbated by the fact that trainees choose their assessors [13]. In addition, when assessors are faculty members, their role as educator is in conflict with their role as evaluator. Likewise, the relationships between the trainee and those peers who complete the MSF questionnaires are varied and might influence the results.

Although there is no way to avoid the effects of relationships, there are some steps that might decrease their influence. Where feasible, using external assessors reduces some of the concerns. Similarly, ensuring anonymity in the context of MSF is important. It may also be useful if the assessor selects the trainee, the patients and the peers, rather than having them chosen by the trainee. However, the effect of self-selection of assessors versus selection by others did not make a difference when studied by Ramsey et al. [54].

Need for Other Forms of Assessment

Observation in the workplace constitutes a powerful tool for assessment and feedback. However, it does not serve all purposes, and other forms of assessment are better suited to some functions. For example, differences in assessors and clinical material make it unwise to compare trainees regionally or nationally purely using workplace assessment. Assessment based on observation in the workplace should not be the sole basis on which to rank trainees and select them for additional educational experiences. Likewise, this form of assessment alone is not suitable for making high-stakes end-of-training decisions such as those for certification in the US. In both instances, assessment based on observation in the workplace needs to be supplemented with a national programme that includes tests of knowledge and clinical skill.

In addition to national ranking and achievement testing, observation in the workplace is best suited to an assessment of integrated skills in the context of patient care. Where problem trainees are identified, this form of assessment is neither ideal nor efficient in determining relative areas of strength and weakness. For diagnostic purposes, it is best to follow with assessments of medical knowledge and clinical skill.

Feasibility

There are significant resource constraints in the workplace setting, and the methods described here have been designed to be as efficient as possible given this context. Nonetheless, there remain serious challenges to carrying out this form of assessment. Clearly, the centralisation of functions, such as collecting and reporting data for MSF, significantly enhances the feasibility of the method. Likewise, a national programme of faculty development is useful.

Despite these efficiencies, local administration strategies are still needed. In the case of the Foundation Programme, trainees are given responsibility for ensuring that the assessments are completed as part of the SLEs [13]. This is effective but creates challenges based on the assessor–trainee relationship. In the US, the assessors have chosen the trainee and patient, but enlisting the cooperation of busy clinician-educators is difficult. This is an area that requires more attention to ensure the validity and feasibility of the methods.

Emerging Discourses

Recently, a number of discourses have emerged which relate to the use of workplace-based assessment. Among them are entrustable professional activities, programmatic assessment design, and rater cognition.

Entrustable Professional Activities

Entrustable professional activities (EPAs) are units of practice or tasks that can be independently executed, measured, and observed, such as taking a medical history, performing a physical examination, and interpreting a series of lab results [31]. Given their nature, workplace-based assessment offers a good vehicle for assessing EPAs and several groups have used this link to bridge the gap between competency-based education and clinical practice [71]. For example, the Association of American Medical Colleges has also recently produced a document called 'Core Entrustable Professional Activities for Entering Residency' (CEPAER) [32]. It identified 13 core EPAs that medical school graduates are expected to be able to perform at a level that permits practice without direct supervision on day one of their postgraduate training. For example, EPA 1 in the CEPAER document is to 'gather a history and perform a physical exam'. The mini-CEX, CSD, and MSF can all contribute information to inform EPA 1. Similarly, DOPS assessing a learner's ability to obtain intravenous access can provide data for EPA 10: 'Recognize a patient requiring urgent care and initiate evaluation and management'.

Mastery of an EPA is marked by the decision to trust a trainee for unsupervised practice. This decision is based on multiple observed WPBAs, with a gradual decrease in the level of supervision. This Statement of Awarded Responsibility scale below is an example that has been adapted by many institutions [34, 72, 73]:

1 Observing the activity.
2 Acting with direct supervision present in the room.
3 Acting with supervision available within minutes.
4 Acting unsupervised (i.e. under clinical oversight).
5 Providing supervision to juniors.

As medical educators seek to document progress along the training continuum, portfolios can help archive and track learner's development. As noted above, they are particularly well suited to providing a repository that supports documentation, aggregation, and analysis based on workplace-based assessment, which is fundamental for competency-based education and EPAs [74]. Longitudinal documentation of achievement of specific goals and learning outcomes using a variety of assessment tools and assessors, in different contexts, helps to show learners their progress across multiple competency domains at a glance. This 'dashboard' approach can potentially be effective in documenting the achievement of specific EPAs, leading to decisions about entrustment and competencies [75].

Programmatic Assessment Design

The increase in the number and type of workplace-based assessments being done as part of educational programmes has led to the need for a systematic approach to what is collected, when it is collected, and strategies for combining the results. Programmatic assessment design is an approach that is driven by the competencies that need to be assessed. When developed, the system offers direction for: (i) combining information from different tests/tools, (ii) aggregating multiple individual assessment points, (iii) compensating for deficiencies by combining several instruments, and (iv) aggregating data used for high-stakes decisions [76–78]. The programmatic assessment design will also detail the assessment purpose and the expertise of stakeholders while providing a 'central governance' to review the design process and employ educational quality improvement plan [76]. A programmatic assessment design is therefore imperative for successful implementation of workplace-based assessment.

Assessor Cognition

The fact that assessors do not necessarily agree even when they are observing the exact same behaviour has been a source of concern for workplace-based methods [62]. This has given rise to three perspectives on why these differences exist [79]. One perspective suggests that assessor variability is due to idiosyncratic but meaningful differences among the assessors, so differences are desirable reflections of context and experience. A second perspective asserts that assessor differences are a result of the application of different frameworks or criteria and is amenable to training, at least in part. A final perspective asserts that the variability is a result of human limitations in cognition and so not amenable to training.

At this point, the research does not substantially support one perspective or another and all three perspectives have something to offer [79]. It would not be surprising if some of the differences between assessors were meaningful and valid, some resulted from cognitive limitations, and some were based on the application of different criteria. As noted above, faculty development programmes are key to helping resolve these tensions by illustrating how the workplace-based assessment tools like mini-CEX, DOPS, and MSF allow documentation of context and complexity of tasks, which should be taken into account while making judgements regarding competence.

Future Research

Although observation and feedback has been central to education for millennia, research on systematic observation and feedback in the medical workplace is in its infancy. A recent systematic review of direct observation concluded that although there are numerous methods, studies of their validity and outcomes are limited [16]. Similarly, a systematic review of the impact of workplace-based assessment on doctors' education and performance found subjective reports of educational impact, but no research in the area [80]. Going forward, addressing these deficits in the literature will be essential even though studies of this type are difficult to do well. In the meantime, there is considerable support in the general education literature for the central role of feedback in achievement [81]. Also, feedback is not possible without observation and assessment.

Conclusion

Assessment in the setting of clinical training, particularly postgraduate training, is not as well developed as assessment in the undergraduate arena. Over the past decade, several methods based on the observation of routine trainee–patient encounters and interactions with colleagues have been proposed. These methods have been used in the UK, the US, and other countries around the world.

The mini-CEX, DOPS, and CSR are based on the observation of a single performance with a patient or a medical record, while MSF captures routine performance. In all instances, judgements about the quality and fitness of the performance(s) are made by the assessor(s), and there is considerable research supporting the validity of these methods.

The opportunity for educational feedback as part of these methods is as important as their contribution to the assessment process. At the completion of each assessment, the observers need to provide an evaluation of the trainees' strengths and weaknesses while enabling them to react to these. Faculty members then need to encourage self-assessment and develop action plans, which will enable the trainees to address any deficiencies. In a sense, the methods bring together summative and formative assessment and create a teaching moment that skilled faculty members can grasp.

A portfolio is a collection of a variety of different assessments and experiences. Because of the variability in their contents, it is not reasonable to generalise from studies of the validity of one portfolio to another. However, it is possible to extrapolate the validity of the portfolio from the validity of the contents (i.e. the WPBAs within) and the quality of the process used to make judgements about it.

Faculty development is a key to the successful use of these methods. A model workshop would consist of training in behavioural observation, performance dimension training, and frame of reference training, along with considerable practice. Periodic but shorter versions of the workshop are needed to maintain faculty involvement and proficiency. In addition, the provision of educational feedback should be an important part of the overall faculty development effort.

Finally, a series of challenges remains, including reliability, equivalence, integrating the formative and the summative, stakes, relationships, the need for other forms of assessment, and feasibility. Given the nature of the methods, some of these (e.g. equivalence) will be very difficult to surmount. Consequently, it will be important to deploy workplace-based methods appropriately and to ensure that they are only one piece of a larger assessment programme.

Acknowledgements

The assessment forms reproduced, by permission, here are for illustrative purposes only and may not reflect versions in current use. Our thanks to the UK Foundation Programme, the American Board of Internal Medicine, the Department of Surgical Education of The University of Ottawa, the Federation of Royal Colleges of Physicians, the American Board of Emergency Medicine, and the University of Sheffield.

References

1 Norman, G.R. (2002). Research in medical education: three decades of progress. *British Medical Journal* 324: 1560–1562.

2 Norcini, J.J. (2005). Current perspectives in assessment: the assessment of performance at work. *Medical Education* 39: 880–889.

3 Miller, G.E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine* 65 (9): s63–s67.

4 Norcini, J.J. (2003). Peer assessment of competence. *Medical Education* 37: 539–543.

5 Norcini, J.J. (2006). Faculty observations of student professional behaviour. In: *Measuring Professionalism* (ed. D. Stern), 147–158. Oxford: Oxford University Press.

6 Elstein, A.S., Shulman, L.S., and Sprafka, S.A. (1978). *Medical Problem-Solving: An Analysis of Clinical Reasoning*. Cambridge, MA: Harvard University Press.

7 Pulito, A.R., Donnelly, M.B., Plymale, M., and Mentzer, R.M. Jr. (2006). What do faculty observe of medical students' clinical performance? *Teaching and Learning in Medicine* 18 (2): 99–104.

8 Silber, C.G., Nasca, T.J., Paskin, D.L. et al. (2004). Do global rating forms enable program directors to assess the ACGME competencies? *Academic Medicine* 79 (6): 549–556.

9 Brennan, R.L. (2001). *Generalizability Theory*. New York: Springer-Verlag.

10 Norcini, J.J., Blank, L.L., Duffy, F.D., and Fortna, G. (2003). The mini-CEX: a method for assessing clinical skills. *Annals of Internal Medicine* 138: 476–481.

11 Martin, J.A., Regehr, G., Reznick, R. et al. (1997). Objective structured assessment of technical skill (OSATS) for surgical residents. *British Journal of Surgery* 84 (2): 273–278.

12 Norcini, J.J. and Boulet, J.R. (2003). Methodological issues in the use of standardized patients for assessment. *Teaching and Learning in Medicine* 15 (4): 293–297.

13 Davies, H., Archer, J., Southgate, L., and Norcini, J. (2009). Initial evaluation of the first year of the Foundation Assessment Programme. *Medical Education* 43: 74–81.

14 Norcini, J.J., Blank, L.L., Arnold, G.K., and Kimball, H.R. (1995). The mini-CEX (clinical evaluation exercise): a preliminary investigation. *Annals of Internal Medicine* 123: 795–799.

15 Day, S.C., Grosso, L.G., Norcini, J.J. et al. (1990). Residents' perceptions of evaluation procedures used by their training program. *Journal of General Internal Medicine* 5: 421–426.

16 Kogan, J.R., Holmboe, E.S., and Hauer, K.E. (2009). Tools for direct observation and assessment of clinical skills of medical trainees. *Journal of the American Medical Association* 302 (12): 1316–1326.

17 Kogan, J.R., Bellini, L.M., and Shea, J.A. (2003). Feasibility, reliability, and validity of the mini-clinical evaluation exercise (mCEX) in a medicine core clerkship. *Academic Medicine* 78 (s10): S33–S35.

18 Durning, S.J., Cation, L.J., Markert, R.J., and Pangaor, L.N. (2002). Assessing the reliability and validity of the mini-clinical evaluation exercise for internal medicine residency training. *Academic Medicine* 77 (9): 900–904.

19 Boulet, J.R., McKinley, D.W., Norcini, J.J., and Whelan, G.P. (2002). Assessing the comparability of standardized patient and physician evaluations of clinical skills. *Advances in Health Sciences Education* 7: 85–97.

20. Holmboe, E.S., Huot, S., Chung, J. et al. (2003). Construct validity of the miniClinical Evaluation Exercise (miniCEX). *Academic Medicine* 78: 826–830.

21 Cruess, R., McIlroy, J.H., Cruess, S. et al. (2006). The professionalism mini-evaluation exercise: a preliminary investigation. *Academic Medicine* 81 (s10): S74–S78.

22 Hatala, R., Ainslie, M., Kassen, B.O. et al. (2006). Assessing the mini-clinical evaluation exercise in comparison to a national specialty examination. *Medical Education* 40 (10): 950–956.

23 Kogan, J.R. and Hauer, K.E. (2006). Brief report: use of the mini-clinical evaluation exercise in internal medicine core clerkships. *Journal of General Internal Medicine* 21 (5): 501–502.

24 Wragg, A., Wade, W., Fuller, G. et al. (2003). Assessing the performance of specialist registrars. *Clinical Medicine* 3 (2): 131–134.

25 Larson, J.L., Williams, R.G., Ketchum, J. et al. (2005). Feasibility, reliability and validity of an operative performance rating system for evaluating surgery residents. *Surgery* 138 (4): 640–647.

26 Goff, B.A., Nielsen, P.E., Lentz, G.M. et al. (2002). Surgical skills assessment: a blinded examination of obstetrics and gynecology residents. *American Journal of Obstetrics and Gynecology* 186 (4): 613–617.

27 Winckel, C.P., Reznick, R.K., Cohen, R., and Taylor, B. (1994). Reliability and construct validity of a structured technical skills assessment form. *American Journal of Surgery* 167 (4): 423–427.

28 Marriott, J., Purdie, H., Crossley, J., and Beard, J.D. (2010). Evaluation of procedure-based assessment for assessing trainees' skills in the operating theatre. *British Journal of Surgery* 98 (3): 450–457.

29 Rekman, J., Hamstra, S.J., Dudek, N. et al. (2016). A new instrument for assessing resident competence in surgical clinic: the Ottawa clinic assessment tool. *Journal of Surgical Education* 73 (4): 575–582.

30 George, B.C., Teitelbaum, E.N., Meyerson, S.L. et al. (2014). Reliability, validity, and feasibility of the zwisch scale for the assessment of intraoperative performance. *Journal of Surgical Education* 71 (6): e90–e96.

31 ten Cate, O. (2005). Entrustability of professional activities and competency-based training. *Medical Education* 39: 1176–1177.

32 Association of American Medical Colleges (2014). Core Entrustable Professional Activities for Entering Residency. Curriculum Developers Guide https://members.aamc.org/eweb/upload/core%20EPA%20Curriculum%20Dev%20Guide.pdf (accessed 30 April 2017).

33 Rekman, J., Gofton, W., Dudek, N. et al. (2016). Entrustability scales: outlining their usefulness for competency-based clinical assessment. *Academic Medicine* 91 (2): 186–190.

34 Chen, H.C., van den Broek, W.E., and ten Cate, O. (2015). The case for use of entrustable professional activities in undergraduate medical education. *Academic Medicine* 90 (4): 431–436.

35 Martin, J.A., Reznick, R.K., Rothman, A. et al. (1996). Who should rate candidates in an objective structured clinical examination? *Academic Medicine* 71 (2): 170–175.

36 Hodges, B., McNaughton, N., Regehr, G. et al. (2002). The challenge of creating new OSCE measures to capture the characteristics of expertise. *Medical Education* 36 (8): 742–748.

37 Hodges, B., Regehr, G., McNaughton, N. et al. (1999). OSCE checklists do not capture increasing levels of expertise. *Academic Medicine* 74 (10): 1129–1134.

38 Rothman, A.I., Blackmore, D., Dauphinee, W.D., and Reznick, R. (1997). The use of global ratings in OSCE station scores. *Advances in Health Sciences Education* 1: 215–219.

39 Regehr, G., MacRae, H., Reznick, R.K., and Szalay, D. (1998). Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Academic Medicine* 73 (9): 993–997.

40 Regehr, G., Freeman, R., Robb, A. et al. (1999). OSCE performance evaluations made by standardized patients: comparing checklist and global rating scores. *Academic Medicine* 74 (s10): S135–S137.

41 Maatsch, J.L., Huang, R., Downing, S., and Barker, B. (1983). Predictive Validity of Medical Specialty Examinations. Final report for Grant HS 02038-04, National Center of Health Services Research. Office of Medical Education and Research and development, Michigan State University, East Lansing, MI.

42 Norman, G.R., Davis, D., Painvin, A., et al. (1989). Comprehensive assessment of clinical competence of family/general physicians using multiple measures. *Proceedings of the Annual Conference on Research in Medical Education*, p. 75. Washington, DC: Association of American Medical Colleges.

43 Solomon, D.J., Reinhart, M.A., Bridgham, R.G. et al. (1990). An assessment of an oral examination format for evaluating clinical competence in emergency medicine. *Academic Medicine* 65 (s9): S43–S44.

44 Norcini, J.J. (2003). Peer assessment of competence. *Medical Education* 37: 539–543.

45 Massagli, T.L. and Carline, J.D. (2007). Reliability of a 360-degree evaluation to assess resident competence. *American Journal of Physical Medicine & Rehabilitation* 86 (10): 845–852.
46 Rees, C. and Shepherd, M. (2005). The acceptability of 360-degree judgements as a method of assessing undergraduate medical students' personal and professional behaviours. *Medical Education* 39 (1): 49–57.
47 Whitehouse, A., Hassell, A., Wood, L. et al. (2005). Development and reliability testing of TAB a form for 360 assessment of senior house officers' professional behaviour, as specified by the General Medical Council. *Medical Teacher* 27 (3): 252–258.
48 Lockyer, J. (2003). Multisource feedback in the assessment of physician competencies. *The Journal of Continuing Education in the Health Professions* 23 (1): 4–12.
49 Rodgers, K.G. and Manifold, C. (2002). 360-degree feedback: possibilities for assessment of the ACGME core competencies for emergency medicine residents. *Academic Emergency Medicine* 9 (11): 1300–1304.
50 Arnold, L., Willoughby, L., Calkins, V., and Eberhart, G. (1981). Use of peer evaluation in the assessment of medical students. *Medical Education* 56: 35–42.
51 Arnold, L. and Stern, D. (2006). Content and context of peer assessment. In: *Measuring Professionalism* (ed. D. Stern), 175–194. Oxford: Oxford University Press.
52 Small, P.A., Stevens, B., and Duerson, M.C. (1993). Issues in medical education: basic problems and potential solutions. *Academic Medicine* 68: S89–S98.
53 Ramsey, P.G., Carline, J.D., Inui, T.S. et al. (1989). Predictive validity of certification by the American Board of Internal Medicine. *Annals of Internal Medicine* 110 (9): 719–726.
54 Ramsey, P.G., Wenrich, M.D., Carline, J.D. et al. (1993). Use of peer ratings to evaluate physician performance. *Journal of the American Medical Association* 269: 1655–1660.
55 Archer, J.C., Norcini, J.J., and Davies, H.A. (2005). Peer review of paediatricians in training using SPRAT. *British Medical Journal* 330: 1251–1253.
56 Whitehouse, A., Hassell, A., Bullock, A. et al. (2007). 360 degree assessment (multisource feedback) of UK trainee doctors: field testing of team assessment of behaviours (TAB). *Medical Teacher* 29: 171–176.
57 Wood, L., Hassell, A., Whitehouse, A. et al. (2006). A literature review of multi-source feedback systems within and without health services, leading to 10 tips for their successful design. *Medical Teacher* 28 (7): e185–e191.
58 Miller, A. and Archer, J. (2010). Impact of workplace based assessment on doctors' education and performance: a systematic review. *British Medical Journal* 341: c5064.
59 Holmboe, E.S., Yepes, M., Williams, F., and Huot, S.J. (2004). Feedback and the mini clinical evaluation exercise. *Journal of General Internal Medicine* 19 (5 Pt 2): 558–561.
60 Holmboe, E.S., Hawkins, R.E., and Huot, S.J. (2004). Effects of training in direct observation of medical residents' clinical competence: a randomized trial. *Annals of Internal Medicine* 140 (11): 874–881.
61 Norcini, J.J. (2001). The validity of long cases. *Medical Education* 35: 735–736.
62 Noel, G.L., Herbers, J.E., Caplow, M.P. et al. (1992). How well do internal medicine faculty members evaluate the clinical skills of residents? *Annals of Internal Medicine* 117: 757–765.
63 Johnson, P. and Raterink, G. (2009). Implementation of a diabetes clinic-in-a-clinic project in a family practice setting: using the plan, do, study, act model. *Journal of Clinical Nursing* 18 (14): 2096–2103.
64 Konopasek, L., Norcini, J., and Krupat, E. (2016). Focusing on the formative: building an assessment system aimed at student growth and development. *Academic Medicine* 91 (11): 1492–1497.
65 Rees, C.E., Cleland, J.A., Dennis, A. et al. (2014). Supervised learning events in the foundation programme: a UK-wide narrative interview study. *BMJ Open* 4 (10): e005980.
66 Academy of Medical Royal Colleges (2016). The Foundation Programme Curriculum 2016. http://aomrc.org.uk/wp-content/uploads/2016/06/Foundation_Curriculum_-2016-1.pdf (accessed 28 March 2017).
67 Hay, J.A. (1995). Tutorial reports and ratings. In: *Evaluation Methods: A Resource Handbook* (ed. S Shannon and G. Norman). McMaster University, Hamilton, Ontario.
68 Ramsey, P.G., Carline, J.D., Blank, L.L., and Wenrich, M.D. (1996). Feasibility of hospital-based use of peer ratings to evaluate the performance of practicing physicians. *Academic Medicine* 71: 364–370.
69 Hall, W., Violato, C., Lewkonia, R. et al. (1999). Assessment of physician performance in Alberta: the physician achievement review. *Canadian Medical Association Journal* 161: 52–57.
70 Norcini, J.J. (2005). Standard setting. In: *A Practical Guide for Medical Teachers* (ed. J.A. Dent and R.M. Harden), 293–301. Edinburgh: Elsevier Churchill Livingstone.
71 Mulder, H., Cate, O.T., Daalder, R., and Berkvens, J. (2010). Building a competency-based workplace curriculum around entrustable professional activities: the case of physician assistant training. *Medical Teacher* 32 (10): e453–e459.
72 ten Cate, O. and Scheele, F. (2007). Competency-based postgraduate training: can we bridge the gap between theory and clinical practice? *Academic Medicine* 82 (6): 542–547.
73 ten Cate, O. (2014). What entrustable professional activities add to a competency-based curriculum. *Academic Medicine* 89 (4): 691.
74 Driessen, E., van Tartwijk, J., Vermunt, J., and van der Vleuten, C. (2003). Use of portfolios in early undergraduate medical training. *Medical Teacher* 25 (1): 18–23.
75 Chertoff, J., Wright, A., Novak, M. et al. (2015). Status of portfolios in undergraduate medical education in the LCME accredited US medical school. *Medical Teacher* 38 (9): 886–896.
76 Timmerman, A. and Dijkstra, J. (2017). A practical approach to programmatic assessment design. *Advances in Health Sciences Education*. Jan 24 January [Epub ahead of print] doi: 10.1007/s10459-017-9756-3.
77 van der Vleuten, C., Schuwirth, L., Driessen, E. et al. (2014). 12 tips for programmatic assessment. *Medical Teacher* 37 (7): 641–646.
78 van der Vleuten, C., Schuwirth, L., Driessen, E. et al. (2012). A model for programmatic assessment fit for purpose. *Medical Teacher* 34 (3): 205–214.
79 Gingerich, A., Kogan, J., Yeates, P. et al. (2014). Seeing the 'black box' differently: assessor cognition from three research perspectives. *Medical Education* 48 (11): 1055–1068.
80 Miller, A. and Archer, J. (2010). Impact of workplace based assessment on doctors' education and performance: a systematic review. *British Medical Journal* 341: c5064.
81 Norcini, J.J. (2010). The power of feedback. *Medical Education* 44: 16–17.

Further Reading

Academy of Medical Royal Colleges (2016). Improving Assessment: Further Guidance and Recommendations. http://aomrc.org.uk/wp-content/uploads/2016/06/Improving_assessment_Further_GR_0616-1.pdf (accessed 30 April 2017).

Barrett, A., Galvin, R., Steinert, Y. et al. (2015). A BEME (Best Evidence in Medical Education) systematic review of the use of workplace-based assessment in identifying and remediating poor performance among postgraduate medical trainees. *Sytstematic Reviews* 4: 65. doi: 10.1186/s13643-015-0056-9

Ten Cate, O. (2013). The nuts and bolts of entrustable professional activities *Journal of Graduate Medicine* 5(1): 157–158.

23 Structured Assessments of Clinical Competence

Katharine A. M. Boursicot[1], Trudie E. Roberts[2], and William P. Burdick[3,4]

[1]Health Professional Assessment Consultancy, Singapore, Singapore
[2]University of Leeds, Leeds, UK
[3]Foundation for Advancement of International Medical Education and Research (FAIMER), Philadelphia, PA, USA
[4]Drexel University College of Medicine, Philadelphia, PA, USA

KEY MESSAGES

- Objective Structured Clinical Examinations (OCSEs) have become widespread in the assessment of clinical competence.
- Authenticity is important in enhancing the validity of the test.
- Developing high-quality OSCE stations takes time and effort.
- OSCEs are often complex and costly to run, but they are a valid, fair, and reliable method of assessing clinical skills.
- OSCEs should be blueprinted to the learning outcomes of the curriculum.
- Training of simulated patients is essential for high levels of consistency.
- Future developments of OSCEs are likely to be in the area of acute medical simulations, which are difficult to test by other means.

Introduction

The reliable and valid assessment of clinical competence has become an increasingly important area of concern in medical education. Various stakeholders, with legitimate interests in the clinical competence of graduates from medical schools and of postgraduate trainees, require evidence that assessments are discriminating between the sufficiently and insufficiently competent at all levels of medical training and education.

While clinical competence is based on a thorough base of specialist medical knowledge [1], the term 'clinical competence' also encompasses other professional practice elements such as history taking and clinical examination skills, skills in practical procedures, doctor–patient communication, problem-solving ability and management skills, relationships with colleagues, and ethical behaviour [2, 3].

Unsuited to testing by written examination, the assessment of clinical competence has historically involved the direct observation of candidates by professional colleagues. With the development of work-based learning methods to assess clinical performance in a more authentic and naturalistic way, there is the potential for confusion over terminology. For the purposes of this chapter we will consider assessments of clinical competence to be measures of what doctors can do in controlled representations of professional practice, i.e. under examination conditions. We will restrict the use of performance assessment to measurements of what doctors do in their professional practice [4, 5]. We shall consider competency-based assessments to be those assessments undertaken *outside* the 'real' clinical environment and performance-based assessment to be those administered *within* the natural clinical setting.

A variety of formats for assessing clinical competence have been developed over the years and in this extended chapter we review the more 'classical' long case and short case formats, and also describe newer formats such as the OSCE and Objective Structured Long Case Examination Record (OSLER). We discuss the stages of planning and implementing OSCEs and offer practical advice on blueprinting, station development, examiner training, simulated/standardised patient training, organisational issues, and standard setting. We do not discuss workplace-based assessment instruments as these are covered comprehensively in Chapter 22 of this book.

The Long Case

In the traditional long case, candidates spend up to one hour with a patient, during which they are expected to take a full formal history and perform a complete examination. The candidate is not observed. On completion of this task the candidate is questioned for 20–30 minutes about the case, usually by a pair of examiners, and may be taken back to the patient to demonstrate clinical signs.

Holistic appraisal of the examinee's ability to assess and manage a real patient is a laudable goal of the long case. However, there are some shortcomings in using one or two long cases as a measure of clinical competence, related to

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

issues of reliability pertaining to examiner and patient factors [6]. The lack of measurement consistency caused by examiner bias and variations in examiner stringency is a major argument against the long case. Reliability is further compromised when there is little prior agreement between pairs of examiners as to what constitutes acceptable competence. Unstructured questioning and global marking without anchor statements compounds the problem. Reliability in the long case encounter is diminished by variability in degree and details of information disclosure by the patient, as well as variability in patients' demeanour, comfort, and health. Furthermore, some patients' illnesses may be straightforward, whereas others may be extremely complex. Examinees' clinical skills also vary significantly across tasks (i.e. task or case specificity) [7] so that assessing examinees on one patient will not provide generalisable estimates of their overall ability [8, 9].

While the authenticity of a long case examination is one of the strengths of the genre, inferring examinees' true clinical skills in the time-constrained environment of actual clinical practice from a one-hour long case encounter is debatable. Additionally, given the evidence of the importance of history taking in achieving a diagnosis [10] and the need for students to demonstrate good patient communication skills, the omission of direct observation in this process is a significant weakness.

Objective Structured Long Case Examination Record

In an effort to address these shortcomings, while at the same time attempting to retain the concept of seeing a 'new' patient in a holistic way, the Objective Structured Long Case Examination Record (OSLER) was developed by Gleeson in the 1990s [11].

The OSLER has 10 key features:
- it is a 10-item structured record
- it has a structured approach – there is prior agreement on what is to be examined
- all candidates are assessed on identical items
- construct validity is recognised and assessed
- history process and product are assessed
- communication skill assessment is emphasised
- case difficulty is identified by the examiner
- it can be used for both criterion- and norm-referenced assessments
- a descriptive mark profile is available where marks are used
- it is a practical assessment with no need for extra time over the ordinary long case.

The OSLER consists of 10 items, which include four on history, three on physical examination, and three on management and clinical acumen. For any individual item, examiners decide on their overall grade and mark for the candidate and then discuss this with their co-examiner and agree on a joint grade. This is done for each item and also for the overall grade and final agreed mark. The recommended time allocation for the OSLER examination is 30 minutes [12].

There is evidence that the OSLER is more reliable than the standard 'long case' [13]. Most recently, Wass et al. demonstrated that assessments using structured long cases could be highly reliable (predicted Cronbach's alpha of 0.84) [14] but this required 10 separate cases and 20 examiners – raising major issues of practicality.

Short Cases

In traditional tests of clinical competence, candidates undertook a series of (usually three to six) short cases. In this type of test, they were taken to a number of patients with widely differing conditions, and asked to examine individual systems or areas and give differential diagnoses of their findings, or to demonstrate abnormal clinical signs or produce spot diagnoses. Although in some ways similar to an OSCE in that they provided a wider range of cases on which the examiner was able to base his or her opinion of the student's ability, there are important differences. Different candidates rarely saw the same set of patients, cases often differed greatly in their complexity and the same two assessors examined the candidate at each case. These cases were not designed to test communication skills, but instead concentrated on clinical examination skills, with communication with the patients merely incidental. The examination was not structured and the examiners were free to ask any questions they wanted. Like the long case there was no attempt to standardise the expected level of performance. For all these reasons, OSCEs have superseded this type of assessment.

Objective Structured Clinical Examinations (OSCEs)

The remainder of this chapter relates to the OSCE, an assessment format in which the candidates rotate sequentially around a series of structured cases located in 'stations', at each of which specific tasks have to be performed. The tasks usually involve a clinical skill, such as history taking, examination of a patient, or a practical skill. The marking scheme for each station is structured and determined in advance. There is a time limit for each station, after which the candidates have to move on to the next task.

The basic structure of an OSCE may be varied in the timing for each station, the use of a checklist or rating scale for scoring, the use of a clinician or standardised patient as examiner, and the use of real patients or manikins, but the fundamental principle is that every candidate has to complete the same assignments in the same amount of time and is marked according to a structured marking schedule.

The terminology associated with the OSCE format can vary – in the undergraduate arena they are more consistently referred to as OSCEs, but in the postgraduate setting a variety of terminology exists. For example, in the UK, the Royal College of Physicians' membership clinical examination is called the Practical Assessment of Clinical Examination Skills (PACES), while the Royal College of General Practitioners' membership examination is called the Clinical Skills Assessment (CSA).

Rationale for the use of OSCEs

The use of OSCEs in the quantitative assessment of competence has become widespread in the field of undergraduate and postgraduate medical education [15–19]) since they were originally described [20], mainly due to the improved fairness and reliability of this assessment format. This has resulted in a fairer test of candidates' clinical abilities, since every candidate is presented with the same challenges in the test, and the scores have become less dependent on who is examining the candidate and which patient is selected for the encounter.

The contemporary view of validity [21, 22] sets out criteria which should be used to evaluate any assessment method. The most applicable is the Kane 'Validity Framework and Argument' [23]; the components of this framework are threefold:

1 a clear statement of intended use or purpose for the assessment
2 meaningful evidence to support and justify the inferences and decisions made on the basis of the outcomes of the assessment
3 the 'argument' or justification for defending the decisions made.

Note that it is inferences based on the test results that are validated, not the test itself [24, 25].

1: Intended use of the OSCE: What is your assessment intended to achieve?

This section requires a clearly defined purpose and intention: for an OSCE, it should be to test clinical and communication skills. It requires demonstration that the OSCE is appropriately designed and delivered for its intended purpose, and that this intended use applies across all stages of the assessment process, from design and delivery to analysis of the data.

2: Meaningful evidence to answer the question: Does the assessment measure what it is intended to measure?

- How you decide on and plan the content of the test = **Content**
- How you construct the test and demonstrate appropriate psychometric properties = **Internal structure**
- How you safeguard fair administration of test and accuracy of scores = **Response process**
- How you decide on the outcomes = **Consequences**

Content

How you decide on and plan the content of the test

The content of an OSCE should adequately sample skills which match the learning objectives of the course for which that OSCE is designed [26]. The sampling should be representative of the whole testable domain for that examination purpose. The best way to ensure an adequate spread of sampling is to use a blueprint method, which we will describe later in the chapter. Inferences about ability to apply clinical knowledge to bedside data gathering and reasoning, and to effectively use interpersonal skills, are most relevant to the OSCE model. Inferences about knowledge, rather than clinically relevant application of knowledge, or clinical and practical skills, are less well supported by this method [27].

Internal Structure

How you construct the test and demonstrate appropriate psychometric properties

An OSCE should demonstrate that there is adequate sampling, stations have been constructed appropriately and are of suitable duration, and the scoring rubric has a clear rationale.

One aspect of enhancing the validity of inferences from an OSCE is that the length of any station should be best fitted to the task to achieve the best authenticity possible. Thus, for example, a station in which blood pressure measurement is tested would authentically be achieved in 5 minutes, whereas taking a history of chest pain or examining the neurological status of a patient's legs would be more authentically achievable in 10 minutes [17, 28, 29].

The psychometric properties of the OSCE should be analysed and include overall examination statistics as well as station-level statistics. The most critical overall analysis is reliability, as this indicates the quality of the test to consistently distinguish between those candidates who are fit to pass and those who are not [30].

Essentially, the OSCE was developed to address the inherent unreliability of classical long and short cases [31–40]. OSCEs are more reliable than unstructured observations in four main ways:

- Structured marking schedules allow for more consistent scoring by examiners according to predetermined criteria; hence reliability is improved.
- Candidates have to perform a number of different tasks across clinical, practical, and communication skill domains – this wider sampling across different cases and skills results in a more reliable picture of a candidate's overall competence. The more stations or cases each candidate has to complete, the more generalisable the test is.
- The reliability of the total test score increases with increasing number and increasing homogeneity of stations or cases. Reliability of sub-scores must be carefully reviewed before reporting.
- As the candidates move through all the stations, each is examined by a number of different examiners, so multiple independent observations are collated. Individual examiner bias is thus attenuated.

It is worth bearing in mind that sampling across different cases makes the most important contribution to reliability; the more stations in an OSCE, the more reliable it will be. However, increasing the number of stations has to be balanced with the practicability of an OSCE exercise. Practically, to enhance reliability it is better to have more stations with one examiner per station than fewer stations with two examiners per station.

Response Process

How you safeguard fair administration of test and accuracy of scores

Safeguards for the fair administration of an OSCE and the accuracy of reported scores are a vital part of ensuring that candidates are tested fairly and that the correct scores are documented and reported. With modern technology, OSCEs can be delivered electronically on iPads, with examiners scoring directly as they observe and with real-time calculation of scores. If OSCEs are scored on paper, there must be safeguards in place to ensure candidate scores are accurately and correctly recorded and transferred to the calculations.

Consequences: How you Decide on the Outcomes

Any summative outcomes and decisions made on the basis of OSCE scoring should be clearly reported and the pass/fail decisions should be robust, fair, and defensible.

Standard Setting

Standard setting or establishing the pass mark is critical for determining who passes and who fails any particular assessment of clinical competence. It is an essential part of the validity evidence which should be gathered to support the decisions made on the basis of an OSCE. The standard or pass mark indicates the minimum score that every candidate has to reach to pass the OSCE. While it is difficult to quantify a concept as complex as clinical competence, the reality is that examinations such as OSCEs are used to discriminate between those who have sufficient clinical skills and those who do not, for a particular level or purpose.

The fundamental principle underlying all standard setting methods is to reach a consensus on professional values and standards [41]. There are many standard setting methods described in the literature [42–44] but many of the traditional ones were developed for multiple-choice questions. It is debatable whether it is appropriate for these methods to be used for complex performance-based examinations such as OSCEs.

As experience with OSCEs has evolved, it is the Borderline Group methods which have become the favoured method of standard setting for OSCEs [45]. It does require some expertise in processing the data, and is more reliable if the examiners are trained [46], but overall it has become regarded as the 'gold standard' for OSCEs [47]. For a full discussion of standard setting methods see Chapter 24.

Educational Impact

The impact on students' learning resulting from a testing process is sometimes referred to as consequential validity. The design of an assessment system can reinforce or undermine learning [48], it is a well-recognised phenomenon that students focus on their assessments rather than the learning objectives of the course. Explicit, clear learning objectives allied with clinical skills assessment content and format can be a very effective way of encouraging students to learn the desired clinical competencies. Objectives that include action verbs like 'demonstrate' or 'perform', which are then linked to OSCEs that measure ability to demonstrate or perform certain skills, will encourage students to practise these skills. By contrast, an assessment system that measures students' ability to answer multiple-choice questions about clinical skills will encourage students to focus on knowledge acquisition. Neither approach is wrong – they simply demonstrate that assessment drives education and that assessment methods need to be thoughtfully applied. There is a danger in using detailed checklists as this may encourage students to memorise the steps in a checklist rather than learn and practise the skill. Rating scale marking schedules encourage students to learn and practise skills more holistically [49].

OSCEs may be used for formative or summative assessment. When teaching and improvement are a major goal of an OSCE, time should be built into the schedule at the end of each station to allow the examiner to give feedback to the student on their performance, providing a very powerful opportunity for student learning [50]. For summative certification examinations, expected competencies should be clearly communicated to the candidates so they have the opportunity to learn the skills prior to taking such examinations.

The increased reliability of the OSCE format over other formats of clinical testing and its perceived fairness by candidates has helped to engender the widespread acceptability of OSCEs among test takers and testing bodies [51]. Consideration should be given to the impact failing a test can have on the candidates, as well as passing.

3: The 'argument' or justification for defending the decisions made

Since Harden's original description in 1979, the use of OSCEs has become widespread in the undergraduate level of testing of clinical competence as well as increasingly in postgraduate assessment. More recently, OSCEs have been used to replace traditional interviews in recruitment processes in both undergraduate and postgraduate settings. For example, for recruitment to general practice training schemes in the UK, candidates go through an OSCE format of scenarios in assessment centres where different exercises are assessed by trained assessors, who observe various job-related competencies, including communication skills, team involvement, and problem-solving ability.

In North America, clinical skills assessment has been accepted on a massive scale. In 1992, the Medical Council of Canada (MCC) added a standardised patient component to its national licensing examination because of the perception that important competencies expected of licensed physicians were not being assessed [45]. Since inception, approximately 2500 candidates per year have been tested at multiple sites at fixed periods of time during the year throughout Canada. The MCC clinical skills examination uses physicians at each station to score the encounter. See Box 23.1 for a detailed description of the historical development of clinical skills assessment in the US.

Blueprinting

For any particular OSCE, the content (i.e. the clinical tasks chosen for the stations) should map onto the learning objectives of the course and the candidates' level of learning. It is only reasonable to test candidates on what they have been taught.

BOX 23.1 FOCUS ON: Development of clinical skills assessment in the US

In the US, the Educational Commission for Foreign Medical Graduates (ECFMG) instituted a performance-based examination in 1998 to assess bedside data gathering, clinical reasoning, interpersonal skills, and spoken English communication skills of foreign medical graduates seeking to enter residency training programmes. From 1998 to 2004, when it was incorporated into the United States Medical Licensing Examination (USMLE), there were 43 642 administrations, including 37 930 first-time takers, making it at the time, the largest high-stakes clinical skills examination in the world. The 11 scored encounters had a standardised format, with each requiring the candidate to elicit a medical history, perform a physical examination, communicate in spoken English with a patient in a clinical setting, and generate a written record of the encounter. In each station, the candidate encountered a unique standardised patient – a lay person recruited and trained to give a realistic portrayal of a patient with a standardised medical and psychosocial history, and standardised findings on physical examination. Each case had a case-specific checklist containing the elements of medical history and physical examination considered pertinent to that particular case. Simulated patients were trained to recognise appropriate queries and/or physical examination manoeuvres, including acceptable equivalents or variants, and to document each checklist item achieved by the candidate. Simulated patients also evaluated each candidate's interpersonal skills and spoken English proficiency. After each encounter, the candidate generated a patient note on which the pertinent positive and negative elements of history and physical examination were recorded, a differential diagnosis constructed, and a diagnostic work-up plan proposed. Performance was evaluated by averaging scores across all encounters and determining the mean for the integrated clinical encounter (data gathering combined with the patient note score) and communication (interpersonal skills and spoken English). Generalisability coefficients for the two conjunctively scored components of CSA were approximately 0.70–0.90.

In 2004, the USMLE adopted the ECFMG clinical skills assessment model and began testing all US medical graduates in addition to foreign medical graduates seeking ECFMG certification. Additional computer and standardised patient training infrastructure was included to ensure comparability across all centres.

The USMLE Step 2 (Clinical Skills) uses 12 standardised patient encounters, each 15 minutes in length, followed by 10 minutes to write a patient note. As in the ECFMG CSA examination, standardised patients document the items asked in the history and performed in the physical examination to specified criteria, and evaluate interpersonal skills and spoken English skills, while physician raters score the patient note. Approximately 35 000 administrations take place each year.

To map the assessment to the learning objectives, the categories of skill to be tested should be mapped on one axis and the elements of the course being tested should be mapped on the other. Usually in OSCEs, the skills domains are categorised into clinical examination skills, practical skills, and communication skills, which can be further subgrouped into history-taking skills and other doctor–patient/colleague interactions. The subject content of the OSCE will be determined to a certain extent by how the elements of the course are categorised, that is, by subject discipline or systems.

Blueprinting is a powerful tool that helps to focus the OSCE designers on the exact nature of what they wish to test and relate this to the teaching. Once this blueprint or framework for an OSCE is agreed, the individual stations can be planned and classified according to this blueprint. This ensures adequate sampling across subject areas and skills, in terms of numbers of stations covering each skill and the spread over the subjects/systems of the course being tested.

The feasibility of testing a particular task also needs to be considered. Real patients with clinical signs can be used to test clinical examination skills, while simulated patients are best for testing communication skills. Simulated patients can also simulate a number of clinical signs (e.g. loss of visual field, localised abdominal pain). Healthy volunteers can be used when testing the technical process of a clinical examination. There are many manikins on the market for testing invasive practical skills, e.g. intravenous cannulation, urethral catheterisation, and arterial blood gas sampling.

It is essential to use a blueprint to plan the content of an OSCE as this helps to ensure that different domains of skill are tested equitably and that the balance of subject areas tested is fairly decided. An example is provided in Box 23.2.

Station Development

It is important to write out station specifications well in advance of the examination date so the stations can be reviewed and trialled prior to the actual assessment. Sometimes stations that seem like a good idea at the time of writing may turn out to be unfeasible in practice. When writing a station specification, the following aspects should be considered:

- Construct: a statement of what that station is supposedly testing, e.g. this station tests the candidate's ability to examine the peripheral vascular system.
- Clear instructions for the candidate: to inform the candidate exactly what task they should perform at that station.
- Clear instructions for the examiners: including a copy of the candidate instructions, to assist the examiner at that station to understand his or her role and conduct the station properly.
- List of equipment required.
- Personnel requirements: whether the station requires a real patient or a simulated patient and the details of such individuals (e.g. age, gender, ethnicity).

BOX 23.2 Example of a system-based blueprint

	History	Explanation	Examination	Procedures
Cardiovascular	Chest pain	Discharge drugs	Cardiac	BP
Respiratory	Haemoptysis		Respiratory	Peak flow
Gastrointestinal	Abdominal pain	Gastroscopy	Abdominal	PR
Reproductive	Amenorrhoea	Abnormal smear	Cervical smear	
Nervous	Headache		Eyes	Ophthalmoscopy
Musculoskeletal	Backache		Hip	
Generic	Pre-op assessment	Consent for post-mortem		IV cannulation

- Simulated patient scenario: if the station requires a particular role to be played.
- Marking schedule: this should include the important aspects of the skill being tested, a marking scheme for each item, and how long the station should last. The marking schedule may be either a checklist or a rating scale as there is good evidence that, despite the apparent objectivity of structured checklists, global rating scales have been shown to be equally as reliable (see Box 23.2). Items can be grouped into the broad categories of process skills, content skills, and clinical management skills.
  - *Process skills*: For clinical examination stations with a real or simulated patient, these could include introduction and orientation, rapport, professional manner, and communicating with the patient appropriately during examination. For history-taking stations, these could include introduction and orientation, listening skills, questioning skills, demonstration of empathy, and appropriate closure. For explanation stations, these could include introduction and orientation, rapport, establishing what the patient knows/understands, demonstration of empathy, appropriate organisation of explanation, checking the patient's understanding, using clear language, and avoiding jargon.
  - *Content skills*: These include appropriate technical steps or aspects of the task or skill being tested.
  - *Clinical management skills*: It may be appropriate to ask the candidate some set questions in relation to the specific case.
- Figures 23.1 and 23.2 provide examples illustrating the checklist and rating scale marking schedules, respectively.

Piloting

Ideally, stations should be piloted before they are used in examinations to ensure that all stations are functional in terms of the following:

- Timing: can the candidates realistically perform the task in the time allotted?
- Difficulty: how difficult is the station?
- Equipment: is all the equipment required available and on the list?
- Is an additional helper required to assist the examiner, e.g. for catherisation, suturing stations?
- Candidate instructions: do the instructions tell the candidate exactly what the task is?
- Examiner instructions: do the instructions tell the examiner how to conduct the station? Does the examiner know what the candidate has been told to do?
- Real patient specifications: are the medical conditions specified?
- Simulated patient scenario: is the age/gender/ethnicity specified? Is there enough information for the simulated patient to learn and play the part effectively?
- Construct validity: is the station testing what it is meant to test? Does the marking schedule reflect the elements of the task appropriately?

Simulated Patient Training

For consistent performances, particularly at communication skills stations, it is best to use well-trained simulated patients. Depending on location, it may be possible to organise a database of actors who assist in the teaching as well as assessment of communication skills. It is desirable to have people across a range of ages and ethnicities, as well as a balanced gender mix. Training and monitoring simulated patients is essential to ensure consistent performance – a significant factor in the reliability of the examination. The simulated patients should be sent their scenarios in advance and then asked to go through their roles with other simulated patients playing the same role, while being supervised by a communication skills teacher and/or a clinician, to develop the role to a suitable standard.

Examiner Training

OSCEs require large numbers of examiners. This can be a strength, as candidates are observed and scored by clinicians, but it is also one of the potential weaknesses of OCSEs, as inconsistency between examiners will reduce fairness and reliability.

Considerable resources are devoted to examiner training. Structured face-to-face training sessions are good for introducing new examiners to OSCEs and scoring processes. The programme for these events is interactive and very much acknowledges the inherent expertise that experienced clinicians bring to the assessment process. These training sessions cover:

- principles of OSCEs
- role of examiners (i.e. to assess not to teach; to conduct vivas, adhere to marking schedules, and respect the role of the simulated patient)

THIS IS A 10-MINUTE STATION	Good	Adequate	Not done/ Inadequate
1 Introduction and orientation (name and role, explains purpose of examination, confirms patient's agreement)	[ ]	[ ]	[ ]
2 Rapport (shows interest, respect and concern, appropriate body language)	[ ]	[ ]	[ ]
3 Appropriately exposes the patient and positions them at 45 degrees		[ ]	[ ]
4 Looks at hands, commenting on peripheral stigmata (i.e. cyanosis, clubbing, splinter haemorrhages, etc.)		[ ]	[ ]
5 Checks the radial pulse, commenting on the rate and rhythm		[ ]	[ ]
6 Asks for patient's blood pressure ***Examiner please give correct BP***		[ ]	[ ]
7 Looks for central stigmata of cardiovascular disease (i.e. anaemia, central cyanosis, hyperlipidaemia)		[ ]	[ ]
8 Examines the JVP correctly (positions the patient's chin and neck; assesses the waveform in the correct area) and comments on findings		[ ]	[ ]
9 Palpates the carotid or brachial pulses, commenting on the character		[ ]	[ ]
Inspects and palpates the precordium: 10 Localises the apex beat, commenting on the position		[ ]	[ ]
11 Examines for RVH		[ ]	[ ]
12 Auscultates in all four cardiac areas		[ ]	[ ]
13 Moves patient to left side and sits patient forward in expiration	[ ]	[ ]	[ ]
14 Comments on heart sounds and times heart sounds against central pulse	[ ]	[ ]	[ ]
15 Comments on any murmurs		[ ]	[ ]
16 Listens to the lung bases		[ ]	[ ]
17 Candidate attempts to assess peripheral pulses ***Examiner please stop candidate***		[ ]	[ ]
18 Checks for ankle/sacral oedema		[ ]	[ ]
19 Presents a brief summary and conclusions	[ ]	[ ]	[ ]
20 Communicates with patient appropriately during examination (explains what they are doing, gains patient's co-operation)	[ ]	[ ]	[ ]
21 Examines patient in a professional manner (gentle, watches for pain, maintains dignity and privacy)	[ ]	[ ]	[ ]
22 Closure (thanks patient, leaves patient comfortable)	[ ]	[ ]	[ ]
23 Candidate cleans hands after examination		[ ]	[ ]

Figure 23.1 Example checklist mark sheet for cardiovascular examination.

- marking video-recorded OSCE stations, followed by assessment with the clinicians of their marking and getting them to think through their mark allocation
- marking 'live stations' with group members playing the candidate, the assessor, and the simulated patient. This demonstrates how stressful this assessment is for the candidate and how difficult it can be to play the part of a good simulated patient
- standard setting procedure used.

The standard setting procedures can be crucial when using a student-centred approach, and all the examiners are integral to the standard setting process. The more the

	Clear fail	Borderline fail	Borderline pass	Clear pass	Excellent
1 Physical examination: inspection, pulses, JVP, carotids, palpation of precordium, auscultation of valve areas	☐	☐	☐	☐	☐
2 Identification and interpretation of physical signs: identifies and interprets signs correctly; makes reasonable diagnosis	☐	☐	☐	☐	☐
3 Management: suggests appropriate investigations, treatment and follow-up	☐	☐	☐	☐	☐

Figure 23.2 Example rating scale mark sheet for cardiovascular examination.

BOX 23.3 FOCUS ON: Simulated patients as assessors

Scoring standardised patient examinations can be done by third-party observers (usually clinicians) or by the standardised patients themselves. Clinician examiners enhance the validity of the assessment because they can apply holistic judgements and integrate subdomains of sequence, logic, and other factors that may be difficult for a non-professional completing a binary checklist to capture. Boulet et al. [38], however, demonstrated that holistic judgements from clinician examiners are similar to aggregate scores from trained standardised patients, at least in assessing a general, entry-level clinician. Clinician evaluator models using holistic scoring models may have greater utility in capturing higher levels of expertise – something that checklist models may not be able to do. Any examiner, whether simulated patient or clinician, must be thoroughly trained and then monitored to ensure consistent use of the score scale, since variability diminishes reliability [46].

Ratings of interpersonal and communication skills provide a unique challenge in determining who is best able to provide the ratings. Although the assessment of doctor–patient communication skills can be accomplished by a clinician or other observers, and can be done 'live' or via video-taped reviews, it is unclear whether someone watching the interplay between a doctor and a patient can adequately measure the complex, multi-dimensional nature of the communication. Many aspects of this communication, especially those that are non-verbal, may be best assessed by the patient or the person trained to be the patient.

Spoken English is another domain that might be better scored by non-clinicians. The generalisability coefficient of this component of the Educational Commission for Foreign Medical Graduates Clinical Skills Assessment (CSA), scored by standardised patients, was 0.94.

From a logistical and cost perspective, the examinee volume for the CSA and now the United States Medical Licensing Examination Step 2 (CS) (approximately 35 000 per year) makes it effectively impossible to entertain using physician examiners. Cost analysis also needs to account for training time and quality assurance for standardised patients or physician raters as well as the different nature of the training needed by each group. It may also be harder to standardise a large number of highly educated, typically independently thinking physicians across five test centres in a year-round testing model.

Currently in the UK, at both the undergraduate and postgraduate levels, examiners are clinicians or other health care professionals.

assessors understand their vital role in this process the more likely they are to do it in a satisfactory way. The use of non-clinicians in assessment is discussed in Box 23.3.

Once examiners have had initial training, it may be helpful to refresh examiners' scoring and standards via interactive online courses, with videos of candidate performance and feedback on examiner scoring.

Working with Real Patients

Patients do not always give the same history each time they are asked to repeat it; they can become tired or unwell and they may develop new signs and symptoms to the ones they originally reported; they may even lose previous clinical findings. However, they can be a most valuable resource and need to be treated as such. Using 'real' patients in OSCEs adds greatly to the validity of the assessment. Ideally, patients should be used to assess the detection of common chronic clinical signs. For each clinical sign assessed several patients will be needed and even the most stoical patient should not be expected to be examined by more than 10 students in the course of a day. Ideally, patients should be swapped in and out of the station to allow them to have sufficient rest time. This is best done by central coordination. See Box 23.4 for details about running OSCEs.

Box 23.4 Operational features

Prior to the OSCE

Recruitment of simulated patients	Once the OSCE has been blueprinted, the simulated patients required should be listed and actors contacted to engage them for the dates of the exam.
Running order of the stations	Stations should be numbered to avoid confusion over mark sheets, equipment, and people involved. Rest stations should be provided: usually one rest per 40 minutes in a circuit is suitable. If many candidates are sitting the OSCE, running multiple circuits of the same stations enables more candidates to be examined at any one time.
Using stations of different lengths	It is best to group stations of the same length together and to run these circuits separately. If there are 5-, 10-, and 15-minute stations, then the candidates should be asked to attend on three separate occasions to undertake each circuit. Mixing stations of different timings in one circuit is possible, but can lead to confusion.
List of all the equipment required	Detailed by station, this is vital for the preparations to be successful. Arrange to go round the circuit the day before the OSCE and to check that all the equipment is correctly set up.
Production and processing of mark sheets	Calculate the numbers required for each station and allow extra for spoilage. Allow time for proofreading. If there are a large number of candidates, it may be worth looking into using sheets that can be processed by electronic scanning after the OSCE. Alternatively, marking by hand will require the organisation of people to mark and ensure that results are entered correctly. Computer systems for automated collection and analysis of station data may be purchased or developed. If a computer system is used, a paper backup should always be available in case the network goes down.
Liaison with clinical skills centre staff	Close cooperation with clinical skills centre technical and teaching staff is vital in the planning. It is useful to draw up a circuit plan to indicate the layout required and for the numbering of the stations to be agreed (see Figure 23.3).

Day of the OSCE

Signs	It is very helpful to put up signs indicating the rooms for the candidates, patients, and the examination, so that people unfamiliar with the venue can find their way easily. Large signs should be used to number all the stations to help candidates follow the circuit successfully.
Timing	An electronic timing programme is the ideal, but a reliable stopwatch and loud manual bell is an acceptable alternative. It is important to ensure that all candidates and examiners can hear the bell so the candidates move onto the next station promptly.
Helpers/Marshalls	A vital part of the smooth running of OSCEs depends on having a small army of helpers to direct the candidates, examiners, simulated patients, and patients to ensure everyone is in the right place at the right time. This should include looking after the welfare of all the people involved on the day.
Catering	Examining, acting, being examined, and helping at OSCEs can be tiring and sometimes stressful work. The very least one can do is provide refreshments for all participants – water for the candidates at rest stations, drinks for all other staff, and lunch for those who spend the whole day assisting or being examined.
Briefing	It is helpful to gather all candidates in a room where they can be registered and briefed about the practical arrangements for the day. Examiners, even if they have attended a training session, should be reminded about how to score the mark sheets and conduct the stations appropriately, and also to switch off their mobile phones.

After the OSCE

Collection of mark sheets	Collection should be organised meticulously, as missing sheets can be very prejudicial to a candidate's overall score. It is also helpful to check the sheets for completeness of scoring and to ask examiners to check they have completed the sheets before leaving.
Care of patients/ simulated patients	A system to ensure that patients have transport to take them home is always appreciated. Arrangements to ensure the simulated patients are paid are also welcome and encourage future participation.
Thank-you letters	Patients, examiners, and helpers are much more likely to take part again if they receive acknowledgement of their contribution to the examination process.

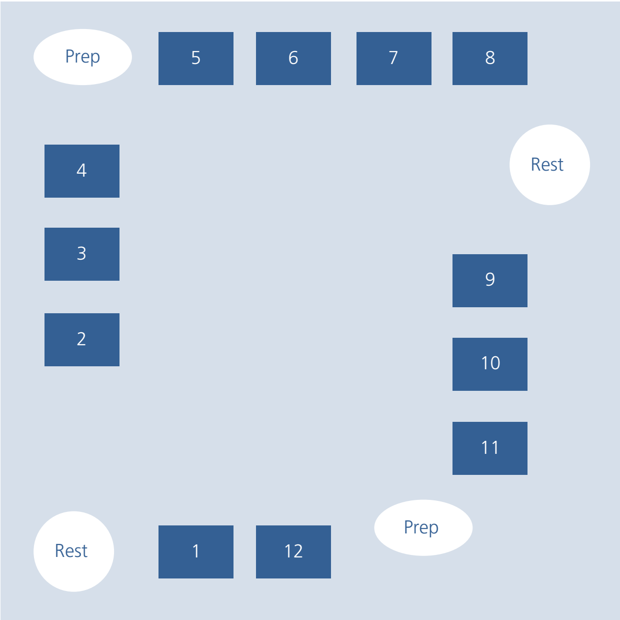


Figure 23.3 Example of a floor plan for an Objective Structured Clinical Examination in a large room.

As workplace-based assessment becomes more widespread, the focus of testing in OSCEs may be directed towards earlier stages of clinical skill acquisition and technical proficiency. Sequential OSCE design may become more common in an effort to control costs and improve efficiency (see Box 23.5). Another development might be that OSCEs become more specialised, possibly focusing on acute clinical scenarios, which would be difficult to assess reliably in vivo. Complex high-fidelity team-working scenarios are being developed in some areas and may become more appropriate as interprofessional training in the postgraduate arena becomes more common.

BOX 23.5 FOCUS ON: Sequential testing in OSCE administration

The majority of health care institutions use models of test–short-term remediation–retest for underperforming students. Whilst these models are typically associated with short-term improvement in candidate performance at retest, such models are costly to deliver (particularly for performance retest with OSCEs). There is increasing evidence that these traditional models are associated with longitudinal underperformance of candidates. Recent work has shown that whilst students often pass the OSCE retest, in the longer term those who have failed continue to perform weakly compared to their peers within a programme of assessment, and often deteriorate [30]. Rather than a traditional OSCE model, sequential testing involves a smaller 'screening' test format, with a further 'sequential' test for candidates who fail to meet the standards of the screening test. Overall pass/fail decisions are then made on the full sequence of tests. This can be used effectively to improve the efficiency of performance assessments like OSCEs that can be very expensive to run.

Conclusions

Assessment of clinical competence is a crucial part of the basis on which decisions are made about the ability of clinicians and doctors in training. But any method of assessing clinical skills should be considered in the context of a wider programme of assessment, which should include the assessment of knowledge, clinical examination skills, practical procedure skills, doctor–patient communication, problem-solving ability, management skills and relationships with colleagues, as well as professional attitudes and behaviour.

One of the most important aspects of assessing clinical skills is the range of sampling across a candidate's skill base; this has to be taken into account when designing any assessment. OSCEs can assess clinical, communication, and practical skills but are still situated in the context of an examination setting. To assess doctors in the context of their professional practice requires the use of different formats in the workplace.

Acknowledgement

Some of the material used in this section has previously been published in an article by one of the authors in *The Clinical Teacher* [52] and is reproduced here with permission.

References

1 Miller, G. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine* 65 (s9): S63–S67.
2 General Medical Council (2003). *Tomorrow's Doctors*. London: GMC.
3 General Medical Council (2013). *Good Medical Practice*. London: GMC.
4 Rethans, J.J., Norcini, J.J., Baron-Maldonado, M. et al. (2002). The relationship between competence and performance: implications for assessing practice performance. *Medical Education* 36: 901–909.
5 Swanson, D.B., Norman, G.R., and Linn, R.L. (1995). Performance based assessment: lessons learned from health professions. *Education Research* 24 (5): 5–11.
6 Norcini, J.J. (2001). The validity of long cases. *Medical Education* 35: 720–721.
7 Norcini, J.J. (2002). The death of the long case? *British Medical Journal* 324: 408–409.
8 Norcini, J.J. (2001). Does observation add to the validity of the long case? [letter]. *Medical Education* 35: 1131–1133.
9 Hampton, J.R., Harrison, M.J.G., Mitchell, J.R.A. et al. (1975). Relative contributions of history taking, physical examination and laboratory investigations to diagnosis and management of medical outpatients. *British Medical Journal* 2: 486–489.
10 Gleeson, F.A. (1997). Assessment of clinical competence using the Objective Structured Long Examination Record (OSLER). *Medical Teacher* 19: 7–14.
11 Gleeson, F.A. (1994). The effects of immediate feedback on clinical skills using the OSLER. *Proceedings of the Sixth Ottawa Conference in Medical Education*, pp. 412–415.

12 van Thiel, J., Kraan, H.F., and van der Vleuten, C.P.M. (1991). Reliability and feasibility of measuring medical interviewing skills. The revised Maastricht history and advice checklist. *Medical Education* 25: 224–229.

13 Wass, V., Jones, R., and van der Vleuten, C.P.M. (2001). Standardized or real patients to test clinical competence? The long case revisited. *Medical Education* 35: 321–325.

14 Sloan, D.A., Donnelly, M.B., Schwartz, R.W., and Strodel, W.E. (1995). The Objective Structured Clinical Examination. The new gold standard for evaluating postgraduate clinical performance. *Annals of Surgery* 222: 735–742.

15 Cohen, R., Reznick, R.K., Taylor, B.R. et al. (1990). Reliability and validity of the objective structured clinical examination in assessing surgical residents. *American Journal of Surgery* 160: 302–305.

16 Davis, M.H. (2003). OSCE: the Dundee experience. *Medical Teacher* 25: 255–261.

17 Newble, D. (2004). Techniques for measuring clinical competence: objective structured clinical examinations. *Medical Education* 38: 199–203.

18 Harden, R.M. and Gleeson, F.A. (1979). Assessment of clinical competence using an Objective Structured Clinical Examination (OSCE). *Medical Education* 13: 41–54.

19 van der Vleuten, C.P.M. (1996). The assessment of professional competence: developments, research and practical implications. *Advances in Health Science Education* 1: 41–67.

20 van der Vleuten, C.P.M. and Swanson, D.B. (1990). Assessment of clinical skills with standardized patients: state of the art. *Teaching and Learning in Medicine* 2: 58–76.

21 Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher* 23 (2): 13–23.

22 Downing, S.M. and Haladyna, T.M. (2004). Validity threats: overcoming interference with proposed interpretations of assessment data. *Medical Education* 38: 327–333.

23 Downing, S.M. (2003). Validity: on the meaningful interpretation of assessment data. *Medical Education* 37: 830–837.

24 Petrusa, E. (2002). Clinical performance assessments. In: *International Handbook of Research in Medical Education* (ed. G.R. Norman, C.P.M. van der Vleuten and D.I. Newble), 673–709. Boston and London: Kluwer, Dordrecht.

25 Kaufman, D.M. (2003). ABC of learning and teaching in medicine: applying educational theory in practice. *British Medical Journal* 326: 213–216.

26 Vargas, A.L., Boulet, J.R., Errichetti, A.M. et al. (2007). Developing performance-based medical school assessment programs in resource-limited environments. *Medical Teacher* 29 (2–3): 192–198.

27 Adamo, G. (2003). Simulated and standardized patients in OSCEs: achievements and challenges 1992–2003. *Medical Teacher* 25: 262–270.

28 Townsend, A.H., Mcllvenny, S., Miller, C.J., and Dunn, E.V. (2001). The use of an Objective Structured Clinical Examination (OSCE) for formative and summative assessment in a general practice clinical attachment and its relationship to final medical school examination performance. *Medical Education* 35: 841–846.

29 Hodges, B. (2003). Validity and the OSCE. *Medical Teacher* 25: 250–254.

30 Pell, G., Fuller, R., Homer, M., and Roberts, T. (2013). Advancing the objective structured clinical examination: sequential testing theory and practice in action. *Medical Education* 47: 569–577.

31 Mavis, B.E. and Henry, R.C. (2002). Between a rock and a hard place: finding a place for the OSCE in medical education. *Medical Education* 36: 408–409.

32 Rymer, A.T. (2001). The new MRCOG Objective Structured Clinical Examination – the examiners' evaluation. *Journal of Obstetrics and Gynaecology* 21: 103–106.

33 Eva, K.W., Rosenfeld, J., Reiter, H.I., and Norman, G.R. (2004). An admissions OSCE: the multiple mini-interview. *Medical Education* 38: 314–326.

34 Lane, P. (2005). Recruitment into training for general practice – the winds of change or a breath of fresh air? *British Medical Journal Career Focus* 331: 153.

35 Whelan, G.P., Boulet, J.R., McKinley, D.W. et al. (2005). Scoring standardized patient examinations: lessons learned from the development and administration of the ECFMG Clinical Skills Assessment (CSA). *Medical Teacher* 27: 200–206.

36 Boulet, J.R., Ben David, M.F., Ziv, A. et al. (1998). High-stakes examinations: what do we know about measurement? Using standardized patients to assess the interpersonal skills of physicians. *Academic Medicine* 73 (10): S94–S96.

37 Whelan, G.P., McKinley, D.W., Boulet, J.R. et al. (2001). Validation of the doctor–patient communication component of the Educational Commission for Foreign Medical Graduates Clinical Skills Assessment. *Medical Education* 35: 757–761.

38 Boulet, J.R., McKinley, D.W., Norcini, J.J., and Whelan, G.P. (2002). Assessing the comparability of standardized patient and physician evaluations of clinical skills. *Advances in Health Sciences Education: Theory and Practice* 7: 85–97.

39 Noel, G.L., Herbers, J.E. Jr., Caplow, M.P. et al. (1992). How well do internal medicine faculty members evaluate the clinical skills of residents? *Annals of Internal Medicine* 117: 757–765.

40 Boulet, J.R., van Zanten, M., McKinley, D.W., and Gary, N.E. (2001). Evaluating the spoken English proficiency of graduates of foreign medical schools. *Medical Education* 35: 767–773.

41 Norcini, J.J. (2003). Setting standards on educational tests. *Medical Education* 37: 464–469.

42 Zeiky, M.J. (2001). So much has changed. How the setting of cutscores has evolved since the 1980s. In: *Setting Performance Standards: Concepts, Methods, and Perspectives* (ed. G.J. Cizek), 19–52. Mahwah, NJ: Lawrence Erlbaum Associates.

43 Cizek, G.J. (2001). *Setting Performance Standards: Concepts, Methods, and Perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates.

44 Angoff, W.H. (1971). Scales, norms and equivalent scores. In: *Educational Measurement* (ed. R.L. Thorndike), 508–600. Washington, DC: American Council on Education.

45 Dauphinee, W.D., Blackmore, D.E., Smee, S.M. et al. (1997). Using the judgements of physician examiners in setting standards for a national multicentre high stakes OSCE. *Advances in Health Science Education* 2: 201–211.

46 Hodges, B. (2003). Analytic global OSCE ratings are sensitive to level of training. *Medical Education* 37: 1012–1016.

47 Smee, S.M. and Blackmore, D.E. (2001). Setting standards for an objective structured clinical examination: the borderline group method gains ground on Angoff. *Medical Education* 35: 1009–1010.

48 Pell, G., Fuller, R., Homer, M., and Roberts, T. (2012). Is short-term remediation after OSCE failure sustained? A retrospective analysis of the longitudinal attainment of underperforming students in OSCE assessments. *Medical Teacher* 34: 146–150.

49 Pugh, D., Halman, S., Desjardins, I. et al. (2016). Done or almost done? Improving OSCE checklists to better capture performance in progress tests. *Teaching and Learning in Medicine* 28 (4): 406–414.

50 Hattie, J. and Timperley, H. (2007). The power of feedback. *Review of Educational Research* 77 (1): 81–112.

51 Swanson, D.B. and Roberts, T.E. (2016). Trends in national licensing examinations in medicine. *Medical Education* 50: 101–114.

52 Boursicot, K. and Roberts, T. (2005). How to set up an OSCE. *The Clinical Teacher* 2: 16–20.

Further Reading

Boursicot, K.L., Etheridge, S.Z., Sturrock, A. et al. (2011). Performance in assessment: consensus statement and recommendations from the Ottawa conference. *Medical Teacher* 33 (5): 370–383.

24 Standard Setting Methods in Medical Education: High-stakes Assessment

André F. De Champlain
Research and Development, Medical Council of Canada, Ottawa, Ontario, Canada

KEY MESSAGES

- Standard setting is a crucial activity for any assessment programme that must render a judgement as to the competency of candidates, whether at the school level or for licensure and certification purposes.
- There is no such thing as a 'gold standard' in determining a cut-score value for a test. Selecting and implementing a rigorous process by which a cut-score value can be arrived at, with appropriate supporting documentation and empirical evidence, is what needs to be defended.
- For the vast majority of medical education assessments, criterion-referenced methods are more appropriate than their norm-referenced counterparts, as the former are based on expert judgement of what constitutes minimal competency. Norm-referenced standards are defensible for selection decisions only.
- For multiple-choice examinations (MCQs), test-centred standard setting methods, such as the Angoff and Bookmark procedures, are most appropriate given the nature of the task. For performance assessments, examinee-centred methods are preferable given the complex, multi-dimensional nature of OSCEs (Objective Structured Clinical Examinations) and workplace-based assessments.
- Regardless of the standard setting method selected, it is imperative to properly document all phases of the exercise, including the objective of the examination, the selection and composition of the panel, as well as the definition of the borderline or minimally proficient candidate.
- Providing evidence to support the stability of the cut-score is integral to supporting *internal* validity. Documenting the impact of applying a cut-score on pass/fail rates, as well the relationship to decisions on other similar assessments, is at the core of the *external* validity argument for the standard.

The Need to Make Decisions

The need to make decisions that assign people, objects, or things into 'classifications' permeates all aspects of daily life, from the mundane to the most significant. For instance, passing an examination to obtain a driver's licence requires meeting a certain level of proficiency with regard to knowledge of traffic laws and performance (passing, parallel parking, etc.). The aim of such a classification is to keep unsafe drivers from getting behind the wheel of a vehicle. Similarly, a jury that renders a verdict in a criminal trial is charged with 'classifying' a defendant as 'guilty' or 'not guilty', after carefully weighing the evidence of a case, i.e. analysing relevant data. The jury analogy seems particularly relevant to standard setting in assessment on a number of counts:

- Both activities require a sufficiently large and representative participant group from the population (whether a citizenry or a profession).
- Both activities necessitate a decision that will be used for classification purposes (rendering a verdict or setting a pass/fail standard).
- The intended use of the information is very similar in each instance (incapacitation and/or rehabilitation in a criminal trial and the corresponding protection of the public and remediation considerations in standard setting).

The need to make a decision is also part-and-parcel of all phases of a physician's professional life, from undergraduate medical education to revalidation. Key decisions occur when awarding or denying an unrestricted licence to practise medicine [1, 2], granting or withholding a credential [3–5], or granting or denying entry into a professional body [6, 7], as well as at the medical school level [8–12]. These decisions are arrived at through a process that is referred to as *standard setting*. Cizek [13] describes standard setting as 'the proper following of a prescribed, rational system of rules or procedures resulting in the assignment of a number

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

BOX 24.1 Definitions

- A *standard* is a qualitative description of a level of performance and can be viewed as a conceptual definition of competence.
- A *cut-score* or passing score corresponds to a number that reflects this standard and can be viewed as an operational definition of competence.

to differentiate between two or more states or degrees of performance' (e.g. pass/fail). This activity is especially critical within the health professions, given the need to ensure the public that graduates as well as holders of certificates and licences possess the knowledge and skill sets that permit safe clinical practice [14, 15]. In spite of this, a basic misconception still persists regarding the terms *cut-score* and *passing standard* (see Box 24.1).

Standard Versus Cut-score

The primary use of any test score in a criterion-referenced setting is to determine whether a candidate has mastered a set of competencies presumed to underlie performance on the examination. Whether at the school level or for licensure and/or certification decisions, standard setting exercises are routinely carried out to identify a passing standard, which is treated as an indicator of mastery or competency in the skill areas deemed important and measured by an examination.

Kane [16] defines a passing standard as a qualitative description of an acceptable level of performance and knowledge required in practice. As such, the passing standard can be viewed as a conceptual or qualitative definition of competence. For example, in a final-year undergraduate OSCE, a standard might stipulate that the borderline candidate demonstrate the data gathering, physical examination, and communication skills necessary for entry into supervised practice. The cut-score, on the other hand, is a number along the score scale that reflects the standard. It is an *operational definition of competence*. In our previous example, expert panellists might decide that a candidate who scores at or above 65% has met the performance standard for the final-year undergraduate OSCE.

Key Considerations in Standard Setting

Standard setting is a process that allows human judgements to be synthesised in a rational and defensible way to facilitate the partitioning of a score scale into two or more categories. Given the emphasis on expert judgement, it is important to underscore that all standards are intrinsically subjective in nature. Consequently, there is no 'gold standard' when it comes to setting a cut-score on an examination. Cut-scores can and will vary as a function of several factors, including, but not limited to, the method selected to set the standard and the panel of participating judges [17–21]. Jaeger [18] best summarised this point by stating that 'a right answer (in standard setting) does not exist, except, perhaps, in the minds of those providing judgement'. Following a systematic process that is supported with appropriate empirical evidence can help standard setting panels translate (policy-based) judgement onto a score scale in a defensible manner, but no method can be used to estimate some 'true' cut-score that perfectly separates masters from non-masters or passers from failers.

In view of the inherent subjectivity of any standard setting process, best practice dictates selection of a panel of judges that broadly represents the target examination population, with respect to background and educational characteristics [22, 23]. The composition of the standard setting panel becomes even more relevant given the complexity of assessments in medical education. Despite their seniority and level of expertise, extensive training of panellists is essential to ensure that the resulting cut-score is reasonable given the objectives of the assessment [24]. If nothing else, training is necessary to ensure that all panellists are in harmony with one another in regard to the goal of the assessment, the purpose of the standard setting exercise, the task that they are asked to complete, and a general definition of what constitutes minimal proficiency or a borderline performance [25]. A typical standard setting training session requires a number of steps including: (i) the provision of sample examination materials to panellists; (ii) a clear presentation of the task that participants are being asked to complete; (iii) a period of discussion allocated to the definition of the borderline candidate; (iv) judgements on a set of exemplars; (v) a discussion period to clarify any misconceptions amongst participants; and (vi) a post-exercise survey on all aspects of training [22].

Despite these caveats, the methods outlined in this chapter will provide systematic steps that can be followed to ensure that the resulting cut-score is defensible and based on informed, rather than capricious, judgements on the part of the expert panel. The difference between a norm-referenced standard and a criterion-referenced standard will first be reviewed prior to an overview of common methods for determining a cut-score on an examination (see Box 24.2).

BOX 24.2 Key considerations

- There is no 'gold standard' in standard setting.
- A standard and accompanying cut-score should reflect expert judgement as to what constitutes competence, supported by several sources of evidence.
- A standard setting panel should be composed of experts who broadly represent all key examination stakeholders with respect to gender, age, specialty, geographical area, etc.
- Thoroughly training panellists on all aspects of the exercise is a task critical to the success of any standard setting exercise, regardless of the method adopted.

Norm-referenced Versus Criterion-referenced Standards

At a very high level, standards can be classified as either norm-referenced or criterion-referenced in nature [26]. A norm-referenced standard is a *relative* standard in that the cut-score is derived from the performances of a comparative group of candidates. There are many examples of norm-referenced standards, such as setting the cut-score at one standard deviation above the mean of the class or fixing the cut-score at the 90th percentile rank of a distribution. The fundamental notion is that the cut-score is set solely as a function of the relative performances of a comparative group. We pass or fail a candidate on an examination purely based on how well (or badly) other test takers performed.

On the other hand, within a criterion-referenced framework, the standard is typically set as a function of the amount of knowledge of the domain that the candidate needs to demonstrate, irrespective of group performance. As such, it is an *absolute* standard. For example, a panel of medical experts might determine that a candidate needs to master 70% of the domain to be deemed minimally competent, based on their professional judgement and the objectives of an examination.

For professional examinations, criterion-referenced standards are generally preferred for a number of reasons. First, a norm-referenced standard tells little to nothing about what a given candidate knows or does not know, since it is entirely based on the relative performance of the group. Second, and more importantly, the cut-score selected in a norm-referenced standard setting exercise will vary as a function of the ability level of the group. Lower cut-scores will result from the performances of less proficient candidates, whereas higher cut-scores will be set with more able cohorts. This, in turn, produces cohorts of candidates who vary in regard to their level of competence. For example, setting a cut-score at one standard deviation below the mean will result in failing about 16% of any cohort, irrespective of what candidates may or may not know. However, it is conceivable that these groups could differ drastically in their knowledge of the domains. Scoring 'near the average' of a distribution can have quite a different meaning if the class is composed of high ability candidates versus less able students. That is, the meaning of a passing performance (and consequently 'minimal competence') can vary as a function of when and with whom the candidate passed.

Consequently, a norm-referenced approach to setting a passing standard is untenable from both political and professional perspectives. The only instance in which it may be acceptable to use a norm-referenced standard is when the selection of a small number of candidates is necessary (e.g. for a restricted number of postgraduate training slots).

Criterion-referenced methods for setting a standard are appealing because they overcome many of these limitations. A cut-score that is set using a criterion-referenced method reflects a level of proficiency that experts representing wide sectors of a given profession agree is indicative of a candidate who possesses the skills and knowledge required for safe practice. For this reason, criterion-referenced methods for setting cut-scores have been successfully employed and defended for several years in the medical licensing arena as well as with other health profession examination programmes [1, 2, 27–29]. The following two sections briefly describe the criterion-referenced standard setting methods in most common use (see Box 24.3).

BOX 24.3 Norm-referencing versus criterion-referencing

- A norm-referenced standard is a relative standard and set as a function of the performance of an arbitrary group of candidates.
- A criterion-referenced standard is an absolute standard and set as a function of what experts believe reflects competence, regardless of the overall performance of any group of candidates.
- With medical education examinations, norm-referenced standards are only appropriate for selection purposes. For the vast majority of decisions (e.g. graduation, passing a clerkship, etc.), criterion-referenced standards are appropriate.

Test-centred Methods

Criterion-referenced test-centred methods are appealing for setting a pass mark on knowledge assessments, such as multiple-choice examinations. In this form of standard setting, experts are asked to judge the level of performance required on each item of the test or task to meet the standard (e.g. minimal proficiency). Common and frequently used test-centred methods include the Angoff, Ebel, Nedelsky, and Bookmark procedures [30] (see Box 24.4).

Angoff Method

In the Angoff procedure, panellists are asked to estimate, on an item-by-item level, the proportion of minimally proficient candidates that *would* answer each item correctly [31]. Effectively this constitutes an assessment of the degree of difficulty of each component part of the test based on expert judgement. These proportions are then summed for each expert judge. Typically, the mean or median sum of item proportions across judges is treated as the cut-score on the examination. Box 24.5 provides a simple illustration of the Angoff procedure based on a five-item examination with three panellists. In this example, panellist cut-scores ranged from 1.35 (or 1/5) to 2.65 (or 3/5). An overall cut-score equal to 1.97/5 (or 2/5) could therefore be selected as the final cut-score.

Modified Angoff methods have also been proposed for determining a standard [9, 32–35]. One adaptation of the Angoff method allows panellists to modify their judgements following a general discussion [36]. Other revisions entail providing normative data (e.g. item difficulty and

BOX 24.4 FOCUS ON: Test-centred methods

- For MCQs, standards are typically set using a test-centred method. Popular test-centred methods include the Angoff, Ebel, Nedelsky, and Bookmark methods.
- Given that panellists are essentially asked to estimate characteristics of each individual item for the minimally proficient candidate in a test-centred standard setting exercise, i.e. difficulty with the Angoff and Bookmark methods, difficulty and relevance with the Ebel method, and additionally 'guessing' with the Nedelsky approach, discussion and broad agreement as to what constitutes a borderline candidate in the training phase is of critical importance.
- The Angoff and Bookmark methods are most commonly used to set a standard on MCQ examinations due to their inherent simplicity.
- The Ebel and Nedelsky methods impose stronger cognitive requirements on the part of panellists that may be difficult to meet with many examinations; respectively determining relevance as well as the likelihood that a borderline candidate will eliminate distractors.

BOX 24.5 Angoff standard setting example

In this five-item test, three judges are involved in standard setting and are asked to estimate, on an item-by-item level, the proportion of 'minimally proficient' candidates who would answer each item correctly.

Judge	1	2	3
Item 1	0.65	0.60	0.75
Item 2	0.60	0.40	0.60
Item 3	0.25	0.10	0.35
Item 4	0.10	0.05	0.55
Item 5	0.30	0.20	0.40
Overall cut-score	1.9 (or 2/5)	1.35 (or 1/5)	2.65 (or 3/5)

Overall cut-score = 1.9 + 1.35 + 2.65 = 5.9/3 = 1.97/5 or 2/5.

discrimination indices) following the initial round of ratings in order to provide panellists with a 'reality performance check' against which to gauge their initial judgements and modify them, if so desired, in a final round [37].

Advantages and Limitations

One main advantage of the Angoff family of methods is that they have been used extensively with a host of examinations, including both MCQ and performance-based assessments [34]. As such, a wealth of evidence and information is available to any researcher interested in carrying out such an exercise. Also, the Angoff method holds a certain amount of intuitive appeal in that panellists are required to review test items and offer judgements based on their expert knowledge of the material and candidates. Finally, the Angoff method is amenable to streamlining such as through the 'Yes/No' method [38], which can simplify the task even more.

On the downside, the Angoff methods have come under heavy criticism due to the inherent nature of the two main tasks that panellists are required to complete, namely to articulate what constitutes minimal proficiency and then consistently estimate proportions of minimally proficient candidates who would correctly answer each test item [35]. Shepard [39] argued that the task presented to panellists was too cognitively challenging and probably beyond the capability of most participants. Others, however, have refuted this claim and ascribed these difficulties to insufficient training of panellists or the absence of performance data to guide judgements [40]. Research conducted by Plake et al. [41] also showed that item performance estimates were consistent within and across panels, as well as within and across years for a high-stakes certification examination. These findings once more underscore the importance of selecting appropriate panels of judges for standard setting exercises and, more importantly, offering extensive training to all experts to eliminate any misconceptions regarding the nature of the task at hand. Despite these limitations, the Angoff family of methods continues to be one of the most prevalent, longstanding, and well researched set of procedures for setting a cut-score on an examination [30].

Ebel Method

The procedure outlined by Ebel extends Angoff's method by asking panellists not only to provide difficulty estimates for each item but also content relevance, given the domains that are presumed to underlie the examination [42]. The cut-score is computed by adding the cross-products of the difficulty and relevance judgements. Box 24.6 provides a simple example of a two-dimensional Ebel grid. In this example, judges felt that 5 of 50 items were essential to the content and 'easy' level of difficulty. In a similar vein, panellists were asked to estimate the proportion of items, in each content relevance/difficulty cell, that the minimally proficient candidate would correctly answer. The resulting cut-score is the sum of the relevance/difficulty cell cross-products. In this example, candidates would need to correctly answer 25/50 items (50%) to pass the examination.

Advantages and Limitations

Ironically, one advantage of the Ebel method for setting a standard, namely that item relevance, in addition to difficulty, can be factored into panellists' judgements, is also its chief weakness. Berk [43], for example, questions the ease with which panellists can separate content (difficulty) and relevance judgements during an exercise, largely based on the argument that these two dimensions are often correlated quite highly. From a test development standpoint, one could also question the merits of including

BOX 24.6 Ebel standard setting example

In this 50-item test, the standard setting panel is invited to consider both the relevance and degree of difficulty of items before estimating the proportion of questions that the minimally proficient candidate would correctly answer in each cell.

	Level of difficulty		
Content relevance	Easy	Average	Difficult
Essential	0.85 (five items)	0.65 (10 items)	0.25 (five items)
Important	0.75 (five items)	0.55 (five items)	0.15 (five items)
Acceptable	0.65 (three items)	0.45 (four items)	0.10 (three items)
Questionable	0.65 (two items)	0.40 (two items)	0.05 (one item)

Cut-score = 0.85(5) + 0.65(10) + 0.25(5) + 0.75(5) + 0.55(5) + 0.15(5) + 0.65(3) + 0.45(4) + 0.10(3) + 0.65(2) + 0.40(2) + 0.05(1) = 25.45/50 (50%).

test items that are not relevant in an examination. In most contexts, the total score is interpreted as an overall reflection of candidates' competencies on a composite of (interrelated) domains. Consequently, items that are deemed irrelevant contribute little to nothing in informing inferences about overall competency (e.g. pass/fail) or standing.

Nedelsky Method

Nedelsky [44] outlined a standard setting method based on the premise that when answering MCQs, minimally proficient candidates first eliminate options that they identify as incorrect based on their knowledge of the material, and then randomly guess amongst remaining choices. The actual cut-score corresponds to the sum across items of the reciprocal of the remaining number of alternatives. To illustrate; assume that a group of panellists estimates that the following number of options would be eliminated, respectively, by the minimally proficient candidate on a five-item, five-option MCQ examination: 2, 1, 3, 3, 4, across each of the items. The Nedelsky cut-score would therefore correspond to:

$$(1/3 + 1/4 + 1/2 + 1/2 + 1/1) = 2.58/5 \text{ or } 3/5(60\%)$$

Advantages and Limitations

The main advantage of the Nedelsky method is that it allows panellists to factor in the quality of the distractors when making their judgements, that is, any partial knowledge that the minimally proficient candidate may possess when answering an MCQ. However, the procedure also suffers from a number of well-documented shortcomings. First, the task imposed on panellists is much more onerous that what is expected in either an Angoff or Ebel exercise. Panellists must not only estimate the probability of a correct response on the part of the minimally proficient candidate, but they must do so in light of options they believe the latter test taker can eliminate either due to poor distractors or partial knowledge. Additionally, probability values that are provided by panellists are de facto restricted due to the nature of the procedure. For example, with a five-option MCQ, the only plausible estimates that judges can provide are: 0.20, 0.25, 0.33, 0.50, and 1.00 [43]. That is, the minimally proficient candidate can eliminate either 0, 1, 2, 3, or 4 options as non-plausible. Finally, and most importantly, the Nedelsky method assumes that the test-taking behaviour of minimally proficient candidates is identical, i.e. they guess in the same fashion from those alternatives not eliminated as implausible. This assumption has been seriously called into question given risk behaviours, differential partial knowledge, and other factors [45, 46]. Though modifications of the procedure have been proposed to address these limitations [47], the Nedelsky method has waned in popularity over the past few decades due to its inherent complexity and few practical benefits over more popular methods.

Bookmark Method

The Bookmark method is also used quite regularly to set a cut-score due to its intrinsic simplicity [48]. With this approach, test items are presented to panellists by order of difficulty from least to most difficult (one item per page in a booklet). Though the original intent of the method was to sequence the items as a function of item response theory (IRT)-based difficulty estimates, it is also possible to adapt the method and order the MCQs by simple *p*-values (proportion of correct responses). Each panellist is required to place a bookmark (a stopping rule) beyond which a minimally proficient candidate would not be expected to correctly answer remaining items. Note that the Bookmark method is also frequently employed for multiple judgements (e.g. determining levels of *basic*, *proficient*, and *advanced*). The final cut-score, in its simplest application, would correspond to the median number of items at the bookmark across panellists. It is important to point out that the original Bookmark procedure also translated this cut-score to the underlying IRT ability metric [48]. Extensions of the method that entail adding the use of performance benchmarks have also been proposed [49, 50]. Readers wishing to obtain more details on these revisions are encouraged to consult these references.

Advantages and Limitations

The main advantage of the Bookmark method is its simplicity and the relatively light cognitive load that is imposed on panellists, at least in comparison to other test-centred methods. Test items are ordered according to difficulty (again, unbeknownst to participants) and panellists are required to place one or several bookmarks to delineate two or more proficiency categories. Another attractive feature of the Bookmark method is that it can be readily applied to multiple-choice and performance examinations as well as mixed-format assessments. Finally, its traditional link to an IRT proficiency metric also holds great appeal given that the majority of large-scale testing programmes implement IRT-based methods for a host of activities, including test construction, scoring, scaling, and equating. As such, the Bookmark standard setting method can easily be integrated into a unified IRT framework.

Despite these advantages, the Bookmark standard setting method does possess a number of limitations that the practitioner should be aware of. First and foremost, the cut-score in a Bookmark standard setting exercise is inextricably linked to the difficulty of the test form. To illustrate, consider a test that is very 'easy' in relation to the proficiency level of candidates. This is often the case with medical licensing and certification examinations where over 90% of first-time test takers typically pass [27]. This 'mis-targeting' can make it impossible for panellists to set an appropriate bookmark. In certain instances, it is plausible that even the last item in a booklet is too easy to distinguish between masters and non-masters when the candidate sample is highly able. As others have mentioned [30], this problem could also crop up with other test-centred methods. The Bookmark approach, by virtue of item difficulty ordering, makes any such problems glaringly obvious. Another practical limitation of this standard setting method is that booklets (i.e. test items if there is one item per page) need to be re-ordered if some items are deleted due to poor performance. A final limitation is that items may not, and in fact are probably not, evenly spaced in terms of differences in difficulty from low to high throughout a test form. Thus, it might be difficult for panellists to identify an actual point along the scale that best discriminates between masters and non-masters, i.e. the bookmark might not be identifiable given gaps in item difficulty. While these limitations do not invalidate the Bookmark method, practitioners should be aware of these potential issues and plan accordingly prior to the actual standard setting exercise.

Examinee-centred Methods

Criterion-referenced examinee-centred methods, on the other hand, involve setting a standard based on global judgements of performance by a group of qualified expert panellists. Given the integrated, multi-dimensional nature of performance assessments in medical education, the latter methods are particularly well suited for setting a cut-score on OSCEs, for example [51]. Two popular examinee-centred standard setting methods are the contrasting groups method and the borderline group method [52, 53]. (see Box 24.7).

BOX 24.7 FOCUS ON: Standard setting for performance assessments

- For performance examinations, such as OSCEs and workplace-based assessments, examinee-centred methods are generally used to set a standard. Common examinee-centred standard setting methods include the contrasting groups and borderline group methods.
- These methods are appealing and well-suited to performance assessment as they allow panellists to provide overall holistic judgements of performance. They require panellists to assign candidates to two or more proficiency categories (e.g. master/non-master, unacceptable, borderline acceptable, clearly acceptable, etc.).
- While appealing, these methods inherently treat the panel as the 'gold standard'. Ample training is therefore necessary to ensure that the task is well understood as well as the definition of borderline performance.
- A number of technical issues need to be considered when implementing any examinee-centred standard setting method, including: (i) determining the costs associated with false-positive and false-negative classifications; (ii) ensuring that the borderline acceptable group is composed of a sufficiently large number of candidates; and (iii) for the contrasting groups method, assuring that panellists are able to assign candidates to one of two categories.

Contrasting Groups Method

In the contrasting groups method, panellists are asked, for each candidate, to review a performance profile (e.g. checklists and rating scales on an OSCE station) and determine whether the test taker is qualified or unqualified to pass the examination. OSCE station scores for both groups of candidates (unqualified and qualified) are then plotted on a graph. The score that best discriminates between both groups of test takers is typically selected as the cut-score [52–54]. A sample contrasting-groups plot is shown in Figure 24.1. In this example, the mid-point of the intersection zone could be selected as the cut-score value if false-positive and false-negative decisions were of equal importance. However, if the intent of the exam is to protect patients from malfeasance, a value in the upper part of the intersection zone would be chosen (minimising false-positive decisions, i.e. minimising the number of passing candidates who do not possess the clinical skills necessary to pass).

Borderline Group Method

In the borderline group method, panellists are also asked to review a performance profile for each candidate and identify unacceptable as well as acceptable performances. Additionally, panellists must designate those candidates that are deemed to lie just at a borderline acceptable performance level. The scores of these borderline acceptable examinees are then plotted on a graph. Typically, the median score value is chosen as the cut-score on the

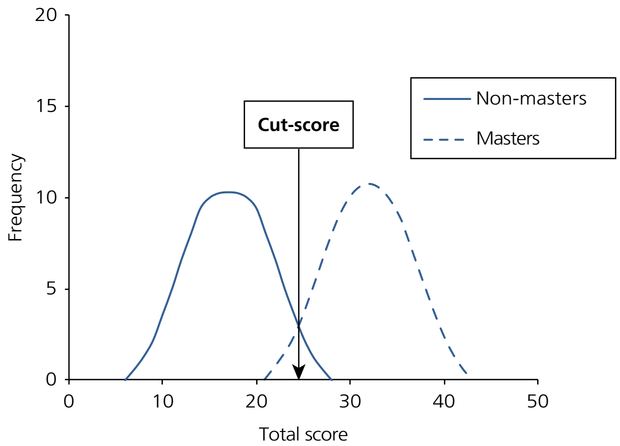


Figure 24.1 Identifying a cut-score using the contrasting-groups method.

examination [1, 53]. One limitation that has been raised with this approach is that the size of the borderline acceptable group might be quite small, thus contributing to a very unstable cut-score (e.g. median) value.

As a means of addressing this shortcoming, the *borderline regression method* was proposed as an alternative, related standard setting method. As it implies, this procedure uses linear regression modelling to predict the cut-score on the score scale as a function of the rating categories (e.g. unacceptable, borderline acceptable, acceptable). That is, the pass mark for a given OSCE station is obtained by regressing candidate scores (e.g. checklist scores) onto the global ratings. Unlike the more traditional borderline group method, all data points are used in determining the cut-score, not only those associated with borderline acceptable candidates [55].

Advantages and Limitations

The contrasting groups and borderline group methods are very similar in that they require panellists to make holistic judgements on the overall performance of candidates by classifying them into two (or more) categories. In fact, one could conceive of the borderline group method as a generalisation of the contrasting groups approach where experts not only need to determine whether a performance is acceptable or unacceptable, but also ‘on the cusp’, i.e. borderline acceptable. Given the high degree of similarity between the methods, it should come as little surprise that they carry the same advantages and limitations.

On the plus side, both methods are often preferred for performance assessment such as OSCEs and workplace-based assessments as they require panellists to complete a task that is more ‘intuitive’, i.e. classify candidates as either unacceptable, acceptable, or borderline acceptable. They are also well suited to these complex assessments given that dimensions on which to make classification judgements are often highly related. As such, these methods provide panellists with the latitude to incorporate all of their considerations when arriving at a classification decision with a candidate. The greater level of flexibility that is afforded by both approaches also potentially constitutes their chief limitation. Both methods treat panellist judgements as intrinsically reliable and valid, i.e. as *the* gold standard. Any factor that can detract from the panellists’ ability to provide such judgements will bias the ultimate cut-score value in a way that is difficult to predict and will lead to a standard that is most certainly unfair to subgroups of candidates. Consequently, the moderator plays a critical role in ensuring that the training offered to panellists can at least minimise this effect to ultimately assure a defensible process for all stakeholders. It is easy to envisage a scenario where panellists, who might very well be familiar with the candidates who they are evaluating, are affected by construct-irrelevant factors when providing their judgements. Such construct-irrelevant factors might include gender, ethnicity, dress, personality, work habits, and a myriad of other extraneous features that are unrelated to ‘competency’, as broadly defined by the examination.

Both the contrasting groups and borderline group methods also rest on the central premise that a sufficiently large group of representative professionals in the field can be identified for an exercise and also trained to complete the task at hand as instructed. Inadequate training can lead to a number of undesirable outcomes, including the propensity to assign disproportionally large number of candidates to the borderline acceptable group [56, 57]. While this may sound appealing, given that the cut-score is derived from the performances of the latter group, classifying nearly all candidates as borderline acceptable seriously raises questions about the quality of the examination, instruction, and other factors, while yielding a cut-score that is again biased in ways that are difficult to ascertain.

Related to this point, the borderline group method does require that the latter group be composed of a sufficiently large number or the resulting cut-score, whether the median score in the simplest case or a predicted value based on more complex statistical modelling (e.g. logistic regression, latent class analysis, etc.), will be unstable and inappropriately reflect ‘minimal competency’. Given the dichotomous nature of the task that is required in a traditional contrasting groups standard exercise, it might also be difficult for panellists to classify candidates as either unacceptable or unacceptable, with no option for a borderline acceptable performance. Plake and Hambleton [56], amongst others, proposed an extension of the method that does allow for a finer gradation of the decision scale. Finally, it is critical, for both methods, that the medical educator clearly set a policy that outlines the consequences of misclassifying a candidate. Treating both false-positive (passing a candidate who should have failed) and false-negative (failing a candidate who should have passed) decisions equally might be quite undesirable in instances where protection of the public is of prime consideration. Under the latter scenario, minimising false-positive classifications is of greater concern. Conversely, in lower-stakes settings, minimising false-negative errors could be perfectly acceptable as a policy. All of the potential limitations associated with the contrasting groups and borderline regression methods, given the immense responsibility that is conveyed upon panellists, again underscore the critical role that the moderator needs to play in such standard setting exercises. Indeed, it is not an exaggeration to state that the moderator can 'make or break' a borderline group or contrasting groups standard setting exercise.

Hofstee Method

The use of criterion-referenced approaches for setting a standard can lead to unacceptable outcomes in the absence of political considerations associated with the decision. That is, the cut-score arrived at following a standard setting exercise should not result in failing or passing an unacceptably large or small proportion of candidates. To illustrate, assume that a given medical specialty examination has consistently failed around 15% of candidates. Further assume that this population is very comparable, ability wise, from year to year. If the cut-score set after an Angoff exercise results in failing 50% of candidates, the standard is unrealistic and might very well be unacceptable from a policy standpoint.

As a means of providing a 'reality check', Hofstee [58] proposed a 'compromise' method that involves asking panellists the following questions, the answers to which are subsequently graphed in a (Hofstee) plot:

- Considering the content as a whole, what are the maximum and minimum tolerable cut-scores? These are typically labelled C_{min} and C_{max} on the Hofstee plot.
- What are maximum and minimum tolerable failure rates? These are usually listed as F_{max} and F_{min} on the Hofstee plot.

An example of a Hofstee plot is provided in Figure 24.2.

In order to create this plot, a cumulative percentage-correct score distribution needs to first be computed. This distribution outlines the cumulative percentage of candidates who would fail at each point along the score scale. Then, the coordinates (C_{min}, F_{max}) and (C_{max}, F_{min}) are plotted and joined by a straight line, as illustrated in Figure 24.2. The point of intersection between this line and the frequency distribution corresponds to the Hofstee cut-score. The cut-score is illustrated by the 'cut' value shown on the x-axis. In the example outlined in Figure 24.2, panellists felt that the cut-score should be no lower than 55 (C_{min}) and no higher than 85 (C_{max}). Similarly, they indicated that the failure rate should be at least 10% (F_{min}) but not higher than 50% (F_{max}). Linking both sets of coordinates and drawing a line down to the x-axis yields a Hofstee cut-score value of 65, which would result in failing about 35% of the candidate cohort. The aim of the Hofstee method is generally to determine whether criterion-referenced standards fall within the vicinity of the Hofstee-based value, i.e. whether they are consistent with political considerations and global impressions of cut-score values and failure rates [59].

Advantages and Limitations

The primary advantage of the Hofstee method is that it allows panellists to offer holistic judgements on cut-score values and failure rates with few to no constraints. Based on their experience, knowledge of the test content, and objective of the examination, panellists must define performance parameter limits. The flexibility and ease with which one can implement the Hofstee method also constitutes its chief limitation. That is, it is not generally viewed as a primary standard setting method but rather as a 'reality check' or fall-back method meant to complement other approaches, whether test- or examinee-centred. Within this supporting context, the Hofstee method can provide valuable information that can help the practitioner gauge whether a cut-score set with a more traditional method gibes with the general expectations of panellists. However, it should generally not be used as a standalone measure given its ad hoc nature. Another more controversial method of standard setting increasingly used in the medical education arena is discussed in the Box 24.8.

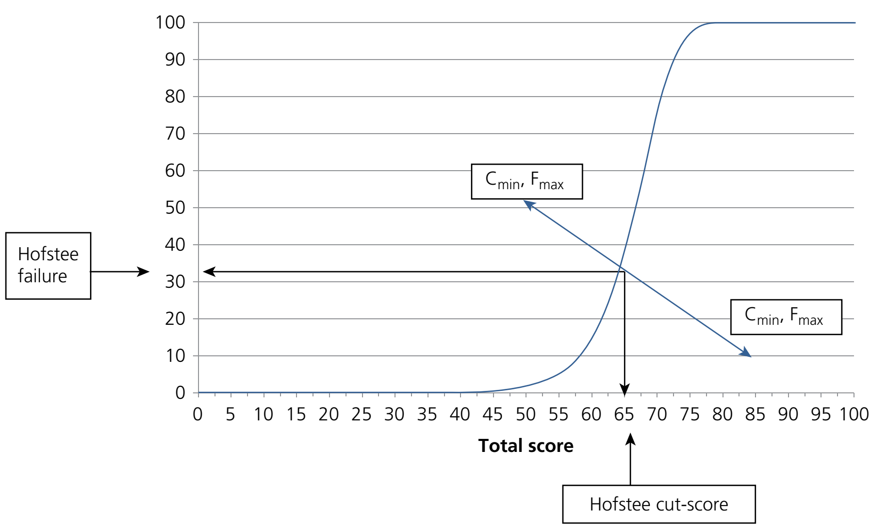


Figure 24.2 Identifying a cut-score using a Hofstee plot.

BOX 24.8 FOCUS ON: The Cohen method

Standard setting methods such as Angoff are resource intensive and time-consuming. Faculty with small staff numbers and limited financial resources can struggle to collect a sufficient number of experts required for reliable and valid methods.

The Cohen method is an alternative form of standard setting increasingly used in medical education where the 'best performing' students (student score at the 95th percentile or P95) are used as a reference point [10]. Medical educators determine what proportion of this high-performing group score is acceptable as a cut score, e.g. 60% × P95. The modified Cohen takes this further proposing that historical data from multiple criterion-referenced exams, within the programme, can personalise this Cohen sum to better reflect the pass mark expected by panels of experts. The modified Cohen is therefore a mixed method, combining both criterion-referenced and norm-referenced data in the creation of the cut score [60]. Users of the method think the Cohen cut score is time efficient and less resource intensive and fair to students in that all students can pass the exam and the cut-score changes with the level of difficulty of the exam. Critics of the Cohen method, however, perceive this cut-score to be norm-referenced, rather than the desired criterion-referenced, as it relies on a pre-determined and relatively arbitrary proportion of the actual cohort performance to create the cut-score.

Selecting a Criterion-referenced Standard Setting Method

The American Educational Research Association 'Standards for Educational and Psychological Testing' [61, p. 53] clearly state that 'there can be no single method for determining cut-scores for all tests or for all purposes, nor can there be any single set of procedures for establishing their defensibility'. Along these lines, Angoff [61] also noted that 'regarding the problem of setting cut-scores, we have observed that the several judgemental methods not only fail to yield results that agree with one another, they even fail to yield the same results on repeated application'.

Despite the fact that no single method can lead to the identification of an 'optimal' cut-score value, as the latter is always embedded in professional judgement, there are nonetheless a number of factors that the medical educator might wish to consider when selecting a standard setting approach. An overview of these factors is presented next.

The extent to which a clear standard setting process is adhered to has the greatest impact on the cut-score. This process, regardless of the method adopted, should include a clear definition of the objective of the examination as well as the standard setting exercise, extensive training of panellists to minimise any misconceptions, as well as a clear outline of what constitutes minimal proficiency or a borderline acceptable performance. However, a number of factors can be considered to select a standard setting method that might be most suitable given the intended aims of the examination and the associated decision that the test score user wishes to make.

One of the first questions to ask is what is the format of the examination? For knowledge-based examinations (e.g. MCQs), test-centred methods are most appropriate given the task that panellists are asked to complete, i.e. estimate a cut-score based on a review of the actual test items. Conversely, for performance assessments, such as OSCEs and workplace-based tasks, examinee-centred methods are more suitable for setting a standard given the complex, multi-dimensional nature of performance. The latter typically entail holistic judgements of performance.

Second, the user may also wish to consider the format of the examination. For example, some standard setting methods (e.g. the Nedelsky method) were developed exclusively for use with MCQs. While some methods can be used with different formats (e.g. Angoff methods), certain assumptions are made that may or may not meet expectations. For example, the Angoff method and its offshoots assume that performance is compensatory in nature, i.e. candidates can compensate for doing poorly in certain parts of the examination by doing well in other sections. These methods would therefore be inappropriate in a conjunctive setting, where different components need to be successfully and independently completed. Other methods (Hofstee, contrasting groups) were developed as test-format invariant.

One erroneous belief that is often promulgated is the one that suggests that combining a multitude of methods when setting a standard will provide a 'better cut-score'. It is important to reiterate that standard setting and the selection of a cut-score are ultimately policy decisions, albeit derived from informed judgement. There is little evidence to suggest that combining multiple methods will lead to a 'better' standard [57]. Since there is no 'correct' cut-score, how can policy makers synthesise results from multiple approaches? This strategy also requires significantly more resources. It is always better to systematically implement one standard setting method rather than provide results from several (poorly) implemented approaches. Again, the process that is followed when arriving at a cut-score is ultimately what needs to be defended. The latter includes properly documenting all phases of a standard setting exercise, clearly describing the selection and training of panellists, as well as providing empirical evidence to support the use of a cut-score. These data typically include the impact of sources of variability (judges, panels, etc.) on the cut-score value as well as the consequences of implementing a cut-score (e.g. the appropriateness of pass/fail rates in light of historical trends). The importance of validating any cut-score is underscored in the next section (see Box 24.9).

The next section provides some practical guidelines to aid in the selection of a standard setting method.

BOX 24.9 HOW TO: Choose a standard setting method

- No standard setting method can yield an 'optimal' cut-score value as this is based on experts' internal construction of what constitutes competence.
- The extent to which a process is systematically implemented and supported with appropriate sources of evidence is much more important than the selection of any standard setting method.
- However, several factors can be considered in the choice of a standard setting method, including the format of the examination (MCQ versus performance assessment).
- Combining several methods will not yield a 'better' standard as the choice of any cut-score is ultimately a policy decision based on a number of considerations.

Gathering Validity Evidence to Support a Cut-score

Regardless of the standard setting method adopted, gathering evidence to validate the resulting standard is a critical step [62, 63]. As stated throughout this chapter, what is ultimately of importance with any standard setting exercise is the extent to which a process is systematically adhered to and can be defended using a number of evidential sources.

The evidence to support *procedural validity* needs to be clearly documented in the standard setting report. This usually comprises the first part of any standard setting report and entails a thorough account of each step of the exercise including:

- An overview of the targeted examination and its purpose.
- A clear articulation of the selected standard setting method implemented with a supporting rationale.
- The process used to select the panel of expert judges, as well as a description of their qualifications and the extent to which they represent the profession as a whole.
- An outline of all phases of the exercise, including the training process, definition of the performance standard, and how data were collected.

Surveying panellists on various aspects of the standard setting exercise constitutes a final important piece of supporting procedural validity evidence. How confident are the panellists in the process and, more importantly, in the resulting cut-score? Evaluating judges' impressions of the training phase as well as the cut-score can provide strong confirmation for any standard setting exercise.

Evidence to support the *internal validity* of the cut-score is also of great importance given the high-stakes nature of most criterion-referenced examinations. That is, how precise is the estimate of the cut-score and how reproducible is it across any facet of interest? With regard to precision, if the cut-score is relatable to an item response theory ability metric, the (conditional) standard error of the proficiency estimate associated with a cut-score can provide a straightforward indication of the stability of the latter value. With an observed score scale (e.g. number-right, percentage-correct, etc.), the practitioner can also estimate the amount of error associated with a cut-score using a compound binomial model [64].

Additionally, the extent to which the cut-score is impacted as a function of the judges participating in an exercise, the panel of judges (if multiple groups are involved), the items/stations selected, etc. can be readily assessed using generalisability theory [8, 65]. This framework allows the medical educator to estimate the amount of variability in scores (including the cut-score) that can be ascribed to any facet or potential source of measurement error as listed above. Similarly, IRT-based rating scale models [66] can also provide useful information with respect to the ability distribution of candidates, difficulty of items/stations, as well as stringency of raters. Regardless of the complexity of the models utilised to gather evidence of internal validity, the aim of this critical source of information is to provide an indication of the stability or precision with which a cut-score is estimated, primarily to provide some boundaries to the practitioner in order to minimise its misuse.

Evidence to support the *external validity* of a cut-score should also be part of any standard setting effort as this relates directly to the impact of implementing a standard. Assessing the reasonableness of the cut-score in light of its impact on failure rates is generally at the core of external validation efforts. For example, assume that a graduation OSCE has typically failed between 10 and 12% of a class. A failure rate of 55%, following a standard setting exercise, would warrant considerable scrutiny of the cut-score and its appropriateness, assuming that the cohort is of comparable ability to past groups and the OSCE of a similar difficulty level.

A comparison of results to other assessments constitutes another important source of external validity for any proposed cut-score. For example, how comparable are pass/fail rates to grades or the status of students on other examinations measuring similar constructs (e.g. a prior OSCE)? Though we would not expect two examinations to measure exactly the same combination of domains, they should nonetheless yield a comparable standing for most candidates.

Conclusions

Standard setting is an intrinsic part of all assessment activities in medical education, from undergraduate training to physician revalidation efforts. Determining whether a candidate has mastered any number of competencies underlying an examination is a key outcome used not only to render individual judgements but also to evaluate programme effectiveness, teaching efficacy, etc. [67, 68].

First and foremost, it is important to reiterate that there is no gold standard and that all cut-scores ultimately reflect informed judgement from a group of content experts on what level of performance constitutes 'competency'. Systematically following a standard setting process and supporting its use with appropriate empirical evidence is therefore central to any such exercise.

This chapter has described a number of standard setting methods that the medical educator might wish to consider based on the nature of their examination as well as practical and financial concerns. Though most of the examples presented in this chapter focused on high-stakes assessment, the same principles and procedures are also helpful in formative assessment environments. For example, in a mastery learning and assessment setting, some of the approaches laid out in this chapter might be useful in determining the behaviours that reflect high achievement standards in a given domain [69, 70].

Irrespective of the method selected to arrive at a cut-score on an examination, several issues need to be addressed prior to undertaking a standard setting exercise. First, the panel of judges should be viewed as a microcosm of all exam stakeholders and as such should mirror any characteristic deemed important by the profession, be that geographical area, medical school location, specialty, gender, or ethnicity. Convening such a broad panel will ensure that views from most members of the profession are incorporated in the exercise, and ultimately, the standard.

Determining a suitable number of panellists for any standard setting panel is also critical. Inviting too few panellists is ill-advised, as the judgements of a single dissenting judge could have an undue impact on the value of the final cut-score. On the other hand, assembling a large panel may not be cost-effective. Consequently, clearly identifying the desired characteristics of the group, as outlined above, can provide valuable information for determining the panel's optimal size. Once set, it is also important that the cut-score for any examination be periodically revisited to ensure its continued appropriateness in light of any changes that may have occurred in the profession, whether political or content-based in nature. Finally, it is important to restate that different standard setting methods will produce different cut-score values. The central aim in any standard setting exercise should be to: (i) defend the choice of a particular method, (ii) meticulously document all steps followed throughout the exercise, and (iii) base the selection of the standard on as much empirical evidence as possible, factoring in global impressions as well as the consequences of adopting a given cut-score. Hopefully, this chapter provides a convenient guiding framework for any medical educator who needs to identify a cut-score for an examination and highlights some of the issues to consider when conducting a standard setting exercise, irrespective of the method adopted.

Acknowledgement

With thanks to Assistant Professor Carmel Tepper of Bond University, Gold Coast, Australia for providing a summary of the Cohen method (Box 24.8).

References

1 Dauphinee, W.D., Blackmore, D.E., Smee, S. et al. (1997). Using the judgments of physician examiners in setting the standards for a national multi-center high stakes OSCE. *Advances in Health Sciences Education: Theory and Practice* 2: 201–211.

2 Melnick, D.E., Dillon, G.F., and Swanson, D.B. (2002). Medical licensing examinations in the United States. *Journal of Dental Education* 66: 595–599.

3 Norcini, J.J. (1994). Research on standards for professional licensure and certification examinations. *Evaluation and the Health Professions* 17: 160–177.

4 Hess, B.J., Weng, W., Lynn, L.A. et al. (2011). Setting a fair performance standard for physicians' quality of patient care. *Journal of General Internal Medicine* 26: 467–473.

5 Sturmberg, J.P. and Hinchy, J. (2010). Borderline competence – from a complexity perspective: conceptualization and implementation for certifying examinations. *Journal of Evaluation in Clinical Practice* 16: 867–872.

6 Mucklow, J. (2011). Development and implementation of the specialty certificate examinations. *Clinical Medicine* 11: 235–238.

7 Lee, R.P., Venkatesh, B., and Morley, P. (2009). Evidence-based evolution of the high stakes postgraduate intensive care examination in Australia and New Zealand. *Anaesthesia and Intensive Care* 37: 525–531.

8 Lagha Richter, R.A., Boscardin, C.K., May, W., and Fung, C.C. (2012). A comparison of two standard-setting approaches in high-stakes clinical performance assessment using generalizability theory. *Academic Medicine* 87: 1077–1082.

9 Jalili, M. and Norcini, J.J. (2011). Comparison of two methods of standard setting: the performance of the three-level Angoff method. *Medical Education* 45: 1199–1208.

10 Cohen-Schotanus, J. and van der Vleuten, C.P. (2010). A standard setting method with the best performing students as point of reference: practical and affordable. *Medical Teacher* 32: 154–160.

11 Ricketts, C., Freeman, A.C., and Coombes, L.R. (2009). Standard setting for progress tests: combining external and internal standards. *Medical Education* 43: 589–593.

12 Boursicot, K.A., Roberts, T.E., and Pell, G. (2006). Standard setting for clinical competence at graduation from medical school: comparison of passing scores across five medical schools. *Advances in Health Sciences Education: Theory and Practice* 11: 173–183.

13 Cizek, G.J. (1993). Reconsidering standards and criteria. *Journal of Educational Measurement* 30: 93–106.

14 Kane, M.T., Crooks, T.J., and Cohen, A.S. (1999). Designing and evaluating standard-setting procedures for licensure and certification tests. *Advances in Health Sciences Education* 4: 195–207.

15 Norcini, J.J. and Shea, J.A. (1997). The credibility and comparability of standards. *Applied Measurement in Education* 10: 39–59.

16 Kane, M.T. (2001). So much remains the same: conception and status of validation in setting standards. In: *Setting Performance Standards: Concepts, Methods, and Perspectives* (ed. G.J. Cizek), 53–88. Mahwah, NJ: Lawrence Erlbaum Associates.

17 Impara, J.C. and Plake, B.S. (2000). A comparison of cut scores using multiple standard setting methods. Paper presented at the meeting of the American Educational Research Association, New Orleans, LA (April 24–28, 2000).

18 Jaeger, R.M. (1989). Certification of student competence. In: *Educational Measurement*, 3e (ed. R.L. Linn), 485–514. New York: Macmillan.

19 Livingston, S.A. and Zieky, M.J. (1989). A comparative study of standard setting methods. *Applied Measurement in Education* 2: 121–141.

20 Longford, N.T. (1996). Reconciling experts' differences in setting cut-scores for pass–fail decisions. *Journal of Educational and Behavioral Statistics* 21: 203–213.

21 Norcini, J.J. and Shea, J. (1992). The reproducibility of standards over groups and occasions. *Applied Measurement in Education* 5: 63–72.
22 Hambleton, R.K. and Pitoniak, M.J. (2006). Setting performance standards. In: *Educational Measurement*, 4e (ed. R.L. Brennan), 433–470. New York: Macmillan.
23 Jaeger, R.M. (1991). Selection of judges for standard setting. *Educational Measurement, Issues and Practice* 10: 3–6. 10, 14.
24 Raymond, M.R. and Reid, J.B. (2001). Who made thee a judge? Selecting and training participants for standard setting. In: *Setting Performance Standards: Concepts, Methods and Perspectives* (ed. G.J. Cizek), 119–157. Mahwah, NJ: Lawrence Erlbaum.
25 Skorupski, W.P. and Hambleton, R.K. (2005). What are panelists thinking when they participate in standard setting studies? *Applied Measurement in Education* 18: 233–256.
26 Norcini, J.J. (1994). Research on standards for professional licensure and certification examinations. *Evaluation and the Health Professions* 17: 236–241.
27 Swanson, D.B., Case, S.M., Waechter, D. et al. (1993). A preliminary study of the validity of scores and pass/fail standards for USMLE steps 1 and 2. *Academic Medicine* 68: s19–s21.
28 Wendt, A. and Kenny, L. (2007). Setting the passing standard for the National Council Licensure Examination for Registered Nurses. *Nurse Educator* 32: 104–108.
29 De Champlain, A.F. (2004). Ensuring that the competent are truly competent: an overview of common methods and procedures used to set standards on high-stakes examinations. *Journal of Veterinary Medical Education* 31: 61–5.
30 Cizek, G.J. (2012). *Setting Performance Standards: Foundations, Methods and Innovations*. New York: Routledge.
31 Angoff, W.H. (1971). Scales, norms, and equivalent scores. In: *Educational Measurement*, 2e (ed. R.L. Thorndike), 508–600. Washington, DC: American Council on Education.
32 Plake, B.S. and Impara, J.C. (1997). Standard setting: an alternative approach. *Journal of Educational Measurement* 34: 353–366.
33 Downing, S.M., Lieska, N.G., and Raible, M.D. (2003). Establishing passing standards for classroom achievement tests in medical education: a comparative study of four methods. *Academic Medicine* 78: s85–s87.
34 Hambleton, R.K. and Plake, B.S. (1995). Using an extended Angoff procedure to set standards on complex performance assessments. *Applied Measurement in Education* 8: 41–55.
35 Ricker, K.L. (2006). Setting cut-scores: a critical review of the Angoff and modified Angoff methods. *The Alberta Journal of Educational Research* 52: 53–64.
36 Jaeger, R.M. (1978). A proposal for setting a standard on the North Carolina high school competency test. Paper presented at the meeting of the North Carolina Association for Research in Education, Chapel Hill, NC.
37 Norcini, J.J. (1988). The effect of various factors on standard setting. *Journal of Educational Measurement* 25: 57–65.
38 Plake, B.S. and Cizek, G.J. (2012). Variations on a theme: the modified Angoff, extended Angoff, and yes/no standard setting methods. In: *Setting Performance Standards: Foundations, Methods and Innovations* (ed. G.J. Cizek), 181–199. New York: Routledge.
39 Shepard, L. (1995). Implications for standard setting of the National Academy of Education evaluation of National Assessment of Educational Progress achievement levels *Proceedings of the Joint Conference on Standard Setting for Large-Scale Assessments*, vol. II, pp. 143–160. Washington, DC: US Government Printing Office.
40 Hambleton, R.K., Brennan, R.L., Brown, W. et al. (2000). A response to 'setting reasonable and useful performance standards' in the National Academy of Sciences' Grading the Nation's Report Card. *Educational Measurement, Issues and Practice* 19: 5–14.
41 Plake, B.S., Impara, J.C., and Irwin, P.M. (2000). Consistency of Angoff-based predictions of item performance: evidence of the technical quality of results from the Angoff standard setting method. *Journal of Educational Measurement* 37: 347–356.
42 Ebel, R.L. (1972). *Essentials of Educational Measurement*. Englewood Cliffs, NJ: Prentice-Hall.
43 Berk, R.A. (1986). A consumer's guide to setting performance standards on criterion-referenced tests. *Review of Educational Research* 56: 137–172.
44 Nedelsky, L. (1954). Absolute grading standards for objective tests. *Educational and Psychological Measurement* 14: 3–19.
45 Gross, L.J. (1985). Setting cutoff scores on credentialing examinations: a refinement in the Nedelsky procedure. *Evaluation and the Health Professions* 8: 469–493.
46 Melican, G.J. and Plake, B.S. (1985). Are correction for guessing and Nedelsky's standard setting method compatible? *Journal of Psychoeducational Assessment* 3: 31–36.
47 Smith, R.M. and Gross, L.J. (1997). Validating standard setting with a modified Nedelsky procedure through common item test equating. *Journal of Outcome Measurement* 1: 164–172.
48 Lewis, D.M., Mitzel, H.C., Green, D.R., and Patz, R.J. (1999). *The Bookmark Standard Setting Procedure*. Monterey, CA: McGraw-Hill.
49 Barton, P.E. (2009). Contents for a new NAEP report: The five largest states. Paper presented for the National Assessment Governing Board. https://nagb.gov/focus-areas/reports/contents-new-naep-report-mega-states.html (accessed 26 June 2018).
50 Pashley, P.J. and Phillips, G.W. (1993). *Toward World-Class Standards: A Research Study Linking International and National Assessments*. Princeton, NJ: Educational Testing Service.
51 Boulet, J., De Champlain, A. and McKinley, D. (2003). Setting defensible performance standards on OSCEs and standardized patient examinations. *Medical Teacher* 25: 245–9.
52 Berk, R.A. (1976). Determination of optimal cutting scores in criterion-referenced measurement. *Journal of Experimental Education* 14: 9475–9469.
53 Livingstone, S.A. and Zieky, M.J. (1982). *Passing Scores: A Manual for Setting Standards of Performance on Educational and Occupational Tests*. Princeton, NJ: Educational Testing Service.
54 Clauser, B.E. and Nungester, R.J. (1997). Setting standards on performance assessments of physicians' clinical skills using contrasting groups and receiver operator characteristic curves. *Evaluation in the Health Professions* 20: 215–238.
55 Wood, T.J., Humphrey-Murto, S., and Norman, G.R. (2006). Standard setting in a small scale OSCE: a comparison of the Modified Borderline-Group Method and the Borderline Regression Method. *Advances in Health Sciences Education* 11: 115–122.
56 Plake, B. and Hambleton, R. (2001). The analytic judgment method for setting standards on complex performance assessments. In: *Setting Performance Standards: Concepts, Methods, and Perspectives* (ed. G.J. Cizek), 283–312. Mahwah, NJ: Erlbaum.
57 Cizek, G.J. and Bunch, M.B. (2007). *Standard Setting: A Guide to Establishing and Evaluating Performance Standards on Test*. Thousand Oaks, CA: Sage Publications.
58 Hofstee, W.K.B. (1983). The case for compromise in educational selection and grading. In: *On Educational Testing* (ed. S.B. Anderson and J.S. Helminck), 107–127. San Francisco, CA: Jossey-Bass.
59 De Gruiiter, D. (1985). Compromise models for establishing examination standards. *Journal of Educational Measurement* 22: 263–9.
60 Taylor, C.A. (2011). Development of a modified Cohen method of standard setting. *Medical Teacher* 33: e678–e682.
61 American Educational Research Association (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association, American Psychological Association, National Council on Measurement in Education.
62 Angoff, W.H. (1988). Validity: an evolving concept. In: *Test Validity* (ed. H. Wainer and H. Braun), 19–32. Hillsdale, NJ: Erlbaum.
63 Kane, M.T. (1994). Validating the performance standards associated with passing scores. *Review of Educational Research* 64: 425–461.
64 Lord, F.M. (1984). *Standard Errors of Measurement at Different Score Levels (Research Report RR-84-8)*. Princeton, NJ: Educational Testing Service.

65 Verhoeven, B.H., van der Steeg, A.F., Scherpbier, A.J. et al. (1999). Reliability and credibility of an Angoff standard setting procedure in progress testing using recent graduates as judges. *Medical Education* 32: 832–837.

66 Andrich, D. (1978). A rating scale formulation for ordered response categories. *Psychometrika* 43: 561–573.

67 Brandon, P.R. (2005). Using test standard-setting methods in educational program evaluation: addressing the issue of how good is good enough. *Journal of MultiDisciplinary Evaluation* 2: 1–29.

68 Bennett, J., Tognolini, J., and Pickering, S. (2012). Establishing and applying performance standards for curriculum-based examinations. *Assessment in Education: Principles, Policy & Practice* 19: 321–339.

69 McGaghie, W. (2015). Mastery learning: it is time for medical education to join the 21st century. *Academic Medicine* 90: 1438–1441.

70 Inui, T.S. (2015). The charismatic journey of mastery learning. *Academic Medicine* 90: 1442–1444.

Further Reading

American Educational Research Association (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association, American Psychological Association, National Council on Measurement in Education.

Cizek, G.J. and Bunch, M.B. (2007). *Standard Setting: A Guide to Establishing and Evaluating Performance Standards on Test*. Thousand Oaks, CA: Sage Publications.

25 Formative Assessment: Assessment for Learning

Diana F. Wood
University of Cambridge School of Clinical Medicine, Cambridge, UK

KEY MESSAGES

- Formative assessment promotes a number of desirable educational outcomes, including learner self-regulation and the development of lifelong learning skills.
- Students with a high level of self-regulation are more effective learners, showing increased resourcefulness, resilience, persistence, and success.
- A well-designed programme of formative assessment linked to overall curriculum aims and the teaching and learning goals of individual modules enhances the learning experience for students.
- Effective feedback is central to the process of formative assessment.
- Teachers in medical education identify the development of constructive feedback skills as the most important aspect of their professional development.
- Formative assessment linked to curriculum design should be an essential component of medical education at all levels.

Introduction

Assessment forms a major element of any teaching and learning programme and should be recognised as integral to the whole educational enterprise, not just delivered as an 'add-on' at the end of a course. Traditionally, in medical education, assessment was used to demonstrate that information had been transmitted in some way from the teacher to the learners, the latter memorising notes diligently and reproducing them as necessary in formal examinations requiring factual recall.

One of the most obvious developments in medical education over the past 20 years or so has been a greater understanding of assessment and the way it can be used to enhance both students' learning and the overall quality of the educational experience. At its most pragmatic level, this reflects our recognition that 'assessment drives the curriculum' [1]. Students learn what is needed to pass examinations and use weighting of assessments as a means to rank the importance of various parts of the syllabus. If assessments are developed independently and added on to the teaching programme to test 'what students have learned' (usually by factual recall), then well-meaning attempts to foster deep learning and understanding of a subject will founder.

Furthermore, students expect to be assessed and tend to use grading systems that compare them with their peers as a means of evaluating the amount of work required of them to perform well in the course [2]. This in itself has provided a challenge to many medical educators faced with introducing minimum-competency 'pass/fail' assessments in which students are not ranked against their peers but against pre-set, minimal-competency guidelines.

The key is to align assessment with the educational desires of the faculty and the aims of curriculum planners. In a well-designed curriculum, faculty members are aware of their educational goals from the outset and build in the design and timing of assessments to ensure that these goals are addressed by the teaching programme.

This chapter predominantly relates to the use of formative assessment in undergraduate medical education. However, the principles described are derived from a variety of sources, including the general educational literature, and can be extrapolated to all levels of medical education. In many cases, the formative assessment methods described here can be directly transferred to postgraduate education and the continuing professional development arena with correction only for the level of the learners in relation to the educational goals.

The following areas will be considered:

- definitions of formative and summative assessment
- teacher and learner perspectives on formative assessment and some of the research evidence underpinning them

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

BOX 25.1 Some functions of assessment

Assessment may be used to:

Measure student learning	Against a pre-set criteria
Grade students	Against a standard
	Against a comparative group
Summarise achievement	For the student
	For the faculty
	For other interested bodies, e.g. university, potential employer, etc.
Indicate readiness to progress	
Provide feedback	On learning
	On why a mark was given by teaching staff
Diagnose specific misunderstandings	
Motivate students to learn	
Focus and direct student learning	
Help students learn more effectively	
Inform the teaching programme	Review what students do not know or understand
	Review teaching and learning methods
Promote staff development	Ensure that faculty are aware of the curriculum goals and understand how assessment forms part of the programme
Contribute to education quality assurance	

- the role of feedback in formative assessment, including examples from experiential learning settings in communication skills teaching
- how formative assessment may be used within a curriculum
- examples of formative assessment in different teaching and learning environments.

Assessment is a complex construct, and recognition of its various purposes will help ensure that an individual educational programme achieves its multiple goals (see Box 25.1). Classically, assessment has been divided into two categories: *formative* and *summative*. In essence, formative assessment provides feedback to learners about their progress, whereas summative assessment measures the achievement of learning goals at the end of a course or programme of study.

Summative assessments measure the achievement of learning goals at the end of a course or programme of study. Summative assessments are formal and used to determine progression to the next stage of a course, to signify the need for remediation, for graduation purposes, or for registration with a national professional body. 'High-stakes assessments' are summative assessments with implications for professional progression.

In general, little feedback is provided to students from summative assessments except in the case of failure. In recent years the distinction between formative and summative assessment has become blurred, with essentially formative workplace-based assessments being collated and used for summative purposes. The term 'assessment for learning' encompasses some of this change, showing how evidence of successful completion of experiential learning tasks can be credited towards summative assessment and progression to the next stage of training.

Much of the literature related to formative assessment derives from studies in primary and secondary schools and in general higher education. However, a focus on the development of learner self-regulation and its benefits for lifelong learning means that the general principles described below can be related to medical education at all levels.

Assessment can be thought of as serving three main functions; *assessment of learning*, *assessment for learning*, and *assessment for quality assurance*. Whilst summative assessment fits most neatly into the first category and formative assessment into the second, in a well-designed educational programme, there is considerable overlap such that the results of ongoing, formative assessment can be used both to measure student learning and to inform institutional quality assurance procedures.

Characteristics of Formative Assessment

Formative assessment refers to any assessment that is designed specifically to provide feedback. It has been defined as follows:

> … encompassing all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged [3, p. 8].

More recently, the same authors refined this definition to include five features of formative assessment that are more directly applicable to medical education [4]:

- clarifying and sharing learning intentions and criteria for success
- engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding

BOX 25.2 FOCUS ON: Benefits of formative assessment

General	Informal
	Ongoing and frequent
	Dynamic
	Non-judgemental
	Part of the overall teaching and learning process
Effects on students	Allows detailed feedback
	Promotes self-directed learning
	Raises self-esteem
	Engages students in the learning process
	Encourages deep learning and understanding
	Motivates learning
	Identifies insecurities
	Offers help with specific remediation
Effects on staff	Allows detailed feedback
	Promotes self-directed learning by the students
	Fosters interactive teaching and learning methods
	Encourages varied and challenging teaching methods
	Identifies students in difficulty early in the curriculum
	Develops teaching skills
	Evaluation feeds into curriculum development

- providing feedback that moves learners forward
- activating students as instructional resources for one another
- activating students as the owners of their own learning.

Some benefits of formative assessment that reflect these features are shown in Box 25.2. Feedback is central to effective formative assessment. In general, formative assessment should be ongoing, frequent, non-judgemental, and carried out in informal settings. For students, the availability of a regular, dynamic interaction with their tutors helps them engage with the learning process, acting as a motivator and encouraging deep learning and understanding. Furthermore, it offers them the opportunity to identify their learning difficulties in a safe environment and to take up remedial assistance if appropriate.

For teachers, formative assessment encourages the development of skills associated with the promotion of self-directed learning in the student. Teachers are motivated by better understanding of their students' needs and by helping them become more self-regulated in their learning. Review of the teaching and assessment programme feeds into curriculum development and forms part of ongoing curriculum evaluation.

Overall, consideration of the effects of formative assessment on students and teachers suggests that it should be a positive experience for both groups. Students are encouraged to engage in active learning and teachers are encouraged to develop skills with which to provide a challenging educational experience in a supportive environment.

Formative assessment can play a major role in the acquisition of lifelong learning skills by helping students self-regulate their learning activities. A well-designed series of formative assessments can make a major contribution to the educational impact of an overall assessment programme, a characteristic that is as important as the reliability and validity of the individual assessments themselves [5]. Thus, in the ideal situation, formative assessment is a two-way process between learner and teacher, placing the student at the centre of the activity.

In reality, assessment is usually seen as the province of teachers, many of whom regard feedback primarily as a means of transmitting information to students. Often, little thought is given to how feedback information received during formative assessment is processed by students. In this respect, assessment has not kept pace with other developments in teaching and learning in higher education, where the emphasis has shifted towards a dialogue between teacher and student, fostering self-direction, and motivation to learn. An assessment process that focuses solely on the teacher's role overlooks the need to help students gain the skills of self-regulation necessary for lifelong learning and ignores the way in which feedback interacts with students' motivation and beliefs. To understand how formative assessment can be most effective, it is therefore necessary to consider the process from the point of view of both teacher and students.

Teacher Perspectives

From the teacher's perspective, the formative assessment process could be described in the following three steps:

1. Review the student's work.
2. Evaluate the work against a reference framework that reflects the pre-set learning objectives and the level expected of students at a particular stage in the course.
3. Make a judgement on the work and provide verbal or written feedback to the student on that judgement.

The apparent simplicity of these steps is misleading, mainly because it disguises the expertise of individual teachers and their differing levels of skill and experience, particularly in giving feedback. Such 'teacher factors' were reviewed by Sadler [6], who identified six important characteristics that highly competent teachers bring to the assessment process. These characteristics are summarised in Box 25.3.

Highly competent teachers are not only knowledgeable but also bring a positive attitude to teaching, with an ability to empathise with their students and a desire to see them improve. Such teachers are reflective about their own skills and show concern for the integrity of the judgements they make. They demonstrate skill in constructing assessments using a variety of methodologies and are aware of assessment criteria and the standards expected of students at

BOX 25.3 Characteristics of highly competent teachers that affect the quality of formative assessment

Characteristic	Effect on formative assessment
Knowledge	Greater knowledge base and understanding of the subject matter than the students
Attitude to teaching	Empathy with students, ability to communicate educational goals, desire to help students improve, concern for the integrity of their own judgements
Skill in constructing assessments	Use of varied assessment tools to develop different skills in students
Knowledge of assessment criteria and appropriate standards	Awareness of standards and appropriate expectations of students' performance at a certain level within the curriculum based on learning outcomes and previous experience of student achievement
Evaluative skills	Ability to make qualitative judgements informed by experience as assessors
Expertise in giving feedback	Identification of strengths and weaknesses, evaluative comments in relation to criteria, suggestions for alternative learning methods, examples of different ways to achieve the goals

Source: Adapted from Sadler [6].

different levels within the curriculum. They learn from their experience in assessment and develop expertise in giving constructive feedback.

Clearly, in any given faculty, the level of expertise will vary between teachers. The importance of the skills of individual teachers in formative assessment reinforces the requirement for assessment to be designed as part of an institutional educational programme, particularly in relation to staff development and appraisal.

Student Perspectives

From the student perspective, formative assessment should be a means to improve performance and aid development as self-directed and motivated learners. The term 'self-regulation' is used to describe the way in which students monitor their learning behaviour by setting and achieving goals, managing resources, and adapting to external feedback. In doing so, students generate their own internal feedback, helping them evaluate their progress towards goals and to adapt their learning processes in the face of obstacles or changes in motivation. Self-regulated learners are aware of their own knowledge, beliefs, and cognitive skills, and they use these to interpret external feedback effectively [7, 8].

Nicol and Macfarlane-Dick reviewed the literature relating to formative assessment and self-regulated learning, elaborating the student's role in developing internal feedback mechanisms and modelling the relationship between internal and external feedback (Figure 25.1) [9]. This model is useful as it illustrates the way in which a task set by the teacher acts as a trigger for internal regulatory processes within the student, drawing on prior knowledge and motivation to learn. The process generates a set of internal outcomes, such as increased understanding or changes in motivational state, in addition to the external outcomes reflected in a piece of work submitted for assessment. External and internal outcomes are linked by external feedback requiring the student to engage actively with such input. Evidence from the literature suggests that students with a higher level of self-regulation are more effective learners, showing increased resourcefulness, persistence, and success.

This student-centred approach to formative assessment described in the general educational literature is consistent with the constructivist approach to learning widely adopted in medical education through formal problem-based learning and other forms of problem-orientated learning and assessment [10]. Translation of social constructivist theories into practice places the relationship between teacher and student and the effective use of feedback at the centre of the educational endeavour [11, 12].

Feedback

Feedback provides the route by which assessment becomes a tool for teaching and learning, and it is central to the concept of formative assessment. Feedback following assessment encourages the student and teacher to work together to improve the student's understanding of a subject. The teacher shows that they are interested in the student's opinions, seeks clarification where appropriate, and, where necessary, encourages the student to approach a topic in a different way. Feedback provided in a non-judgemental and open fashion allows the student to feel more confident to discuss their difficulties and plan better approaches to learning where necessary.

It has been recognised for many years and across all educational sectors that effective feedback is positively correlated with student achievement [13], although it is also clear from early studies that the quality of feedback is vital. Poor-quality feedback may have no effect or may even be detrimental [3].

Effective Feedback

Feedback can be defined as a way in which learners become aware of the gap between their current level of knowledge

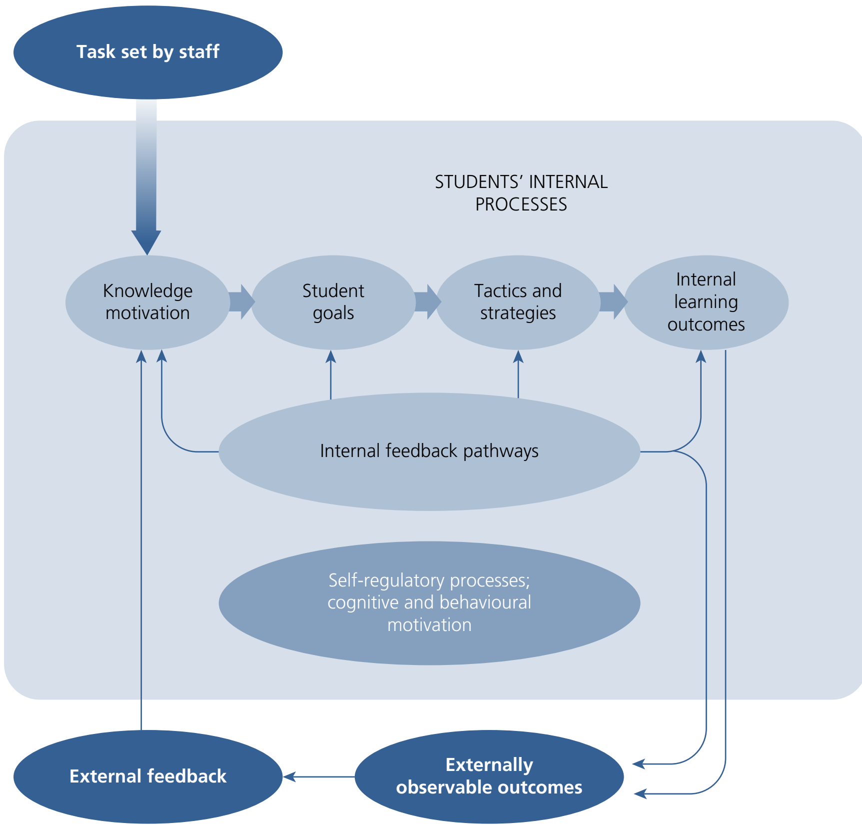


Figure 25.1 A model of self-regulated learning illustrating the relationship between external factors and internal self-regulation in the student. *Source*: Adapted from Nicol and Macfarlane-Dick [9].

or skill and the desired goal. It provides guidance towards reaching the goal, but effective feedback is achieved only when the student takes action to narrow the gap [14, 15]. This implies not only that the educational goals are clearly described, but also that students are able and empowered to take the necessary action to achieve them. This in turn means that effective feedback is a collaboration between teachers and learners rather than just a function of teaching per se.

Nicol and Macfarlane-Dick propose seven principles of good feedback that can facilitate the development of self-regulation (see Box 25.4) [9]. These emphasise the need for learner and teacher to work together towards learner self-regulation. They clarify the teacher's role in providing information to the student about their performance and also, crucially, about what is expected of them and how to recognise the gap between current and expected attainment. These principles raise a number of issues pertinent to medical education that can be considered under the heading 'Education for feedback'.

Education for Feedback

On entry into medical school, students are generally very well motivated and academically capable, having achieved high standards in national school exit examinations, prior higher degrees, and/or medical school selection examinations. Despite this, some students experience early failure, contributing to internal demotivation and a cycle of further failure. Paradoxically, for the small number of students who underachieve in medical school, the learning habits developed to produce high-level performance in pre-medical school examinations may be the reason for their failure at the undergraduate level [16]. In particular, students who have learnt previously in a didactic teaching environment and become successful at memorising facts may be challenged by small group learning, problem-solving, dealing with 'grey areas' of knowledge, and scoping vast amounts of information to contain their learning goals within reasonable limits.

A programme of formative assessment and feedback introduced early in a course can go a long way towards preventing the onset of the cycle of failure and demotivation in these students. A well-designed formative assessment programme will ensure that students are aware of their goals and the ways in which these might be achieved. However, it is incumbent on the faculty to explain this process to the students at the beginning of the course. Individual and group feedback can be difficult for students and, if handled badly, can be detrimental to their progress. Educating students about formative assessment and feedback is essential to ensure maximum gain from the process. This in itself requires planning and thought – it is unlikely that a single lecture at the start of a course will be effective. Integrating the educational process with early feedback sessions and modelling good feedback within a learning group is more likely to be educationally valuable [17].

BOX 25.4 Good feedback practice

- Helps clarify what good performance is
- Facilitates the development of self-assessment (reflection) in learning
- Delivers high-quality information to students about their learning
- Encourages teacher and peer dialogue around learning
- Encourages positive motivational beliefs and self-esteem
- Provides opportunities to close the gap between current and desired performance
- Provides information to teachers that can be used to help shape teaching

Source: Adapted from Nicol and Macfarlane-Dick [8].

Much of the research into feedback in medical education comes from the field of teaching communication skills, and these principles have been carried forward into other experiential and reflective learning environments. In the following section, two well-known approaches to feedback developed in the field of communication skills teaching are compared to provide examples of how effective feedback can be used in practical teaching situations.

Feedback in Experiential Learning Settings

Experiential learning in one-to-one or small group settings forms the basis of communication skills teaching programmes, usually in the form of observed simulated consultations. Students learn through frequent practice accompanied by feedback and reflection. This type of teaching can be particularly challenging for students who may feel exposed when required to perform difficult communication tasks in front of their tutors and peers. It places demands on teachers, who need to be aware of the dynamics within a learning group and who need to be trained and capable of handling students' responses and reactions. In these situations, feedback should not only be constructive, but is best delivered within a framework that is known and accepted by both teachers and students.

One widely used set of guidelines for providing feedback during teaching about consultation skills was described by Pendleton et al. in 1984 [18] and has become known as 'Pendleton's Rules'. The stimulus for developing these guidelines was primarily the observation that feedback in medical education is traditionally negative, pointing out students' errors, while failing to draw attention to their strengths and successes. Application of this type of feedback to the experiential learning settings being introduced in communication skills teaching was more destructive than constructive, leading students to develop negativity about the whole teaching process and resent the use of role play and other observational teaching methods. In experiential learning it is clearly important that students should feel that they are in a safe environment. Actions must be confidential to the teacher and the learning group, and students should be supported in identifying their own strengths and weaknesses and helped in addressing areas of concern.

Pendleton's Rules stress the need for safety in the learning environment by emphasising the need to discuss the learners' strengths before commenting on their weaknesses, and to make recommendations rather than criticise. Furthermore, in each part of the process, the learner makes the first comments – this self-evaluation not only encourages them to develop skills of reflection but also enables the teacher to assess these skills and address any difficulties students may have in self-reflection.

Pendleton's Rules applied to a small group learning session following a simulated consultation model can be summarised as follows from the teacher's point of view:

- clarify any issues of fact
- ask the learner to comment on what went well and why
- ask the group to discuss what went well and why, and add comments
- ask the learner to comment on what went less well and how it could be done differently
- discuss what could be done differently, and how, with the whole group.

There are a number of advantages to this approach to feedback. From the student's point of view, it provides a consistent framework in a safe environment – the student knows what to expect at the end of an observed consultation. The emphasis on self-assessment helps the student become more reflective about learning. The 'rules' force the student to think about positive aspects of their performance and to become aware of their individual strengths in communication. The requirement for positive comments by the student, the teacher, and the group means that no student receives only negative feedback, and any adverse comments must be presented in a constructive way as recommendations for change. The overall effect is that the feedback experience should enhance motivation to learn and encourage the development of self-regulation. Finally, from the teacher's point of view, it provides a simple structure within which much can be achieved – this is particularly important for relatively inexperienced teachers.

There are, however, a number of disadvantages to Pendleton's Rules, mainly related to their enforcement of a strict order for the way in which feedback is given. By ensuring that each student receives positive feedback at the beginning of the process, the individual student's own agenda may be overlooked – students themselves feel that within the time available during a teaching session, there is little opportunity for the constructive criticism they desire. Interestingly, this may reflect much-needed change in the culture of teaching and learning in medical schools over recent years, and the improved methods used by teachers for experiential learning activities. Increasingly, students expect their opinions to be sought by the faculty, and the culture of persistently negative feedback prevalent at the time of the development of Pendleton's Rules is disappearing. Students appreciate the opportunity to reflect on their successes, but are anxious to receive advice on how they might improve their performance.

The agenda-led outcome-based analysis (ALOBA) of the consultation, described by Silverman et al. in 1996 [19],

BOX 25.5 Agenda-led, outcomes-based analysis of the consultation (ALOBA)

Task for the teacher	Reason
Organise the feedback	
Identify the learner's agenda	Helps the learner to express their views on the consultation and describe what help they would like from the group
Discuss the outcomes that both learner and patient were trying to achieve	The learner starts to recognise the importance of their own desired outcomes and those of the patient
Allow the learner to comment first	Encourages self-assessment and reflection
Involve the whole group in problem solving	All students become more analytical of the consultation and reflect on how they might perform in the same situation
Group feedback	
Invite feedback from all members of the group	Helps all students develop feedback skills, including making specific non-judgemental comments
Ensure balanced feedback	Allows all students to support the learner by considering both what went well and what was less successful
Suggest alternatives rather than make prescriptive comments	The learner can consider alternative approaches and how they might work
Be supportive, act as a role model	All students can observe the use of constructive feedback
Ensure that feedback leads to greater understanding	
Rehearse suggestions	Allows the learner time to try out alternatives and for group comment on the effects
Use the consultation as learning material	All group members can contribute to the session and can learn as much as the learner under observation
Develop a wider discussion	Allows the introduction of concepts and research evidence to the group
Structure, summarise, and record	Provides structure for the teaching allowing maximum learning benefits for the students; record the learning to inform future sessions

Source: Adapted from Silverman et al. [19].

provides an alternative mechanism for giving feedback in small group and one-to-one experiential learning situations. The ALOBA approach is built around the students' own agenda, allowing them to identify their individual problems in the context of their own and the patient's desired outcomes for a consultation. It provides opportunities for a group to give feedback, thus encouraging the development of feedback skills in all the learners in a small group setting. Finally, it allows the teacher to introduce a wider discussion of theoretical concepts and research evidence. The principles of the ALOBA method are shown in Box 25.5.

For a teacher using the ALOBA method, the task can be divided into three sections:

- organising the feedback
- group feedback
- ensuring that feedback leads to greater understanding.

Following a real or simulated consultation, the feedback process is organised in terms of the learner's agenda, requiring them to identify the problems they have encountered and the help they would like from the group. The student should first identify what outcomes they wished to achieve from the consultation and, with the rest of the group, should consider the patient's agenda in terms of outcomes (this part of the process may include the opinions of simulated or trained patients where present). The learner can then be asked to comment on the process, and the whole group is asked to join in the problem-solving process, identifying the issues, feeding them back to the learner, and generating solutions.

In the second part of the process, specific feedback is invited from all members of the group. It is the teacher's responsibility to ensure that feedback is balanced, non-judgemental and descriptive in nature, and that the group offers suggestions and alternatives rather than prescriptive comments. This is a particular opportunity for the teacher to be seen as a role model, providing constructive criticism in a supportive environment.

In the final part of the process, the teacher has more freedom to optimise the learning opportunities of the session. Thus, the learner may be given time to rehearse suggestions made by the group, allowing other group members to see the effects of their suggestions. The teacher may take the opportunity to widen the discussion, introducing aspects of their own experience or drawing on the research evidence for a particular aspect of the consultation. Finally, the teacher should summarise the session, providing structure and offering suggestions for further learning within an appropriate conceptual framework. Recording the learning that has occurred in a session is another useful activity that can form the foundation of future sessions.

The advantages of the ALOBA method for providing feedback are, paradoxically, that having placed the individual student and their own agenda at the centre of each learning experience, the session becomes of more value to all the students involved. By offering the opportunity for students to consider the problems inherent in a consultation and engage in problem solving, all participants are more involved in the learning process. Whereas Pendleton's Rules may result in the learner becoming the passive recipient of feedback from all the other participants, the ALOBA technique ensures that everyone is equally engaged in the process. From the teacher's point of view, the ALOBA method also provides an opportunity to introduce some of the concepts underlying good communication skills, providing a theoretical structure for the students to understand their learning. However, the ALOBA method does require more experienced teachers and may be daunting for the less skilful.

Other methods for giving feedback in experiential learning have been described [20–22], all of which contain aspects of the models described above.

Helpful and Unhelpful Feedback

Feedback is central to the process of formative assessment. Constructive feedback can enhance the learning experience and promote learner self-regulation; destructive feedback can have profoundly negative effects on learning. For many teachers in medical education, the development of constructive feedback skills is seen as the most important aspect of their professional development [23]. The principles of constructive feedback for experiential learning situations have been described elsewhere [24–26] and are summarised in Box 25.6. In essence, helpful feedback is specific, non-judgemental, behavioural, and descriptive, and is provided within a supportive educational environment close to the time of the learning experience.

Similar criteria can be applied to feedback given in other formative assessments. For example, when marking written pieces of work, simply giving a grade or making a comment such as 'Good work' is less helpful to the learner than a description of why the work is good and suggesting other issues that might have been included or arguments that might have been presented. A number of systems for classifying students' responses have been described, of which the most well-known is the structure of the observed learning outcome (SOLO) taxonomy [27] (see Box 25.7). The levels described in this scheme are not content specific and can be applied to students' work at any stage in a curriculum, assuming the teacher is aware of the aims of the module and the level of attainment expected. Student work that scores in levels 4 and 5 shows evidence of categorising and structuring knowledge, characteristics associated with deep learning. Feedback offered to the students within this (or a similar) framework

BOX 25.6 HOW TO: Give helpful feedback in experiential learning

Unhelpful feedback	Reason	Helpful feedback	Reason
'Your body language wasn't very good at the start'	Judgemental	'At the beginning you were looking at the computer screen records and not at the patient as she started to tell her story.'	Descriptive, detailed, behavioural
'You weren't very empathetic'	Non-specific	'You didn't acknowledge the problems she has dealing with her husband's illness.'	Identifies specific problem
'You're very abrupt'	Personality issue	'You interrupted a lot, for example…' (give specific points in consultation)	Behavioural, specific
'I think it would be better if you did it this way'	Advice	'Have you thought about trying it like this?'	Generating alternatives
'I don't think you heard everything with your hearing problem'	Hearing problem not resolvable in this situation	'You have always discussed your hearing problems with us. Was there any point at which you thought it was affecting the consultation?'	Supportive, possibly can be changed by altering the environment
'You didn't notice how upset she was'	Judgemental	'At one point she was looking down and appeared quite upset. You quickly continued by asking her direct questions about her medication and she never returned to the problem of what was upsetting her. Did you notice that?'	Descriptive, non-judgemental, specific
'It was really good'	Non-specific	'At the start you asked an open question and then allowed her to tell her story. You left silences so that she continued in her own words.'	Positive, specific, descriptive

BOX 25.7 The SOLO taxonomy to classify the structural complexity of students' written work

Level	Descriptor
1 Prestructural	Use of irrelevant information or no meaningful response
2 Unistructural	Answer focuses on one relevant aspect only
3 Multistructural	Answer focuses on several relevant features, but they are not coordinated
4 Relational	The several parts are integrated into a coherent whole: details are linked to conclusions; meaning is understood
5 Extended abstract	Answer generalises the structure beyond the information given: higher-order principles are used to bring in a new and broader set of issues

Source: From Biggs and Collis [27].

is more helpful to their learning than simple judgemental statements.

Finally, it is useful to check with the students that they have understood the feedback they have been given and that their interpretation of the feedback is correct. The students' perception of the feedback they have received may vary greatly from that of the teacher who gave the feedback [28] and only by checking can these discrepancies be addressed.

One of the major difficulties encountered in medical education is the delivery of feedback during clinical teaching sessions. Whilst students are able to give and receive feedback in a formal way in classroom or simulation-based teaching, they may have difficulty recognising feedback on their clinical performance with real patients, when a teacher may offer feedback within a general discussion of a clinical case. Clinicians teaching in clinical environments need to be careful to ensure that they give deliberate and specific feedback, even if it is brief. Faculty development is essential to ensure that clinical teaching staff are equipped to deliver timely and effective feedback during the course of clinical work [24, 26, 29, 30].

Formative Assessment in the Curriculum

A programme of formative assessment with effective feedback can be used to develop self-regulation in learners, leading to better outcomes in terms of their learning and overall success. Formative assessment should be considered as part of a teaching institution's assessment strategy alongside summative assessments. The following section reviews the way in which a programme of formative assessment can be designed within a curriculum and considers some examples of different types of formative assessment used in medical education.

Formative Assessment and Module Design

A programme of formative assessment should be built into the design of a teaching module that has explicit learning outcomes (which can be assessed). The module design should show clearly how evaluation of teaching, learning, and assessment will be performed, allowing development of the teaching programme in the future. A number of schemata for module design incorporating assessment, feedback, and evaluation have been described, such as the one shown in Figure 25.2.

This basic pattern of module design is appropriate for all types of learning in medical education, including classroom-based activities and experiential learning [31]. The example given in Figure 25.2 is based around threshold criteria (pass/fail assessments) but can be adapted to include grading, where appropriate. The model takes into account not only the aims of the teaching module but also the level of attainment expected at the particular point in the curriculum. Module designers can therefore translate the level descriptors into learning outcomes and hence threshold assessment criteria. At that point the assessment methods are designed alongside the teaching and learning strategy for the module. Having delivered the module and performed the assessment as designed, evaluation of the module includes the teaching methods and the appropriateness of the learning outcomes and assessments used.

Using this format for curriculum design, assessment forms an integral part of the teaching programme and can easily be blueprinted against curriculum content and teaching methodologies. This 'constructive alignment' of the curriculum ensures that assessments facilitate learning by being linked explicitly to the learning outcomes such that internally coherent assessments (both formative and summative) are embedded in curriculum design and review [32, 33]. Linking the assessment programme explicitly to the design and evaluation of a single module should ensure that evaluation of the assessment process itself is not overlooked [34].

Examples of Formative Assessment in Undergraduate Medical Education

Recognition of the value of formative assessment as a means of enhancing teaching and learning in medical education has led to an increase in its use in undergraduate programmes. The ability of formative assessment to identify students with difficulties and then to offer them remedial teaching is an important feature of the widening use of formal formative assessment. The principles of good

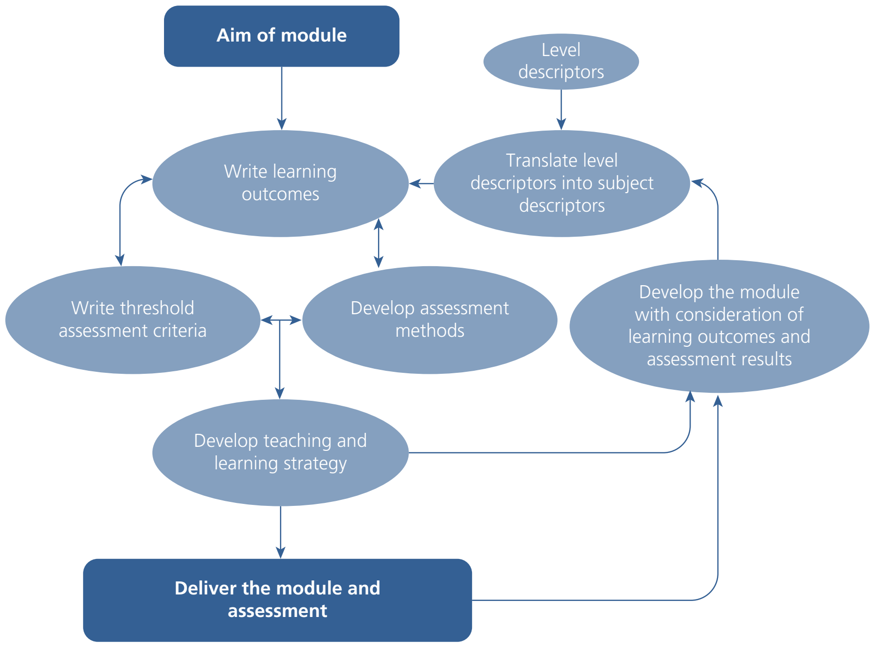


Figure 25.2 Curriculum module development and review. *Source*: Adapted from Moon [31].

formative assessment can be applied to all areas of assessment in medical education, including:

- knowledge testing
- testing competence – practical, communication, and clinical skills
- experiential learning settings – hospital clinical placements, general practice, community placements
- portfolios.

Often the most appropriate formative assessments include elements of different learning activities (e.g. an Objective Structured Clinical Examination [OSCE] plus written work), allowing the faculty to give a rounded assessment of student performance and provide help in specific areas. This section focuses on examples of formative assessment in three areas of undergraduate medical education:

- hospital clinical placements
- teaching consultation skills in general practice
- assessment of portfolios and reflective writing.

Hospital Clinical Placements

Much of the teaching and learning in medical schools occurs in the context of hospital clinical placements (clerkships). One of the most well-recognised and disheartening aspects of traditional medical education programmes was the ability of students to pass through a clinical teaching programme and only be identified as having problems when they failed a summative assessment or even their final examinations (often not to the surprise of teaching staff). Formative assessment can identify struggling students earlier in the course and, coupled with appropriate identification of the learning difficulties and additional teaching, can result in improved student performance [16].

Most medical schools expect students on clinical placements to receive feedback on their performance, usually in the form of a grade, which is regarded as highly subjective by the students. Furthermore, learners report a lack of regular feedback or describe feedback that they perceive to be poorly given or unfair, and they may become defensive, especially to feedback given by non-medical clinicians such as nurses or paramedical staff [35]. It is less common to have a formative assessment process in which the 'firm grades' form part of the overall assessment. In my own medical school, we have developed assessment programmes that combine formative and summative elements and which are supplemented by additional teaching tailored to students' needs (Box 25.8). Other schools have made similar changes to their assessment programmes with improved student performance [16, 36].

However, the format of the assessment programme is crucial – the introduction of an in-training assessment (ITA) consisting of a range of assessment formats to an internal medicine clerkship was not found to increase the number of supervisions or the quality of feedback received by the students [37]. That particular ITA was complex, requiring the student to undertake a number of supervised encounters or presentations related to 13 core competencies, and was accompanied by feedback. The commitment of the senior clinicians to such a programme needs to be very high as it is ongoing and time consuming. Asking too much of busy clinicians by way of formative assessment and feedback may be counterproductive. It may prove to be more valuable to organise centralised assessment formats assessing competence, which can be used in either formative or summative ways, and to ask for workplace-based assessment in the assessment of performance. Workplace-based assessment is discussed in Chapter 22 of this book.

Consultation Skills in General Practice

General practice placements often provide particularly good environments for formative assessment with appropriate

BOX 25.8 HOW TO: Combine formative and summative assessments as part of a programme of assessment

The University of Cambridge standard undergraduate medical course is a six-year programme. In the first three years the emphasis is on core medical sciences and all students undertake a Bachelor's degree, usually (but not exclusively) in one of the biomedical sciences. The final three years of the course focus on clinical medicine. This programme is divided into three year-long stages: Core Clinical Practice, Specialist Clinical Practice, and Applied Clinical Practice. In each year there are formative assessments which identify students with difficulties, allowing them to receive additional teaching in the area of concern. Groups of formative assessments are combined to form hurdles that the students must pass prior to entry into the high-stakes summative end of year exams. As an example, the assessment programme in Psychiatry is described below.

The assessment consists of:

- Full record of adequate attendance.
- Record of four long-case histories having been presented to medical staff during the placement.
- Completed Record of Experience ('log-book').
- Individual meeting with the supervisor at the end of the clinical placement to discuss the feedback from teaching clinicians, review the Record of Experience, and summarise the student's performance.
- Written paper comprising 30 Single Best Answer questions; feedback is provided by the specialty lead teacher immediately after the exam with a discussion of the correct and incorrect answers.

This set of formative assessments is combined as a summative hurdle, which students must pass prior to entering the high-stakes end of year examinations.

At the end of years four and five, each student has a one-to-one interview with a senior member of the teaching faculty. Detailed feedback is given to them in the light of their own reflection on their performance. Students who identify specific difficulties, such as in practical or communication skills, are then referred for additional tuition during the succeeding part of the course. A record is made of the interview and students followed up by the faculty.

The advantages of this combination approach are that:

- Each student is provided with a detailed review of their progress.
- A range of assessment modalities is used to inform the discussion.
- Detailed knowledge of each student's performance held by the clinical supervisor is acknowledged and relayed to senior faculty members.
- Students are able to 'benchmark' themselves and evaluate the amount and type of work they need to do in order to succeed in clinical medicine.
- Students with difficulties are clearly identified and can be offered additional targeted support.

feedback. In general, students are attached to a practice in small numbers and may be given the opportunity to see patients on their own or in observed consultations. A wide range of assessment methods for postgraduate trainees in general practice have been described, including observed consultations, review of video-taped consultations, multi-rater assessment, and evaluation by peers and patients. General practitioner teachers are therefore skilled in assessment and feedback, and many of these methods can be extended to the undergraduate curriculum. Work reported from the Department of General Practice in Leicester suggests that postgraduate assessment methods can be transferred successfully with high levels of reliability, validity, and educational impact [38].

Portfolios

Much has been written about the use of portfolios in medical education, and there remains a debate about what the term actually means and what should be contained in a medical student's learning portfolio. To some, it is a repository for assessment grades, written pieces of work, and lecture notes. Others have tried to harness the potential of portfolio learning to encourage students to develop reflective practice and adult learning skills [39, 40]. Online learning portfolios have the potential to enhance the progression towards adult learning.

Portfolios lend themselves to use in formative assessment as they can be the centre of discussion at student progress meetings. Inclusion of a variety of assessments within a portfolio is helpful if they have been marked according to appropriate criteria and with effective feedback given as part of the overall formative assessment programme. The attraction of the portfolio as a means of assessment is strong – the collection of a series of pieces of work or assessments of competencies together with self-reflective pieces and evidence of professional development is a unique addition to the assessment opportunities open to medical educators. Portfolios are now increasingly used in formative assessment, and it may be possible to combine this with a summative element in the portfolio assessment [41, 42]. For a full discussion of portfolios in medical education see Chapter 18 in this book.

Conclusions

A programmatic approach to assessment within a curriculum that is constructively aligned and with staff and students who understand the learning outcomes and the goals of each assessment is desirable and can address all three of the overarching aims of assessment [43, 44]. In medical education, formative assessment is a valuable part of the assessment programme. A well-designed programme of formative assessment linked to overall curriculum aims and the teaching and learning goals of individual modules enhances the learning experience for students and promotes desirable educational outcomes, including learner self-regulation and the development of lifelong learning skills.

References

1 Newble, D.I. and Jaeger, K. (1983). The effect of assessments and examinations on the learning of medical students. *Medical Education* 17: 165–171.

2 Rust, C. (2002). The impact of assessment on student learning: how can research literature practically help to inform the development of departmental assessment strategies? *Active Learning in Higher Education* 3: 128–144.

3 Black, P. and William, D. (1998). Assessment and classroom learning. *Assessment in Education* 5: 7–74.

4 Black, P. and William, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability* 21: 5–31.

5 van der Vleuten, C. (1996). The assessment of professional competence: developments, research and practical implications. *Advances in Health Science Education* 1: 41–67.

6 Sadler, D.R. (1998). Formative assessment: revisiting the territory. *Assessment in Education* 5: 77–84.

7 Butler, D.L. and Winne, P.H. (1995). Feedback and self-regulated learning: a theoretical synthesis. *Review of Educational Research* 65: 245–281.

8 Cassidy, S. (2011). Self-regulated learning in higher education: identifying key component processes. *Studies in Higher Education* 36: 989–1000.

9 Nicol, D.J. and Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education* 31: 199–218.

10 Schuwirth, L.W.T. and van der Vleuten, C.P.M. (2004). Changing education, changing assessment, changing research? *Medical Education* 38: 805–812.

11 Rushton, A. (2005). Formative assessment: a key to deep learning? *Medical Teacher* 27: 509–513.

12 Wood, D.F. (2012). Formative assessment. In: *Oxford Textbook of Medical Education* (ed. K. Walsh), 478–488. Oxford: Oxford University Press.

13 Hattie, J.A. (1987). Identifying the salient factors of a model of student learning: a synthesis of meta-analyses. *International Journal of Educational Research* 11: 187–212.

14 Ramaprasad, A. (1983). On the definition of feedback. *Behavioural Science* 28: 4–13.

15 Sadler, D.R. (1989). Formative assessment and the design of instructional systems. *Instructional Science* 18: 119–144.

16 Sayer, M., Chaput de Saintonge, M., Evans, D., and Wood, D. (2002). Support for students with academic difficulties. *Medical Education* 36: 643–650.

17 Henderson, P., Ferguson-Smith, A.C., and Johnson, M.H. (2005). Developing essential professional skills: a framework for teaching and learning about feedback. *BMC Medical Education* 5: 11. https://bmcmededuc.biomedcentral.com/articles/10.1186/1472-6920-5-11 (accessed 28 June 2018).

18 Pendleton, D., Schofield, T., Tate, P., and Havelock, P. (1984). *The Consultation: An Approach to Learning and Teaching*. Oxford: Oxford University Press.

19 Silverman, J.D., Draper, J., and Kurtz, S.M. (1996). The Calgary–Cambridge approach to communication skills teaching. 1. Agenda led outcome-based analysis of the consultation. *Education for General Practice* 7: 288–299.

20 Neher, J.O., Gordon, K., Meyer, B., and Stevens, N. (1992). A five-step 'microskills' model of clinical teaching. *Journal of the American Board of Family Practice* 5: 419–424.

21 Brukner, H.A., Altkorn, D.L., Cook, S. et al. (1999). Giving effective feedback to medical students: a workshop for faculty and house staff. *Medical Teacher* 21: 161–165.

22 Wall, D. (2004). Giving feedback effectively. In: *Teaching Made Easy: A Manual for Health Professionals* (ed. K. Mohanna, D. Wall and R. Chambers), 147–158. Oxford: Radcliffe Medical Press.

23 Hewson, M.G. and Little, M.L. (1998). Giving feedback in medical education: verification of recommended techniques. *Journal of General Internal Medicine* 13: 111–116.

24 Ende, J. (1983). Feedback in clinical medical education. *Journal of the American Medical Association* 250: 777–781.

25 Kurtz, S., Silverman, J., and Draper, J. (2005). *Teaching and Learning Communication Skills in Medicine*, 123–129. Oxford: Radcliffe Publishing.

26 Roy Chowdhury, R. and Kalu, G. (2004). Learning to give feedback in medical education. *The Obstetrician and Gynaecologist* 6: 243–247.

27 Biggs, J.B. and Collis, K.F. (1982). *Evaluating the Quality of Learning: The SOLO Taxonomy*. New York: Academic Press.

28 Sender Liberman, A., Liberman, M., Steinart, Y. et al. (2005). Surgery residents and attending surgeons have different perceptions of feedback. *Medical Teacher* 27: 470–472.

29 Sargeant, J. and Mann, K. (2010). Feedback in medical education: skills for improving learner performance. In: *ABC of Teaching and Learning in Medicine* (ed. P. Cantillon and D.F. Wood), 29–32. Oxford: Wiley-Blackwell.

30 Ramani, S. and Krackov, S.K. (2012). Twelve tips for giving feedback effectively in the clinical environment. *Medical Teacher* 34: 787–791.

31 Moon, J.A. (2004). *A Handbook of Reflective and Experiential Learning*, 149–157. London: Routledge Falmer.

32 Biggs, J.B. (2003). *Teaching for Quality Learning at University*, 2e. Buckingham: Open University Press/SRHE.

33 Gitomer, D.H. and Duschl, R.A. (2007). Establishing multilevel coherence in assessment. *Yearbook of the National Society for the Study of Education* 106: 288–320.

34 Fowell, S.L., Southgate, L.J., and Bligh, J.G. (1999). Evaluating assessment: the missing link? *Medical Education* 33: 276–281.

35 Higgins, R.S.D., Bridges, J., Burke, J.M. et al. (2004). Implementing the ACGME general competencies in a cardiothoracic surgery residency program using 360-degree feedback. *Annals of Thoracic Surgery* 77: 12–17.

36 Denison, A.R., Currie, A.E., Laing, M.R., and Heys, S.D. (2006). Good for them or good for us? The role of academic guidance interviews. *Medical Education* 40: 1188–1191.

37 Daelmans, H.E.M., Hoogenboom, R.J.I., Scherpbier, A.J.J.A. et al. (2005). Effects of an in-training assessment programme on supervision of and feedback on competencies in an undergraduate internal medicine clerkship. *Medical Teacher* 27: 158–163.

38 McKinley, R.K., Fraser, R.C., van der Vleuten, C., and Hastings, A. (2000). Formative assessment of the consultation performance of medical students in the setting of general practice using a modified version of the Leicester assessment package. *Medical Education* 34: 573–579.

39 Snadden, D. (1999). Portfolios – attempting to measure the unmeasurable? *Medical Education* 33: 478–479.

40 Hays, R.B. (2004). Reflecting on learning portfolios. *Medical Education* 38: 800–804.

41 Friedman Ben David, M., Davis, M.H. et al. (2001). AMEE Medical Education Guide 24. Portfolios as a method of student assessment. *Medical Teacher* 23: 535–551.

42 Driessen, E., van der Vleuten, C., Schuwirth, L. et al. (2005). The use of qualitative research criteria for portfolio assessment as an alternative to reliability evaluation: a case study. *Medical Education* 39: 214–220.

43 Schuwirth, L.W. and van der Vleuten, C.P. (2011). Programmatic assessment: from assessment of learning to assessment for learning. *Medical Teacher* 33: 478–485.

44 van der Vleuten, C.P.M., Schuwirth, L.W.T., Driessen, E.W. et al. (2012). A model for programmatic assessment fit for purpose. *Medical Teacher* 34: 205–214.

Further Reading

Wood, D.F. (2012). Formative assessment. In: *Oxford Textbook of Medical Education* (ed. K. Walsh), 10–14. Oxford: Oxford University Press.

Schuwirth, L.W. and van der Vleuten, C.P. (2011). Programmatic assessment: from assessment of learning to assessment for learning. *Medical Teacher* 33: 478–485.

Sargeant, J. and Mann, K. (2010). Feedback in medical education: skills for improving learner performance. In: *ABC of Teaching and Learning in Medicine* (ed. P. Cantillon and D.F. Wood), 29–32. Oxford: Wiley-Blackwell.

26 Selection into Medical Education and Training

Fiona Patterson[1,2], *Eamonn Ferguson*[3], *and Lara Zibarras*[2,4]
[1]University of Cambridge, Cambridge, UK
[2]Work Psychology Group, London, UK
[3]University of Nottingham, Nottingham, UK
[4]City, University of London, London, UK

KEY MESSAGES

- Historically, medical school admissions have relied primarily on indicators of academic achievement to select students.
- There is relatively little research investigating selection issues in medical education and training other than the prediction of exam performance.
- Assessment for selection is significantly different from summative assessments of education or training, such as written examinations.
- Best practice selection involves a thorough job analysis and the use of evidence from validation studies to drive continual improvement of accuracy and fairness.
- Establishing the predictive validity of a selection method presents many conceptual and practical problems. The validation process may take several years and piloting is essential.
- At postgraduate level, selection ratios for specialties differ widely. This has implications for the design of specific selection systems.

Introduction

After an individual submits his or her medical school application, selection is the first assessment within the medical education and training pathway. The intention is to predict who will become a competent clinician at the outset of a long developmental journey. With limited student places and large numbers of applicants, medical schools have tended to rely on academic criteria, with the assumption being that with high academic ability, the other skills and attributes required to be a competent clinician are trainable. However, along with academic ability, medical students must have other important skills, values, and personal qualities (e.g. compassion). These constructs are often categorised into *cognitive skills* (i.e. academic ability, clinical knowledge) and *non-cognitive skills* (i.e. personal qualities such as empathy, communication, integrity). Conceptually, a key issue is whether medical schools should aim to select individuals who will make successful students or those who will make competent clinicians. Clearly, success as a student and competence as a clinician are not mutually exclusive, but the former is not necessarily a precursor of the latter.

Research suggests that medical school selection criteria vary between schools both intra- and internationally. This diversity of entry criteria is at odds with recent job analysis research suggesting that there is commonality in the knowledge, skills, and attributes required to be a competent clinician, irrespective of the specialty practised [1].

Context of Medical Selection

The assessment paradigm used to understand a selection context is different to that of professional examinations. In examinations, the aim is to assess end-of-training capability, where judgements are made by trained examiners about an individual's capacity to perform a job with competence. In theory, all candidates can pass the assessment. By contrast, in selection settings, if the number of candidates outweighs the number of available posts, then the assessment is geared towards ranking individuals. If the competition is very high, competent candidates may not be awarded a post.

Assessment in selection uses a 'predictivist paradigm', where the intention is to *predict* who will become a competent clinician (i.e. to identify those individuals who will successfully complete training, *before* training commences). Although there are several similarities, the parameters for designing and validating a robust selection system and selection methods are different from other assessment settings. Importantly, the criteria used to judge the effectiveness of a selection system are potentially more complex.

In evaluating professional exams, the *reliability* of the assessments is viewed as the 'gold standard'; to ensure a trainee is safe for subsequent independent practice. In selection, the *predictive validity* of the assessments is the gold

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

standard, as students and trainees enter supervised education, and recruiters wish to appoint those most likely to succeed in training. Selection research in health care has tended to focus largely on reliability (e.g. how many stations are required for a multiple mini-interview process to be reliable), and here we remind readers that it is quite possible to be *reliably wrong*, such that greater attention should be paid to establishing the predictive and construct validity of selection systems.

Internationally, selection into medicine continues to be highly competitive. Practically, this means that it can be highly resource intensive, resulting in practical challenges for recruiters. Medical selection is also 'high-stakes' since compared to many other professions, the length of training required to practise as a clinician is long and costly, normally exceeding 15 years from medical school through to senior appointment. When reviewing selection issues within the training pathway, a distinction must be drawn between selection into medical school (*pre-employment*) and postgraduate training (*employment*). Importantly, the latter is governed by specific employment law, for which there are also significant international differences [2]. In addition, selecting the wrong person for a job can have serious consequences for an organisation, the employee involved, and, perhaps most importantly in medicine, the patient.

As a result, depending on the methods used for selection, there can be a major risk of litigation if these are perceived to be unfair by key stakeholders. As such, Patterson and colleagues argue that the design of selection systems should account for *political validity* – the reactions of stakeholders to the criteria and methods used [3, 4]. This includes the reactions of applicants and recruiters, but also those of wider stakeholders, such as government, regulatory bodies, and the general public, who play an important role in decision-making at policy level.

In this chapter we first outline the key concepts associated with selection processes and the relative accuracy of selection methods for medical education and training. We discuss why medicine provides a unique occupational context and summarise international perspectives on selection practices, referencing both undergraduate and postgraduate training. Key concepts underpinning selection research are described and we summarise the research evidence on the reliability and validity of current selection methods. Finally, some considerations for a future research agenda are presented.

Key Concepts

Selection Process

Figure 26.1 summarises the main elements of designing and implementing a selection process. This starts by conducting a thorough analysis of the relevant knowledge,

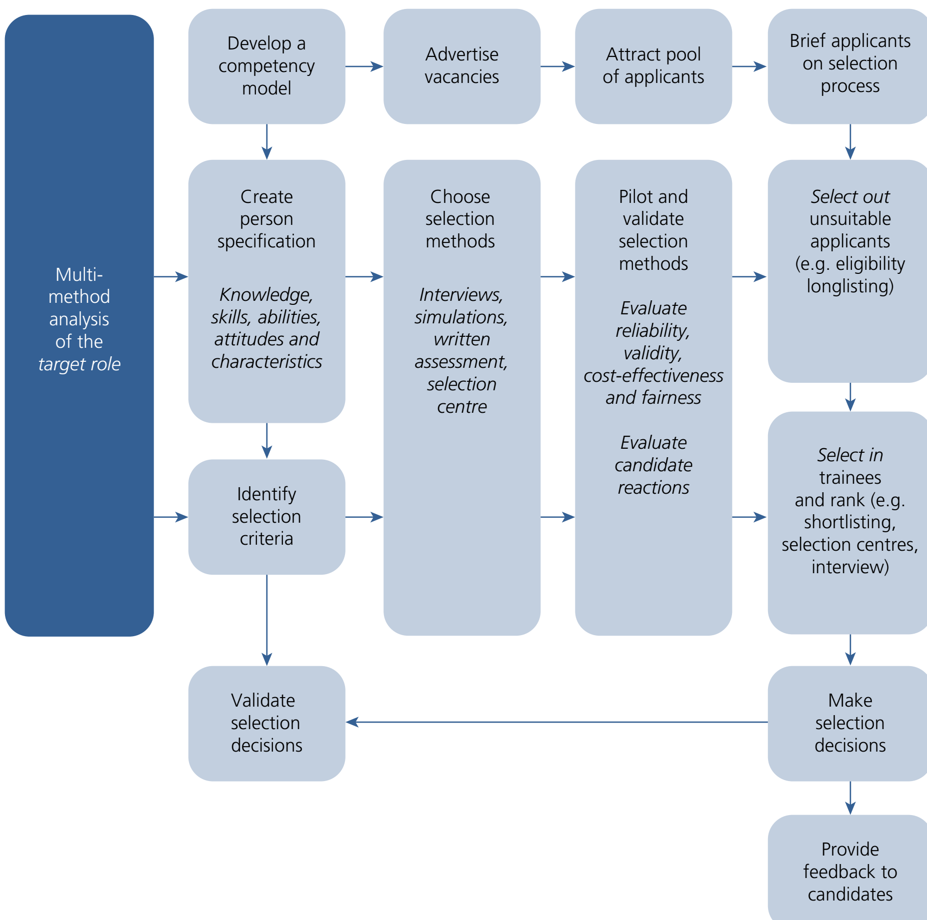


Figure 26.1 The process of selection.

skills, abilities, and attitudes associated with performance in the target role. This information is used to construct a person specification (and job description where appropriate) and used to decide which selection methods will elicit applicant behaviour related to the selection criteria. Outputs from this analysis should detail the responsibilities in the target job and provide information about the competencies and characteristics required of the job holder.

In deciding to apply for a post (or a place at medical school), applicants engage in *self-selection* where they make an informed judgement about whether the role suits their skills, abilities, and values. Once selection decisions are made and accepted applicants enter training, the trainees' performance on the original selection criteria should be used to examine the predictive validity of the selection methods (i.e. to what extent are scores at selection associated with assessment of in-training and work performance?).

Figure 26.1 also shows that best practice selection is a 'two-way' selection process. To attract the best trainees, both medical schools and hospitals have become increasingly aware that candidates' reactions to the selection process are important particularly in relation to perceptions of fairness. Since large resources are often spent on selection procedures, the utility of the selection procedures should be evaluated. In addition, information collected at selection (i.e. entry point to training) can be used to design tailored development plans for trainees.

The rudiments of best practice selection are clear, yet research shows that two elements in the process are often not conducted effectively. First, many organisations do not conduct thorough job analyses to identify the key knowledge, skills, and behaviours associated with competent performance in target job roles. This is particularly important when exploring potential differences between medical specialties. Second, validation studies are rarely conducted in organisations as they are time-consuming and difficult to administer. It often means tracking trainees' performance over several years, from medical school selection through to senior posts. In medical education and training, far more validation research has occurred in undergraduate selection, exploring the predictive validity of various cognitive factors (prior academic performance or knowledge tests) with respect to exam performance [5]. The criteria used to judge performance at medical school are more readily observed as there are standardised assessments involved, such as examinations. By contrast, the research literature is relatively sparse when considering selection for either postgraduate training or non-cognitive factors [5].

In summary, research demonstrates that best practice selection is an iterative process. Results from evaluation and validation studies should be used to review the original selection criteria and the choice of selection methods to make continual improvements to enhance accuracy and fairness of selection systems.

Evaluation Criteria

Before judging how well selection methods work, it is necessary to understand the framework used for determining best practice. Box 26.1 lists criteria for judging the 'quality' of selection procedures which should be reviewed when designing and implementing selection systems.

BOX 26.1 Evaluation criteria for selection procedures

1. Reliability and validity of selection tools
2. Employee/candidate reactions
3. Ease of interpretation
4. Generality of use
5. Cost and value
6. Practicality/administrative convenience
7. Legality
8. Fairness
9. Educational impact
10. Mechanisms for generating feedback
11. Arrangements for future validation, evaluation, and development
12. Ongoing availability of analytical expertise

When choosing the selection method(s) it is important that the output (score) is consistent/stable (*reliable*) and relevant/accurate (*valid*), and that the method is *objective*, *standardised*, administered by trained professional(s), and monitored. Evaluation of the system is essential to ensure that selection tools are also fair, defensible, cost-effective, and feasible. Feedback is used to make continual improvements to the selection system to enhance accuracy and fairness. For postgraduate training there are legal reasons for ensuring accurate selection procedures are used, as is essential for compliance with employment law.

Validity

No single validation study will provide a definitive answer regarding the validity of a selection method. Each validation study is conducted on a sample of relevant people at a particular point in time, using a specific selection method. A specific challenge, especially for validating selection methods that focus on non-academic selection criteria, is in identifying appropriate outcome variables. However, certain factors such as the sampling, the methods, the timing of the study and so on, will influence the results in some way. Applying statistical models such as generalisability theory allows for the contribution of such components to the overall effect size to be estimated [6]. To estimate the validity of a particular selection method, more than one study design is needed, to minimise the error. Most selection systems combine several predictors (selection methods), such as an applicant's score on an interview and academic achievements. In this respect, a major challenge in validity research for selection is that medical schools tend to weight selection methods differently and so conducting large scale longitudinal validation studies (rather than smaller single-site studies) can be problematic.

In validation studies, a key question is how much does adding another predictor (i.e. selection method) increase the predictive power of the selection process? This is known as *incremental validity*. Specifically, recruiters may want to know how accuracy is improved, for example due to using

BOX 26.2 FOCUS ON: Validity in selection

Faith validity	This is 'blind' faith that a selection method works because someone plausible said so.
Face validity	The selection tool content appears relevant to the target role (determined by the applicants).
Content validity	The content of the selection tool is judged to be directly relevant to the target role by subject matter experts.
Criterion validity: *Concurrent*	A form of criterion-related validity in which data on the predictor and criterion are obtained at the same time. High correlations between predictor and criterion scores indicate concurrent validity.
Criterion validity: *Predictive*	This is the extent to which a predictor measure (e.g. a selection test score) is correlated to a criterion measure (e.g. work performance). High predictive validity indicates that a selection measure gives an accurate indication of candidates' future performance on the criterion.
Incremental validity	This is an empirical issue to determine how much additional value using another assessment provides.
Construct validity	An indication of the extent to which the test or procedure measures the construct that it is intended to measure (such as empathy, clinical expertise).
Political validity	An indication of the extent to which various stakeholders and stakeholder groups (such as employers, parents, government departments, society, the regulator) consider the tool(s) to be appropriate and acceptable for use in selection.

a personality assessment rather than relying solely on interview scores. Information on the incremental validity of a specific selection method is valuable as it allows organisations to conduct a cost–benefit analysis of using additional tools. Box 26.2 provides a list of the different forms of validity for reference.

Predictive Validity and the 'Criterion Problem'

The way to collect criterion-related validity data, i.e. how well scores on the selection method predict some future outcome or criterion, is to use a predictive (or follow-up) design. This design involves collecting predictor information (e.g. interview ratings, test scores) for candidates and then following up for performance data (e.g. during their first year of employment or exams at medical school). Predictive validity is assessed by examining the correlation between scores at selection (Time 1) and criterion data collected at Time 2 (perhaps through relevant work-based assessments, examinations, etc.). It is unusual in field studies to obtain validity coefficients over $r = 0.5$ [7].

Conducting validation studies in practice presents some problems. One major problem regards accessing the appropriate criterion (outcome) data to validate the selection process. Often the criteria used to measure performance in the job role do not match the criteria used for selection. Conversely, sometimes the criterion and predictor are very similar (e.g. using the Medical College Admission Test or other knowledge-based test to predict exam performance in medical school), which may lead to common method variance and content overlap. Ideally, predictor scores should only be used to make selection decisions *after* a predictive validation study has been conducted. Practically, this is difficult to achieve so piloting is essential to conduct an appropriate validation. Box 26.3 presents three sources of error that are important to consider when conducting validation studies in selection – sampling, measurement precision, and restriction of range issues. This is not intended as an exhaustive list of sources of error, which also includes issues such as selection bias, reverse causation, and missing variable problems.

Candidate Reactions

Candidate reactions to different recruitment methods are critically important [8]. Considerable research has determined applicants' views on selection methods and has explained the different factors that influence applicant reactions using organisational theories of justice.

Distributive justice focuses on perceived fairness regarding equity (whether the selection outcome is consistent with the applicant's expectation) and equality (the extent to which applicants have the same opportunities in the selection process). *Procedural justice* refers to the formal characteristics of the selection process, such as information and feedback offered, job-relatedness of the procedures and methods, and recruiter effectiveness [9]. Four main factors seem to account for positive applicant reactions, where selection methods: (i) are based on a thorough job analysis and appear job relevant, (ii) are not personally intrusive, (iii) do not contravene procedural or distributive justice expectations, and (iv) allow applicants to meet in person with the recruiters. Other research shows that applicants prefer multiple opportunities to demonstrate their skills (as in selection centres) and prefer selection systems that are administered consistently for all applicants. In particular, when competition ratios are high, applicant reactions and candidate expectations of 'fair play' are crucial.

BOX 26.3 FOCUS ON: Sources of error in validation studies

Sampling error

If relatively small samples are used in many validation studies, the results obtained may be unduly influenced by the effects of small numbers of people within the sample with unusual results. As sample size increases, more reliable results are obtained.

Poor measurement precision

The measurement of attributes at both the predictor (i.e. selection method) and criterion (i.e. job performance) stage of the validation process is subject to unsystematic error. This error (unreliability) in the scores will reduce the observed correlation between predictor and criterion. This means that as reliability decreases, the maximum possible correlation between predictor and criterion will decrease.

Restricted range of scores

The sample used in a validation study may not provide the full possible range of scores on the predictor and/or criterion measures. A restricted range of scores limits the size of the linear correlation between two variables. So, like unreliability, range restriction in a sample reduces the magnitude of the observed correlation coefficient.

Fairness

Fair selection and recruitment is based on: (i) having objective and valid criteria (developed through an appropriate job analysis), (ii) accurate and standardised assessment by trained personnel, and (iii) monitored outcomes. Research has explored the extent to which selection procedures are fair to different subgroups of the population (such as ethnic minorities or women [10]). However, a test is not unfair or biased simply because members of different subgroups obtain different scores on the tests. Men and women have different mean scores for height: this does not mean that rulers are unfair measuring instruments. However, it would be unfair to use height as a selection criterion for a job, if the job could be done by people of any height, since it is important for selection criteria to be job related. Normally, the extent to which a selection method is related to job performance is estimated by validation research, and it is clear, therefore, that fairness and validity are closely related.

Values-based Recruitment

Values-based recruitment (VBR) involves attracting and recruiting students, trainees, and employees based on the extent to which their values align with the values of the organisation in which they work [11]. Compassion, benevolence, respect, and dignity are important for any health care professional to ensure the provision of high-quality care and patient outcomes. This concept has international relevance, and in the UK has been explored extensively following two government enquiries [12, 13] which highlighted major concerns about compassionate care (or lack thereof) within health care roles. The workforce plays a vital role in delivering safe and compassionate care, and values are important in facilitating this. The purpose of VBR is to ensure that people are recruited who have the right skills and the right values to support effective delivery of high-quality patient care and outcomes.

Selection Methods

Researchers have reviewed the use of numerous selection methods across several different occupational groups [14]. Given the multifaceted nature of the role of a doctor, recruiters are likely to use multiple selection methods to assess applicants. Therefore, recruiters must decide whether a job applicant must score highly on all selection criteria (non-compensatory) or whether high scores on some criteria can make up for low scores on another (compensatory). In practice, recruiters might assign different weightings to various selection criteria, depending on the nature of the job role. For example, if clinical knowledge is the most important criterion and an applicant does not achieve a certain score, their application may not be considered further.

Although many of these methods have been piloted for selection into medicine, the interview tends to be the most common method used in both undergraduate and postgraduate selection. In the following sections we provide an overview of the research evidence in relation to the most common selection methods.

Interviews

Interviews are ubiquitous in the selection processes of a variety of professions [15]. They can be used at different stages of the selection process, either as the sole method of selection, or in conjunction with other methods. Interviews vary in terms of (i) purpose, (ii) duration, (iii) mode of administration (telephone, face-to-face, or video conference), (iv) number of interviewers (one-to-one or panel), (v) degree of structure (unstructured, semi-structured, or structured), and (vi) number of sessions (single or multiple). Research consistently shows that structured interviews tend to have much higher reliability and criterion-related validity than unstructured interviews, when they are based on thorough role analysis and have validated scoring criteria [16–19]. Box 26.4 provides a summary of best practice in designing structured interviews.

BOX 26.4 HOW TO: Run a structured interview

- Relate questions to the person specification (based on a thorough job analysis)
- Ask the same questions of each candidate, limit prompting, and use follow-up and probing questions to elicit evidence
- Use relevant questions and design as either situational, competency-based, biographical, or knowledge questions
- Use longer interviews, or a larger number of questions, to control the input of ancillary information
- Do not allow questions from the candidate until after the interview (when the information to make selection decisions has been collected)
- Rate each answer and use standardised rating scales (increase specificity)
- Use detailed anchored rating scales and take detailed notes
- Use multiple interviewers where possible (but ensure efficiency)

Meta-analytical studies (statistical combination of results from various studies to identify generalisable findings) have found structured interviews to be valid predictors of job performance [15]. Research evidence also suggests that structured interviews have incremental validity over cognitive ability tests [20] and they generally yield small ethnic group differences [21]. Adding structure to an interview may also increase the chances of an organisation successfully defending a lawsuit [22]. A limitation of structured interviews involving multiple assessors is that inter-rater reliability may be modest where assessors are not properly trained; yet when properly trained and when using standardised questions and validated scoring criteria, the interview has an acceptable level of inter-rater reliability [23, 24].

Unstructured interviews are still widely used in many countries for employee selection, despite their low reliability, low predictive validity, and poor legal defensibility [25]. Unstructured interviews are prone to bias and error, including: (i) stereotyping, (ii) making a judgement solely on first impressions rather than allowing all candidates the chance to demonstrate their skills (e.g. 'I know if he or she is the right person immediately'), (iii) halo and horns effects (i.e. selectors being unduly influenced by one positive or negative characteristic of the applicant), and (iv) similar-to-me bias, where interviewers rate most favourably interviewees who are similar to themselves [26]. Moreover, unstructured interviews may assess different characteristics for different candidates, meaning the content validity can be variable [27]. All of these factors are likely to distort interviewers' ratings of candidates.

In recent years, Multiple Mini-Interviews (MMIs) have become increasingly popular in medical selection. The MMI is an interview format comprising multiple stations, which is based on the format of the Objective Structured Clinical Examination (OSCE). MMIs typically involve a one-to-one interview, as well as role play and interactive tasks, focused on a range of domains and lasting for 5–10 minutes each [28]. Some evidence suggests that graduate and female applicants may outperform school leavers and male applicants on MMIs, respectively [29]. However, overall the research on MMIs suggests they often have a good level of reliability [29, 30]. Accordingly, MMIs have now been incorporated into medical school and postgraduate medical selection systems in many countries [30–33]. Moreover, research suggests that candidate reactions to MMIs are favourable [29, 34, 35].

However, some evidence suggests that interview methods may not always be strong predictors of performance at medical school [36]. Other research suggests that performance on MMIs offers good predictive validity in relation to licensure examinations [29, 37–39]. Currently, the literature provides support for the validity, reliability, and acceptability of MMIs and some structured interviews in medical selection, but comparatively less evidence supports unstructured interview methods. Further research on MMIs may be necessary to assess group differences in performance and the relative financial feasibility of the selection method as interviews in general are a relatively resource-intensive selection method.

References and Referee Reports

Large-scale empirical studies consistently show that references tend to be unreliable and ineffective at predicting job performance [40–42]. Despite this, references are widely used in selection across occupations, including medicine, and it is likely that they will continue to be used as an additional guide [43]. In practice, employers tend to value references, but they can be poor at differentiating between candidates fairly. Research on the content of references suggests that the writers of reports tend to apply positive and negative attributions homogenously across applicants, making it impossible for admissions committees to differentiate between applicants on the basis of these data [36].

A 2006 study found that the vast majority of medical schools in England used referees' reports as part of their selection process [38]. However, their reliability is questionable given changes in data legislation, which removed the confidentiality that existed previously [44]. In studying predictive validity, Ferguson et al. [41] showed that references obtained though UCAS (the central organisation that processes applications for full-time undergraduate courses at UK colleges and universities) did not predict preclinical or clinical performance. However, medical schools differ in terms of the weight they place on references obtained through the UCAS application. Some medical schools may actually ignore information contained in referees' reports for fear of unduly biasing selection decisions [38]. Despite the limitations of using referees' reports, they remain widespread in medical student selection [38]. Moreover, despite evidence of poor reliability and validity, referees' reports may still be viewed positively by some medical selection professionals [45]. This finding contradicts the prevailing opinion among researchers in the field that referees' reports are not of use in medical student selection [38, 42, 45].

Personal Statements and Autobiographical Submissions

Personal statements, essays, and other autobiographical submissions are often included in application forms, as an alternative to curriculum vitae, to facilitate shortlisting of candidates. The information obtained through application forms is collected systematically, allowing employers to assess objectively the candidates' suitability for a given post and make fair comparisons across applicants. Application forms may include questions on biographical information, educational background, previous work experience, and competencies identified through a job analysis. Some research evidence suggests that medical school application forms may be predictive of subsequent performance [46, 47]. However, other evidence suggests that application forms have low reliability compared to other selection methods [48]. Further research suggests that application forms are not predictive of performance in the clinical aspects of medical training or performance at medical school overall [41, 49]. The reliability and validity of application forms not completed under invigilated examination conditions may be contaminated by factors such as the length of time spent completing the form, and the potential influence and assistance of third parties. Therefore, application forms are not likely to reflect the medical school candidates' ability as well as other methods; a conclusion supported by research showing applicants present themselves in ways they perceive to be desirable but not necessarily accurate [50, 51]. These findings contrast with the intended function of application forms, which is to provide objective data to make selection decisions.

Academic Records

Academic criteria are a major component of selection to medical school in most countries. In the UK, for example, selection for admission to medical school is based on predicted or actual A-level results (a school-end examination designed to assess knowledge in various subjects, usually taken at age 18 years). One problem with using A-level grades for selection is in discriminating between students who obtain similarly high results [52]. Another concern is that medical school entry is socially exclusive, partly because A-level results might reflect type of schooling and 'social class' [53]. Research also suggests that predicted A-level grades may be inaccurate in more than 50% of cases [54]. In the USA and Canada, students apply to medical school at postgraduate level (graduate entry). However, academic grades such as Grade Point Average (GPA) remain the main criterion for selection, although they are usually considered in combination with other predictors, such as aptitude tests.

Some authors have shown that academic criteria such as A-level grades correlate with drop-out rates, career progression, postgraduate membership, and fellowship exams [5, 55–57]. These findings contrast with earlier studies that questioned the long-term predictive validity of academic records [58]. Whilst pre-admission academic grades such as A-level or GPA are undoubtedly related to academic performance at medical school, their relationship with long-term outcome measures of a clinicians' performance is less obvious, partly because of the 'criterion problem', discussed earlier [59]. Candidates selected purely on high academic performance are much more likely to drop out than candidates selected using a series of selection methods aimed at exploring commitment to studying medicine [60]. Therefore, using academic records to select candidates into medicine is complicated by a number of factors, including grade inflation, bias towards higher socio-economic classes, and an uncertain relationship between academic attainment and subsequent performance as a doctor.

General Mental Ability and Aptitude Tests

Tests of general mental ability (GMA) and specific cognitive abilities (e.g. numerical, verbal, and spatial reasoning) are increasingly popular in selection both in the USA and in the UK [61, 62]. Internationally, GMA and cognitive ability tests are robust predictors of job performance and training success across a range of occupations [63, 64]. However, there are concerns regarding fairness since GMA tests can produce adverse impact, with marked racial differences in test performance [65]. Specific ability tests tend to show smaller group differences [66].

Aptitude tests are standardised tests designed to measure a person's ability to develop skills or acquire knowledge. They are used to predict future performance in a given activity. Like GMA tests, aptitude tests measure an individual's overall performance across a broad range of mental abilities. In addition, aptitude tests also often include items that measure more specialised abilities (such as verbal and numerical skills).

Aptitude tests that include specific ability tests and a knowledge component are increasingly popular in medicine. In the UK, concerns over the discriminatory power of A-levels led to the introduction of additional selection methods such as specific medical knowledge tests [52] and intellectual aptitude tests (e.g. the Oxford Medicine Admission test). The UKCAT, comprising reasoning and decision-making tests and situational judgement, is now used in selection by the majority of medical and dental schools in the UK. The use of aptitude tests for medical school selection is also increasing in several other countries [33]. The outlook is somewhat different at postgraduate level where aptitude tests are rarely (if at all) used. This is not surprising given that most applicants have already passed an aptitude test for entry into medical school. At this stage, cognitive ability is a necessary but not a sufficient condition to predict who will be a competent physician.

In a selection context – especially with respect to widening access – it is important to distinguish between GMA in terms of *crystallised* intelligence (i.e. knowledge-based acquired via schooling) and *fluid* intelligence (i.e. biologically based cognitive skills such as processing speed, inductive reasoning, etc.) [67, 68] It is argued that tests of fluid intelligence should be used in medical school selection to widen access (i.e. to identify 'raw talent' independent of education). However, there are problems with this approach because commonly used tests of 'intelligence' assess crystallised intelligence [68], which may be more related to type of schooling. In addition, the assessment of 'raw talent' would have to be combined with an assessment of desire to study medicine and other non-cognitive skills (determined by a job analysis), otherwise students may fail, drop out, or under-perform at medical school.

Research evidence suggests that the UKCAT has predictive validity in relation to performance in the first two years of medical school [69]. Similarly, the MCAT, which is widely employed in North America, has significant predictive validity [70, 71]. Finally, the BioMedical Admissions Test, used by some UK universities to select medical students, has predictive validity in relation to subsequent performance [72]. Despite evidence of predictive validity, the strength of the predictive relationship between aptitude tests and subsequent clinical performance is relatively weak [70, 71, 73]. High-quality, longitudinal studies are required to examine the precise relationship between scores on selection aptitude tests and subsequent performance in medical school and as a clinician.

Personality Inventories

The last 20 years have seen a substantial increase in the use of personality assessments in personnel selection for a range of jobs [74]. Personality assessments are generally self-report inventories where candidates respond to statements by rating the extent to which they agree, or by indicating how accurate an item is as a description of their personality. Over many decades of research, researchers have agreed a general taxonomy of personality traits, the 'Big Five' model, which is based on five factors or traits: *extraversion* (i.e. outgoing, sociable, impulsive), *emotional stability* (i.e. calm, relaxed), *agreeableness* (i.e. trusting, cooperative, helpful), *conscientiousness* (i.e. hardworking, dutiful, organised) and *openness to experience* (i.e. artistic, cultured, creative).

Research shows important relationships between measures of personality and job or academic performance [74]. For example, personality traits defined as 'dysfunctional' are significantly associated with negative outcomes for medical students, such as lower academic performance [75]. Other researchers have found personality measures to be a useful addition to medical school selection and is predictive of performance at medical school [76]. Significant associations have been reported between some Big Five personality traits and performance across various aspects of medical school performance [77]. Conscientiousness, for example, is a positive predictor of preclinical knowledge and exam results [41, 49, 78] and offers incremental validity over knowledge-based assessments [41, 49]. However, while positively associated with preclinical knowledge, conscientiousness is also a significant negative predictor of clinical skill [41, 79]. Therefore, the relationship between personality traits and performance in medical education and training may be complex, possibly non-linear and changes with the dynamic changing nature of the job [79, 80]. This implies that a trait used to select at one time point may not be predictive of all aspects of later job performance. Thus again the difference between selecting for a medical student or clinician are not necessarily the same thing as the skill and ability requirements necessarily change between being a student versus being a practising clinician.

Other researchers have presented evidence suggesting that personality measures were not useful predictors of medical school performance. For example, no significant association was reported between the Myers–Briggs Type Indicator and performance on the MCAT [81]. Similarly, further research showed the Personal Qualities Assessment (PQA) was not correlated with success as a medical student [82]. These mixed findings can be considered with findings that certain personality characteristics may have differential costs and benefits over time. For example, evidence shows the validity of personality measures in predicting medical school grades increases over the course of medical education and training [83]. For example, there may not be advantages to being open and extraverted for early academic performance, yet these traits gain importance for later academic performance [79, 84]. Therefore, earlier studies may have under(over)-estimated the predictive value of some personality traits. Furthermore, there is no consensus among researchers about whether conscientiousness is increasingly advantageous or disadvantageous over the course of medical education and training, with research presenting contradictory findings [79].

The use of personality assessment for job applicants remains controversial. Critics argue that the predictive validity of personality traits for job performance is often low and badly understood [85]. Further, personality tests used by organisations are often poorly chosen [86], and 'faking' can compromise the validity of personality tests [87, 88]. However, there is also evidence to suggest that faking or socially desirable responding does not compromise the predictive validity of personality tests [89]. In medicine, concerns over the strong reliance on academic predictors have led to the search for alternative selection methods. Specifically, there is a growing interest in the role of personality in selection at undergraduate level. Nevertheless, best practice suggests that personality assessment should be used to drive questioning within an interview, and should not be used in isolation to make selection decisions.

Selection Centres

Selection centres (SCs), also known as assessment centres, are a selection method used widely in non-medical selection contexts. They involve a combination of selection methods, such as written exercises, interviews, and work simulations, to assess candidates across several key skills, attitudes, and behaviours (e.g. empathy; as identified in the job analysis). Candidates are assessed in groups or individually by multiple assessors.

The SC is different from an OSCE. In an OSCE, each station assesses a candidate on *one* key skill, usually observed by one assessor. By contrast, the SC includes *multiple* situations (interview, work simulation, written exercise, etc.) where candidates demonstrate a key skill, observed by several trained assessors. Thus, a fairer (multiple opportunities to perform) and more reliable (multiple observations of behaviours by multiple observers) assessment can be made. With careful design, the increased reliability results in greater validity and more positive candidate reactions.

SCs have become widely used as a tool for recruitment [61]. SCs are especially popular for graduate recruitment, and only recently has this approach been used in medicine [90]. In the UK, Patterson et al. have pioneered the use of SCs, initially in the selection of general practitioners, and results have shown good predictive validity [90]. This work has been extended to select doctors for postgraduate training in other specialties such as for obstetrics and gynaecology and for paediatrics [91, 92]. SCs have been piloted in the UK for graduate entry to medical school [93], and also for medical student selection internationally, with positive reports on the reliability and internal validity of the method [33, 94, 95]. Evidence is also emerging for the predictive validity of SCs when used in medical specialty training [96], but further research is needed on the predictive validity of SCs in medical student selection. Such research may be constrained by the financial cost and logistical complexity of implementing SCs [94].

A carefully designed and administered SC can be effective at predicting job performance across a wide range of occupations [97, 98]. Gains are made in reliability and validity because SCs make use of a combination of different exercises (using a multi-trait, multi-method approach) and use standardised scoring systems to measure the selection criteria. Scoring should be directly linked to the selection criteria (not the exercise scores) and the information gathered should be interpreted in context by appropriately trained assessors. Unfortunately, many fail to understand this fundamental difference between OSCE-style examinations and SCs in the selection context. Well-executed SCs have incremental validity over cognitive ability tests [99, 100], and tend to be viewed positively by candidates [9]. Careful design and implementation is crucial for the SC to live up to its reputation and to be cost-effective [101].

Situational Judgement Tests

Situational judgement tests (SJTs) are a measurement methodology designed to measure candidates' judgement in role-relevant settings (see Box 26.5). SJTs can be constructed in a variety of formats but in general they present candidates with a scenario and a list of possible responses. The candidate is asked to consider the situation and make judgements about the possible responses. The candidates' responses are scored against expert responses. An example of a SJT item used for health care selection purposes is displayed in Box 26.6. For a review of the research evidence relating to the use of SJTs and their relevance for selection into the health care professions we direct readers to Patterson et al.'s recent review [102].

In the UK, the use of SJTs in medical selection and assessment has become widespread. Numerous high quality meta-analyses, review articles, and cross-sectional studies have been published that assess the effectiveness of SJTs [103–109]. These studies suggest SJTs have criterion validity and incremental validity over academic ability and personality assessment. Moreover, evidence suggests SJTs have low adverse impact against minority groups, and are perceived favourably by candidates.

Other researchers have reported procedural issues in the use of SJTs in medical student selection. For example, the mode of administration may impact on the validity of an SJT [110]. Similarly, the response instructions included in an SJT and the construction of different SJT forms may impact

BOX 26.5 FOCUS ON: Situational judgement tests in medical selection

In 2012, a review of the methods by which medical students are selected into UK foundation training recommended updates to the selection methods used. Although the existing system worked well, several concerns were raised regarding the use of personal statements on the application from; including the low reliability of personal statements, a lack of adequate standardisation, plus the risk of plagiarism and costs in terms of the time it takes to score them.

The recommendations were to design an SJT to assess several non-academic attributes and employability for a training post (to replace the application form questions and personal statement); and to use the SJT in conjunction with a measure of educational performance to assess academic competence, clinical knowledge, and skills.

Pilots of the SJT involving more than 1000 UK students found it a valid and reliable method of selection in this context. The SJT was based on a multi-method job analysis and was developed in consultation with junior doctors and with clinicians who work with junior doctors. This was to ensure that the scenarios were relevant, realistic, and fair. The SJT targets five professional attributes: commitment to professionalism, coping with pressure, effective communication, patient focus, and working effectively as part of team. The test was launched successfully for live recruitment in 2013 and further information can be found at www.isfp.org.uk

In the SJT there are two item formats:

1. Rank five possible responses in the most appropriate order.
2. Select the three most appropriate responses for the situation.

The choice of response options reflects the scenario content. For example, the nature of some scenarios and the possible responses lend themselves to ranking items (the ability to differentiate between singular actions that vary in appropriateness in response to a scenario), whereas some scenarios lend themselves to multiple-choice items (where it is necessary to tackle more than one aspect in response to a scenario).

Applicants must answer what they 'should' do in the scenario described, not what they 'would' do. This is because SJT research shows that questions asking an applicant what they 'would' do are more susceptible to coaching.

BOX 26.6 Situational judgement test: Question example

You review a patient on the surgical ward who has had an appendectomy done earlier in the day. You write a prescription for strong painkillers. The staff nurse challenges your decision and refuses to give the medication to the patient.

Choose the THREE most appropriate actions to take in this situation:

A Instruct the nurse to give the medication to the patient
B Discuss with the nurse why she disagrees with the prescription
C Ask a senior colleague for advice
D Complete a clinical incident form
E Cancel the prescription on the nurse's advice
F Arrange to speak to the nurse later to discuss your working relationship
G Write in the medical notes that the nurse has declined to give the medication
H Review the case again

Answer: B C H

Rationale: Ensuring patient safety is key to this scenario. It is important that the nurse's decision is discussed with her as there may be something that was missed when first reviewing the patient (B). Therefore, it would also be important to review the patient again (H). Also, relating to this is the importance of respecting the views of colleagues and maintaining working relationships, even if there is disagreement. As there has been a disagreement regarding patient care, it is important to seek advice from a senior colleague (C).

BOX 26.7 WHERE'S THE EVIDENCE: Effectiveness of selection methods [25]

Selection method	Reliability	Validity	Candidate acceptability	Promotes widening access
Academic records	High	High	High	Low
Structured interviews/MMIs	Moderate to high	Moderate to high	High	Moderate
Situational judgement tests	High	High	Moderate to high	High
Aptitude testing	High	Various	Moderate	Moderate
Personality tests	High	Moderate	Low to moderate	N/A
Traditional interviews	Low	Low	High	Low
Personal statements	Low	Low	High	Low
References	Low	Low	High	Low

Source: Patterson et al. [14].

on their validity [111]. Other researchers have presented contradictory evidence on the susceptibility of SJTs to coaching, faking, and practice effects [104, 112]. However, it is important to note that SJTs are a measurement method and that they can be designed to minimise coaching effects in high-stakes selection settings such as in medicine.

From a practical perspective, research suggests that SJTs can often usefully and feasibly be incorporated into existing selection systems [103, 113, 114]. SJTs tend to have high face validity, and are rated favourably by candidates [9, 107, 115] and evidence suggests they are a cost-effective option in medical selection [104]. In summary, using SJTs in medical student selection is supported by the weight of published research on this selection method. Although the research base is relatively small compared to more established selection methods, there is consensus on the reliability, validity, and utility of SJTs as selection methods in medicine. Further research on this selection method should focus on the finding that the predictive validity of SJTs increases throughout medical education and training and the predictive validity of medical school entrance SJTs on subsequent performance as a doctor.

Opportunities for Further Research

Box 26.7 summarises the predictive validity research evidence for different selection methods. The evidence on each of the techniques listed also includes an estimate of extent of usage across all occupational groups and applicant reactions. We have also summarised the extent to which each selection method might address widening access and diversity issues. Note that there are international differences in the extent of usage for various techniques, which is governed by international differences in employment law.

Conclusion

Research into medical selection is relatively new and there remain uncharted territories for exploration. Medicine continues to change and associated skills relevant to many specialties are changing. For example, in surgery, the use of laparoscopes and other technologies has transformed many surgical procedures. With international developments in technology and new treatment regimens, the pace of change is likely to change in the future. Since the career path for a physician is long and complex, it is difficult to define appropriate selection criteria for physicians of the future. Going forward, research must involve more future-focused job analysis studies to define the knowledge, skills, abilities, and attitudes relevant for physicians in general, and to explore any differences between specialties. In many countries, the enhanced focus on patient satisfaction has highlighted the need for empathy and communication skills, where physicians increasingly work in partnership with patients.

The selection gateways to progressing in medical training should be accompanied by accurate career information for individuals. Self-selection is crucially important and further research is warranted in this area. In the selection literature, very little research exists at more senior level appointments and so future research must address this, particularly at the consultant level, where competencies required may also include leadership of multi-professional teams, resource management, and political awareness [1].

More research is needed in the area of candidate reactions in medical education and training. What combination of selection tests is seen as fair and valid by candidates? What are the implications for widening access to medicine? What is the effect of candidate reactions (emotional, anxiety, perceived justice, and fairness) on test performance and test validity? Whilst there is a growing research literature on these topics in other professions [8], they are yet to be explored within medicine.

Future research should also explore how to best design a selection system across the whole training pathway. There may be generic skills required across all specialities (i.e. the basic skills for being a doctor, including cognitive, non-academic, and behavioural skills) which should guide the design of selection criteria for recruitment to undergraduate medical courses at the outset of training. It is unlikely that one part of the medical career path (e.g. undergraduate to initial specialty training) should fully equip an applicant with all the skills to progress from one stage to the next (e.g. from specialty training to senior appointments), especially if candidates are not selected to have the core aptitude at the outset.

Designing an accurate selection system is a complex process. Medical education and training is a long process, and the predictive validity of selection methods may not be consistent at different points in the career pathway. In other words, one factor may be an important predictor for undergraduate training but may not be predictive for aspects of specialty training. For example, openness to experience is important to general practitioner training performance, but is not important for undergraduate training performance. The real challenge is to integrate this knowledge and to weight and sequence different selection methods appropriately so as to develop selection systems that are valid from undergraduate selection through to specialty training.

Future research must account for established theoretical models of adult intellectual development and skill acquisition, which integrate cognitive and non-cognitive factors. One such model is PPIK theory [67], which asserts that adult intellectual ability is a function of *process* (basic mental capacities such as processing speed), *personality* (e.g. extraversion, conscientiousness), *interests* (e.g. preferences for science or art), and *knowledge* (e.g. factual knowledge as contained in A-levels). PPIK theory proposes a developmental trajectory to understand adult intellectual functioning, where personality, intellect, and interests operate alongside each other. For example, a person's interests are likely to influence the types of knowledge they seek out. This approach may help us to understand what motivates people to study medicine and their choice and aptitude for a specialty later in training. The idea of trait complexes (overlapping cognitive and non-cognitive traits) should be considered in medical selection [67]. Clusters of traits may be identified that overlap to define areas of competence and preference. In future, it may be possible to identify trait complexes that are unique to success in undergraduate medical education and in later specialty training.

There are many opportunities for research in medical selection for both undergraduate and postgraduate medical education and Patterson et al. [106] have proposed a future research agenda. To summarise, key topics for future research in selection are job analysis, longitudinal validity studies, organisational justice, exploring trait overlaps, and the temporal dynamics of training.

References

1 Patterson, F., Tavabie, A., Met, D. et al. (2013). A new competency model for general practice: implications for selection, training, and careers. *British Journal of General Practice* 63 (610): e331–e338.

2 Jefferis, T. (2007). Selection for specialist training: what can we learn from other countries? *BMJ* 334 (7607): 1302–1304.

3 Patterson, F. and Zibarras, L.D. (2011). Exploring the construct of perceived job discrimination in selection. *International Journal of Selection and Assessment* 19 (3): 251–257.

4 Patterson, F., Lievens, F., Kerrin, M. et al. (2012). Designing selection systems for medicine: the importance of balancing predictive and political validity in high-stakes selection contexts. *International Journal of Selection and Assessment* 20 (4): 486–496.

5 Ferguson, E., James, D., and Madeley, L. (2002). Factors associated with success in medical school: systematic review of the literature. *BMJ* 324: 952–957.

6 Bloch, R. and Norman, G. (2012). Generalizability theory for the perplexed: a practical introduction and guide: AMEE Guide No. 68. *Medical Teacher* 34 (11): 960–992.

7 Robertson, I.T. and Smith, M. (2001). Personnel selection. *Journal of Occupational and Organizational Psychology* 74 (4): 441–472.

8 Hausknecht, J.P., Day, D.V., and Thomas, S.C. (2004). Applicant reactions to selection procedures: an updated model and meta-analysis. *Personnel Psychology* 57 (3): 639–683.

9 Patterson, F., Zibarras, L.L., Carr, V. et al. (2011). Evaluating candidate reactions to selection practices using organisational justice theory. *Medical Education* 45 (3): 289–297.

10 Lievens, F., Patterson, F., Corstjens, J. et al. (2016). Widening access in selection using situational judgement tests: evidence from UKCAT. *Medical Education* 50 (6): 624–636.

11 Patterson, F., Prescott-Clements, L., Zibarras, L. et al. (2015). Recruiting for values in healthcare: a preliminary review of the evidence. *Advances in Health Sciences Education: Theory and Practice* 21 (4): 859–881.
12 Francis, R. (2013). Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry. http://www.midstaffspublicinquiry.com/sites/default/files/report/Executivesummary.pdf (accessed 21 June 2017).
13 Cavendish, C. The Cavendish Review: An Independent Review into Healthcare Assistants and Support Workers in the NHS and Social Care Settings. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/236212/Cavendish_Review.pdf (accessed 21 June 2017).
14 Patterson, F., Knight, A., Dowell, J. et al. (2016). How effective are selection methods in medical education and training? Evidence from a systematic review. *Medical Education* 50: 36–60.
15 Hartwell, C.J. (2014). The structured employment interview: narrative and quantitative review of the research literature. *Personnel Psychology* 67 (1): 1–82.
16 Campion, M.A., Pursell, E.D., and Brown, B.K. (1988). Structured interviewing: raising the psychometric properties of the employment interview. *Personnel Psychology* 41: 25–42.
17 Goho, J. and Blackman, A. (2006). The effectiveness of academic admission interviews: an exploratory meta-analysis. *Medical Teacher* 28: 335–340.
18 McDaniel, M.A., Whetzel, D.L., Schmidt, F.L., and Maurer, S. (1994). The validity of employment interviews: a comprehensive review and meta-analysis. *Journal of Applied Psychology* 79: 599–615.
19 Wiesner, W.H. and Cronshaw, S.F. (1988). A meta-analytic investigation of the impact of interview format and the degree of structure on the validity of the employment interview. *Journal of Occupational Psychology* 61: 275–290.
20 Cortina, J.M., Goldstein, N.B., Payne, S.C. et al. (2000). The incremental validity of interview scores over and above cognitive ability and conscientiousness scores. *Personnel Psychology* 53 (2): 325–351.
21 Huffcutt, A.I. and Roth, P.L. (1998). Racial group differences in employment interview evaluations. *Journal of Applied Psychology* 83 (2): 179–189.
22 Posthuma, R.A., Morgeson, F.P., and Campion, M.A. (2002). Beyond employment interview validity: a comprehensive narrative review of recent research and trends over time. *Personnel Psychology* 55 (1): 1–81.
23 Donnon, T., Oddone-Paolucci, E., and Violato, C. (2009). A predictive validity study of medical judgment vignettes to assess students' noncognitive attributes: a 3-year prospective longitudinal study. *Medical Teacher* 31 (4): e148–e155.
24 Patrick, L.E., Altmaier, E.M., Kuperman, S., and Ugolini, K. (2001). A structured interview for medical school admission, Phase 1: initial procedures and results. *Academic Medicine* 76 (1): 66–71.
25 Williamson, L.G., Campion, J.E., Malos, S.B. et al. (1997). Employment interview on trial: linking interview structure with litigation outcomes. *Journal of Applied Psychology* 82: 900–912.
26 Quintero, A.J., Segal, L.S., King, T.S., and Black, K.P. (2009). The personal interview: assessing the potential for personality similarity to bias the selection of orthopaedic residents. *Academic Medicine* 84 (10): 1364–1372.
27 Albanese, M.A.P., Snow, M.H.P., Skochelak, S.E.M.M. et al. (2013). Assessing personal qualities in medical school admissions. *Academic Medicine* 78 (3): 313–321.
28 Knorr, M. and Hissbach, J. (2014). Multiple mini-interviews: same concept, different approaches. *Medical Education* 48 (12): 1157–1175.
29 Dowell, J., Lynch, B., Till, H. et al. (2012). The multiple mini-interview in the UK context: 3 years of experience at Dundee. *Medical Teacher* 34 (4): 297–304.
30 Roberts, C., Walton, M., Rothnie, I. et al. (2008). Factors affecting the utility of the multiple mini-interview in selecting candidates for graduate-entry medical school. *Medical Education* 42 (4): 396–404.
31 Harris, S. and Owen, C. (2007). Discerning quality: using the multiple mini-interview in student selection for the Australian National University Medical School. *Medical Education* 41 (3): 234–241.
32 Hofmeister, M., Lockyer, J., and Crutcher, R. (2009). The multiple mini-interview for selection of international medical graduates into family medicine residency education. *Medical Education* 43 (6): 573–579.
33 Gafni, N., Moshinsky, A., Eisenberg, O. et al. (2012). Reliability estimates: behavioural stations and questionnaires in medical school admissions. *Medical Education* 46 (3): 277–288.
34 Kumar, K., Roberts, C., Rothnie, I. et al. (2009). Experiences of the multiple mini-interview: a qualitative analysis. *Medical Education* 43: 360–367.
35 Humphrey, S., Dowson, S., Wall, D. et al. (2008). Multiple mini-interviews: opinions of candidates and interviewers. *Medical Education* 42 (2): 207–213.
36 Ann Courneya, C., Wright, K., Frinton, V. et al. (2005). Medical student selection: choice of a semi-structured panel interview or an unstructured one-on-one interview. *Medical Teacher* 27 (6): 499–503.
37 Prideaux, D., Roberts, C., Eva, K. et al. (2011). Assessment for selection for the health care professions and specialty training: consensus statement and recommendations from the Ottawa 2010 Conference. *Medical Teacher* 33 (3): 215–223.
38 Parry, J., Mathers, J., Stevens, A. et al. (2006). Admissions processes for five year medical courses at English schools: review. *BMJ* 332 (7548): 1005–1009.
39 Eva, K.W., Reiter, H.I., Trinh, K. et al. (2009). Predictive validity of the multiple mini-interview for selecting medical trainees. *Medical Education* 43 (8): 767–775.
40 McCarthy, J.M. and Goffin, R.D. (2001). Improving the validity of letters of recommendation: an investigation of three standardized reference forms. *Military Psychology* 13 (4): 199–222.
41 Ferguson, E., James, D., O'Hehir, F. et al. (2003). Pilot study of the roles of personality, references, and personal statements in relation to performance over the five years of a medical degree. *BMJ* 326 (7386): 429–432.
42 Stedman, J.M., Hatch, J.P., and Schoenfeld, L.S. (2009). Letters of recommendation for the predoctoral internship in medical schools and other settings: do they enhance decision making in the selection process? *Journal of Clinical Psychology in Medical Settings* 16 (4): 339–345.
43 Ryan, A.M., Sacco, J.M., McFarland, L.A., and Kriska, S.D. (2000). Applicant self-selection: correlates of withdrawal from a multiple hurdle process. *Journal of Applied Psychology* 85 (2): 163–179.
44 Hughes, P. (2002). Can we improve on how we select medical students? *Journal of the Royal Society of Medicine* 95: 18–22.
45 Bates, B.P. (2002). Selection criteria for applicants in primary care osteopathic graduate medical education. *Journal of American Osteopathic Association* 102 (11): 621–626.
46 Peskun, C., Detsky, A., and Shandling, M. (2007). Effectiveness of medical school admissions criteria in predicting residency ranking four years later. *Medical Education* 41 (1): 57–64.
47 Benbassat, J. and Baumal, R. (2007). Uncertainties in the selection of applicants for medical school. *Advances in Health Sciences Education: Theory and Practice* 12 (4): 509–521.
48 Dore, K.L., Kreuger, S., Ladhani, M. et al. (2010). The reliability and acceptability of the multiple mini-interview as a selection instrument for postgraduate admissions. *Academic Medicine* 85 (10 Suppl): S60–S63.
49 Ferguson, E., Sanders, A., O'Hehir, F., and James, D. (2000). Predictive validity of personal statements and the role of the five-

factor model of personality in relation to medical training. *Journal of Occupational and Organizational Psychology* 73 (3): 321–344.

50 White, J.S., Lemay, J.-F., Brownell, K., and Lockyer, J. (2011). "A chance to show yourself" – how do applicants approach medical school admission essays? *Medical Teacher* 33 (10): e541–e548.

51 White, J., Brownell, K., Lemay, J.-F.F., and Lockyer, J.M. (2012). "What do they want me to say?" the hidden curriculum at work in the medical school selection process: a qualitative study. *BMC Medical Education* 12 (1): 17.

52 McManus, I.C., Powis, D., Wakeford, R. et al. (2005). Intellectual aptitude tests and A levels for selecting UK school leaver entrants for medical school. *BMJ* 331 (7516): 555–559.

53 Nicholson, S. (2005). The benefits of aptitude testing for selecting medical students. *BMJ* 331 (7516): 559–560.

54 Hayward, G., Sturdy, S., and James, S. (2005). Estimating the reliability of predicted grades: UCAS report. Cited in Hopkin, R. and Green, S. (2011) Commentry. Predicting A level grades using AS level grades. www.cambridgeassessment.org.uk/Images/109920-commentary-predicting-a-level-grades-using-as-level-grades.pdf (accessed 21 June 2017).

55 Arulampalam, W., Naylor, R., and Smith, J. (2004). Factors affecting the probability of first year medical student dropout in the UK: a logistic analysis for the intake cohorts of 1980–92. *Medical Education* 38 (5): 492–503.

56 McManus, I.C., Smithers, E., Partridge, P. et al. (2003). A levels and intelligence as predictors of medical careers in UK doctors: 20 year prospective study. *BMJ* 327 (7407): 139–142.

57 Lumb, A.B. and Vail, A. (2004). Comparison of academic, application form and social factors in predicting early performance on the medical course. *Medical Education* 38 (9): 1002–1005.

58 Reede, J.Y. (1999). Predictors of success in medicine. *Clinical Orthopaedics and Related Research* 362: 72–77.

59 McManus, C. (1997). From selection to qualification: how and why medical students change. In: *Choosing Tomorrow's Doctors* (ed. I. Allen, P. Brown and P. Hughes), 60–79. London: Policy Studies Institute.

60 Neill, L.O., Hartvigsen, J., Wallstedt, B. et al. (2011). Medical school dropout – testing at admission versus selection by highest grades as predictors. *Medical Education* 45 (11): 1111–1120.

61 Zibarras, L. and Woods, S.A. (2010). A survey of UK selection practices across different organization sizes and industry sectors. *Journal of Occupational and Organizational Psychology* 83: 499–511.

62 Ryan, A.M., Mcfarland, L., Baron, H., and Page, R. (1999). An international look at selection practices: nation and culture as explanations for variability in practice. *Personnel Psychology* 52 (2): 359–392.

63 Salgado, J.F., Anderson, N., Moscoso, S. et al. (2003). A meta-analytic study of general mental ability validity for different occupations in the European community. *Journal of Applied Psychology* 88 (6): 1068–1081.

64 Bertua, C., Anderson, N., and Salgado, J.F. (2005). The predictive validity of cognitive ability tests: a UK meta-analysis. *Journal of Occupational and Organizational Psychology* 78 (3): 387–409.

65 Murphy, K.R. (2002). Can conflicting perspectives on the role of g in personnel selection be resolved? *Human Performance* 15 (1–2): 823–854.

66 Kehoe, J. (2002). General mental ability and selection in private sector organizations: a commentary. *Human Performance* 15 (1): 97–106.

67 Ackerman, P.L. (1996). A theory of adult intellectual development: process, personality, interests, and knowledge. *Intelligence* 22 (2): 227–257.

68 Blair, C. (2006). How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. *Behavioral and Brain Sciences* 29: 109–125.

69 Wright, S.R. and Bradley, P.M. (2010). Has the UK clinical aptitude test improved medical student selection? *Medical Education* 44 (11): 1069–1076.

70 Julian, E.R. (2005). Validity of the Medical College Admission Test for predicting medical school performance. *Academic Medicine* 80 (10): 910–917.

71 Donnon, T., Paolucci, E.O., and Violato, C. (2007). The predictive validity of the MCAT for medical school performance and medical board licensing examinations: a meta-analysis of the published research. *Academic Medicine* 82 (1): 100–106.

72 Emery, J.L., Bell, J.F., and Vidal Rodeiro, C.L. (2011). The BioMedical Admissions Test for medical student selection: issues of fairness and bias. *Medical Teacher* 33 (1): 62–71.

73 McManus, I.C., Ferguson, E., Wakeford, R. et al. (2011). Predictive validity of the Biomedical Admissions Test: an evaluation and case study. *Medical Teacher* 33 (1): 53–57.

74 Barrick, M.R. and Mount, M.K. (2012). Nature and use of personality in selection. In: *The Oxford Handbook of Personnel Assessment and Selection* (ed. J.N. Butcher), 225–251. Oxford: Oxford University Press.

75 Knights, J.A. and Kennedy, B.J. (2007). Medical school selection: impact of dysfunctional tendencies on academic performance. *Medical Education* 41 (4): 362–368.

76 Lumsden, M.A., Bore, M., Millar, K. et al. (2005). Assessment of personal qualities in relation to admission to medical school. *Medical Education* 39 (3): 258–265.

77 Haight, S.J., Chibnall, J.T., Schindler, D.L., and Slavin, S.J. (2012). Associations of medical student personality and health / wellness characteristics with their medical school performance across the curriculum. *Academic Medicine* 87 (4): 476–485.

78 Lievens, F., Coetsier, P., De Fruyt, F. and De Maeseneer J. (2002). Medical students' personality characteristics and academic performance: a five-factor model perspective. *Medical Education* 36 (11): 1050–6.

79 Ferguson, E., Semper, H., Yates, J. et al. (2014). The "dark side" and "bright side" of personality: when too much conscientiousness and too little anxiety are detrimental with respect to the acquisition of medical knowledge and skill. *PLoS One* 9 (2): e88606.

80 Ferguson, E. and Lievens, F. (2017). Future directions in personality, occupational and medical selection: myths, misunderstandings, measurement, and suggestions. *Advances in Health Sciences Education: Theory and Practice* 22 (2): 387–399.

81 Sefcik, D.J., Prerost, F.J., and Arbet, S.E. (2009). Personality types and performance on aptitude and achievement tests: implications for osteopathic medical education. *Journal of the American Osteopathic Association* 109 (6): 296–301.

82 Dowell, J., Lumsden, M., Powis, D. et al. (2011). Predictive validity of the personal qualities assessment for selection of medical students in Scotland. *Medical Teacher* 33 (9): e485–e488.

83 Lievens, F., Ones, D.D.S., and Dilchert, S. (2009). Personality scale validities increase throughout medical school. *Journal of Applied Psychology* 94 (6): 1514–1535.

84 Nettle, D. (2006). The evolution of personality variation in humans and other animals. *American Psychologist* 61 (6): 622–631.

85 Tett, R.P., Jackson, D.N., Rothstein, M., and Reddon, J.R. (1999). Meta-analysis of bidirectional relations in personality-job performance research. *Human Performance* 12 (1): 1–29.

86 Murphy, K.R. (2005). Why don't measures of broad dimensions of personality perform better as predictors of job performance? *Human Performance* 18 (4): 343–357.

87 Birkeland, S.A., Manson, T.M., Kisamore, J.L. et al. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment* 14 (4): 317–335.

88 Rosse, J.G., Miller, J.L., and Stecher, M.D. (1994). A field study of job applicants' reactions to personality and cognitive ability testing. *Journal of Applied Psychology* 79 (6): 987–991.

89 Li, A. and Bagger, J. (2006). Using the BIDR to distinguish the effects of impression management and self-deception on the criterion validity of personality measures: a meta-analysis. *International Journal of Selection and Assessment* 14 (2): 131–141.

90 Patterson, F., Ferguson, E., Norfolk, T., and Lane, P. (2005). A new selection system to recruit general practice registrars: preliminary findings from a validation study. *BMJ* 330 (7493): 711–714.

91 Randall, R., Davies, H., Patterson, F., and Farrell, K. (2006). Selecting doctors for post-graduate training in paediatrics using a competency-based assessment centre. *Archives of Disease in Childhood* 91: 444–448.

92 Randall, R., Stewart, P., Farrell, K., and Patterson, F. (2006). Using an assessment centre to select doctors for postgraduate training in obstetrics and gynaecology. *Obstetrics & Gynecology* 8 (4): 257–262.

93 Kidd, J., Fuller, J., Patterson, F., and Carter, Y. (2006). Selection Centres: Initial description of a collaborative pilot project. *Proceedings for the Association for Medical Education in Europe (AMEE) Conference*, Genoa, Italy.

94 Ziv, A., Rubin, O., Moshinsky, A. et al. (2008). MOR: a simulation-based assessment centre for evaluating the personal and interpersonal qualities of medical school candidates. *Medical Education* 42 (10): 991–998.

95 ten Cate, O. and Smal, K. (2002). Educational assessment centre techniques for entrance selection in medical school. *Academic Medicine* 77 (7): 737.

96 Gale, T.C.E., Roberts, M.J., Sice, P.J. et al. (2010). Predictive validity of a selection centre testing non-technical skills for recruitment to training in anaesthesia. *British Journal of Anaesthesia* 105 (5): 603–609.

97 Schmidt, F. and Hunter, J. (1998). The validity and utility of selection methods in personnel psychology: practical and theoretical implications of 85 years of research findings. *Psychology Bulletin* 124: 262–274.

98 Anderson, N.R., Lievens, F., Van Dam, K. et al. (2006). A construct-driven investigation of gender differences in a leadership-role assessment center. *Journal of Applied Psychology* 91 (3): 555–66.

99 Lievens, F., Harris, M.M., Van Keer, E. and Bisqueret, C. (2003). Predicting cross-cultural training performance: the validity of personality, cognitive ability, and dimensions measured by an assessment center and a behavior description interview. *Journal of Applied Psychology* 88(3): 476–89.

100 Krause, D.E., Kersting, M., Heggestad, E.D., and Thornton, G.C. (2006). Incremental validity of assessment center ratings over cognitive ability tests: a study at the executive management level. *International Journal of Selection and Assessment* 14 (4): 360–371.

101 Rupp, D.E., Hoffman, B.J., Bischof, D. et al. (2015). Guidelines and ethical considerations for assessment center operations. *Journal of Management* 41 (4): 1244–1273.

102 Patterson, F., Zibarras, L., and Ashworth, V. (2016). Situational Judgement tests in medical education and training: research, theory and practice: AMEE Guide No. 100. *Medical Teacher* 38 (1): 3–17.

103 Patterson, F., Baron, H., Carr, V. et al. (2009). Evaluation of three short-listing methodologies for selection into postgraduate training in general practice. *Medical Education* 43 (1): 50–57.

104 Lievens, F., Peeters, H., and Schollaert, E. (2008). Situational judgment tests: a review of recent research. *Personnel Review* 37 (4): 426–441.

105 Christian, M.S., Edwards, B.D., and Bradley, J.C. (2010). Situational judgment tests: constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology* 63 (1): 83–117.

106 Lievens, F. and Patterson, F. (2011). The validity and incremental validity of knowledge tests, low-fidelity simulations, and high-fidelity simulations for predicting job performance in advanced-level high-stakes selection. *Journal of Applied Psychology* 96 (5): 927–940.

107 Zibarras, L.D. and Patterson, F. (2015). The role of job relatedness and self-efficacy in applicant perceptions of fairness in a high-stakes selection setting. *International Journal of Selection and Assessment* 23 (4): 332–344.

108 Cousans, F., Patterson, F., Edwards, H. et al. (2017). Evaluating the complementary roles of an SJT and academic assessment for entry into clinical practice. *Advances in Health Sciences Education* 22 (2): 401–413.

109 Patterson, F., Cleland, J., and Cousans, F. (2017). Selection methods in healthcare professions: where are we now and where next? *Advances in Health Sciences Education: Theory and Practice* 22 (2): 229–242.

110 McDaniel, M.A., Morgeson, F.P., Finnegan, E.B. et al. (2001). Predicting job performance using situational judgment tests: a clarification of the literature. *Journal of Applied Psychology* 86 (4): 730–740.

111 McDaniel, M., Hartman, N., Whetzel, D., and Grubb, W. (2007). Situational judgment tests, response instructions, and validity: a meta-analysis. *Personnel Psychology* 60 (1): 63–91.

112 Patterson, F., Ashworth, V., Zibarras, L. et al. (2012). Evaluations of situational judgement tests to assess non-academic attributes in selection. *Medical Education* 46 (9): 850–868.

113 Ahmed, H., Rhydderch, M., and Matthews, P. (2012). Can knowledge tests and situational judgement tests predict selection centre performance? *Medical Education* 46: 777–784.

114 Clevenger, J., Pereira, G.M., Wiechmann, D. et al. (2001). Incremental validation of situational judgment tests. *Journal of Applied Psychology* 86 (3): 410–417.

115 Koczwara, A., Patterson, F., Zibarras, L.L. et al. (2012). Evaluating cognitive ability, knowledge tests and situational judgement tests for postgraduate selection. *Medical Education* 46 (4): 399–408.

Further Reading

Patterson, F., Knight, A., Dowell, J. et al. (2016). How effective are selection methods in medical education and training? Evidence from a systematic review. *Medical Education* 50: 36–60.

Prideaux, D., Roberts, C., Eva, K. et al. (2011). Assessment for selection for the health care professions and specialty training: consensus statement and recommendations from the Ottawa 2010 Conference. *Medical Teacher* 33 (3): 215–223.

Patterson, F., Prescott-Clements, L., Zibarras, L. et al. (2015). Recruiting for values in healthcare: a preliminary review of the evidence. *Advances in Health Sciences Education: Theory and Practice* 21 (4): 859–881.

Part 4

Research and Evaluation

27 Philosophical Research Perspectives and Planning your Research

Jan Illing[1] *and Madeline Carter*[2]

[1]School of Medical Education, Newcastle University, Newcastle, UK
[2]Department of Psychology, Faculty of Health and Life Sciences, Northumbria University, Newcastle, UK

KEY MESSAGES

- Research is the practice of critical or scientific inquiry.
- Research activities overlap with audit, quality improvement, and evaluation; but there are some important distinctions related to purpose and scope.
- Philosophical perspectives provide the assumptions and frameworks that guide research.
- A philosophical perspective encompasses ontology, epistemology, and methodology.
- Research can use multiple methods and can combine qualitative and quantitative data.
- Research is expected to be ethical, to minimise the risk of harm or discomfort to people, which in educational or social research is more likely to take the form of psychological distress than physical injury.

Introduction

This chapter aims to provide a map to explain the fundamental philosophical concepts that underpin knowledge creation during research and to provide information on the practical issues a researcher should consider before starting out on their project.

What is Research?

Research has been defined as 'a search or investigation directed to the discovery of some fact by careful consideration or study of a subject; a course of critical or scientific inquiry' [1]. This definition may sound straightforward, in that most researchers would agree that they are involved in a critical inquiry of something, but some would argue that their aim is not to establish facts but to increase or change understanding about something.

How does Research Differ from Audit, Quality Improvement, and Evaluation?

There is considerable overlap in research, audit, quality improvement (QI), and evaluation activities, and the degree of overlap may be dependent on the specific project. However, there are some important distinctions, often focused on project scope and purpose [2]. The designation may also influence the need for ethical review, as research typically requires ethical approval whereas other activities may not.

'Research is concerned with discovering the right thing to do; audit with ensuring that it is done right' [3, p. 905]. Audit may be ongoing and uses routine data to improve local practice, whereas research is often a one-off activity in which more complex data is collected with a view to generalisability or transferability. Although both may use similar methods, audit is often regarded as less rigorous [3]. Audit is typically linked to a quality improvement cycle that uses data for benchmarking purposes, takes specific actions to improve, and monitors performance against agreed standards [2].

Quality improvement (QI) initiatives have been defined as 'small-scale cycles of interventions that are linked to assessment and that have the goal of improving the process, outcome, and efficiency of complex systems in health care' [4, p. 2276]. QI projects often share a number of features with research and, in practice, some activities may be regarded as both QI and research. Both frequently adopt a systematic approach to investigation, use working hypotheses on how to improve processes, and utilise a similar range of research methods and analytic tools; prompting debate on the ethical requirements of QI [5]. However, for some, research may be more associated with developing independent and generalisable or transferable knowledge in discrete projects, whereas QI may be more associated with implementation and a continuous process of modification, assessment, and change with the aim of rapid improvement of care and local practice [5].

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

According to Clarke [6], what differentiates evaluation from research is the question of purpose. 'An evaluation is action orientated. It is conducted to determine the value or impact of a policy, programme, practice, intervention or service, with a view to making recommendations for change.' Robson [7] states that 'to evaluate is to assess the worth or value of something'. Following this definition, evaluation is about setting out to make a judgement. Going back to our definition of research, there is no mention of research leading to judgement, but to the identification of findings by critical inquiry. Evaluation research is part of research, but in evaluation the aim involves assessing the worth of something (see Chapter 30). Commentaries on similarities and differences between research and evaluation are summarised by Cohen [8]. Similarities focus on methods and the skills of the evaluators involved. Differences focus on motivation, purpose, and objectives, with researchers having a greater emphasis on advancing knowledge, satisfying curiosity, and contributing to theory rather than solving a problem or informing a decision. Politics may have a stronger role in evaluation, as evaluations are typically commissioned by a client and the evaluator has less autonomy. Research is judged by its contribution to the field and internal and external validity, whereas evaluations are judged by the criteria of utility and credibility. However, Cohen [8] acknowledged the blurring of these distinctions, particularly due to the politicised nature of research agendas, funding, and the growth of applied research.

What is Scholarship?

Crites et al. [9] described scholarly activities as 'inquiries guided by an academic tradition and the dissemination of the results of inquiries to allow peer judgment of their merit, erudition and utility; the cumulative description for all these activities is scholarship' (p. 658).

Traditionally, perspectives on scholarship focused on research and publication. However, in his seminal work, Boyer [10] sought to broaden definitions of scholarship and proposed four categories to capture different elements of scholarship. These included:

- *The scholarship of discovery*: original research and the discovery of new knowledge in order to better understand the world.
- *The scholarship of integration*: the placement of isolated research into a wider context and making connections within or between disciplines, or between findings obtained by different approaches.
- *The scholarship of application*: how theory and practice interact to inform each other, and how knowledge can be used and applied.
- *The scholarship of teaching*: communicating knowledge effectively to others; stimulating others to become active learners and encouraging students to be critical, creative thinkers, with the capacity to go on learning after their college days are over.

The quality of all types of scholarship can be assessed using six standards proposed by Glassick [11]. High-quality scholarship should involve:

- clear goals (clear and important purpose, achievable objectives)
- adequate preparation (understanding of the field, necessary skills, and resources)
- appropriate methods (suitable selection, application, and modification of methods)
- significant results (work adds to the field, goals achieved)
- effective presentation (appropriate communication of work)
- reflective critique (critical evaluation).

This chapter will focus primarily on the scholarship of discovery, but it is important to acknowledge the broader definitions and the calls for more equitable recognition of all types of scholarship in the literature [12–14].

Theoretical Frameworks in Education and the Social Sciences

Kneebone [15] published a personal view about his attempt to engage with the education and social science literature. He wrote: 'At first and to my great surprise I found this literature almost impenetrable, of course it was peppered with unfamiliar words … I had the disquieting sensation of moving into alien territory, where familiar landmarks had disappeared.' Kneebone came to the realisation that all of his medical training had been based within one view of science, the positivist paradigm, and that this was a very narrow and limited view. He ended with a plea to include an exploration of what the humanities have to offer the medical curriculum, and also with explicit guidance on how to gain access to this world. The aim of this chapter is to make this other 'world' penetrable.

The focus of this particular section is to present some of the frameworks within which research in education and the social sciences are conducted. Quantitative research in education and social science is typically represented by the social survey and experimental methods, whereas qualitative research uses techniques such as observation and interview. Alignment between the method selected and the research question posed is crucial, and each type of approach signals to the reader the framework within which the research is expected to be read and judged. Research methods are addressed in detail in Chapters 28 and 29 of this book.

In the past, the scientific method applied to the study of the natural sciences was considered appropriate and desirable for the study of education and the social sciences. Early textbooks focused on the scientific method, and other methods such as participant observation were deemed less scientific and weak by comparison, and consequently of lower status. From the 1970s, the debate over the appropriateness of the natural science model for social sciences inquiry gained momentum. Arguments centred on the differences in focus; people in education and the social sciences, and objects in the natural sciences. The terms 'qualitative' and 'quantitative' signified more than different methods of collecting data; they indicated different assumptions about research in the social world.

The debate may have gathered momentum following Kuhn's [16] work on the history of science. Of particular importance is Kuhn's idea of a *paradigm*, a set of beliefs and dictates that influence what should be studied, how the research should be conducted, and how the results should be interpreted. It can be compared to viewing the social world through a particular lens and encompasses ontology, epistemology, theory, and methods. Paradigms cannot be proven but rely on argument, persuasion, and usefulness. A paradigm is defined as 'a conceptual or methodological model underlying the theories and practices of a science or discipline at a particular time; [hence] a generally accepted world view' [1].

With the exception of positivism, all the other perspectives discussed below are still in their formative stages of development and will be referred to as philosophical perspectives rather than paradigms.

BOX 27.1 Key terms

- *Ontology:* the study of being. It is concerned with the nature of existence and the structure of reality. With regard to social inquiry, this is often taken to mean the assumptions that a particular theoretical perspective makes about the nature of social reality.
- *Epistemology:* the theory of knowledge, its origins and nature, and the limits of knowledge.
- *Methodology:* the research design or plan that shapes the methods to be used in the study. The methodology provides a rationale for the choice of methods used in a study.
- *Methods:* the techniques used for data collection.
- *Philosophical perspective:* the philosophical framework and assumptions that lie behind the methodology.

Philosophical Perspectives in Research

Philosophical perspectives are taken here to mean the philosophical stances that underpin research methodology. Philosophical perspectives are the starting point from which assumptions about the research are based; they influence how the study is conducted, the researcher's role, and the type of knowledge that is produced. Each perspective will also have a particular set of criteria to be used in evaluating a piece of research. There has been a great deal written about the different perspectives, and much of it has focused narrowly on only one perspective without guiding the reader to where each perspective sits in relation to others. What is offered here is an overview of the conventional positivist and post-positivist perspectives, as well as more recent perspectives. For more detailed exposition, see these selected references [17–20].

Each perspective takes a particular *ontological* and *epistemological* position that informs the resulting research methods. *Ontology* is the study of being, and is concerned with the nature of existence and the structure of reality. It raises questions about the nature and form of reality and what can be known about it. In the social world is there a 'real' and single reality? Are there multiple realities dependent on whose view is taken? *Epistemology* focuses on the nature of the relationship between the researcher and what is to be known. The epistemological question is dependent on the answer to the ontological question. For example, when reality is assumed to be 'real', then what can be known about it can be independent of any relationship between the researcher and the subject of inquiry, and knowledge can be said to be objective. Therefore, the concept of objectivity in research assumes the existence of a 'real' world. However, if the answer to the ontological question is that reality is socially constructed and there is no single 'real' version, then the answer to the epistemological question becomes subjective, as each researcher (and research participant) has his or her own version of reality and there is no single true version, only a socially constructed reality. The methodological approach taken comes secondary to the answers to the ontological and epistemological questions (and focuses on the methods by which knowledge can be acquired on the subject of inquiry). If a 'real' reality is assumed, then this implies that the researcher can collect objective data and the ability to control variables becomes feasible (see Box 27.1).

Positivism

Positivism has been the dominant perspective in the physical and social sciences, going back to the Enlightenment in the seventeenth century, and is identified with the study of the natural world and with quantitative methods. Positivism is linked to empirical science, offering assurances that knowledge is unambiguous, accurate, and certain. 'Positive' comes from '*something that is posited*', a science that is firmly grounded, not something that is arrived at from speculation. Auguste Comte (1798–1857) is attributed as the founder of positivism, although the ideas on establishing scientific laws from observation and experiment are reported much earlier in the work of Francis Bacon (1561–1626). What is posited in positivist science is what is scientifically observed following use of the scientific method. Comte's positivism bids us to look for regular characteristics and constant relationships to facts and to laws that can be scientifically established using the scientific method of observation, experimentation, and comparison. The 'verification principle', which became a central tenet of positivism, is attributed to Ludwig Wittgenstein (1889–1951). The verification principle focuses on the importance of verifying statements via the use of the scientific method and the resulting outcomes. Today, positivism is still linked to empirical science. The confidence in science is reflected in the belief that science is both accurate and certain, in contrast to values, opinions, and feelings, which are empirically unverifiable and of no interest to positivism.

Ontology, Epistemology, and Methodology

The ontology of positivism is realism. Reality is assumed to exist in an 'absolute' sense, and the aim is to explain the social world in terms of laws, often including cause and effect. The epistemology of positivism is objectivism.

Positivism maintains that objects in the world have meaning both prior to and independently of any consciousness of them. Positivism maintains that there are 'facts' that can be accurately collected about the social world, which are independent of individual interpretation and are 'true'. Researchers can be objective in the collection and interpretation of data. The researcher takes on the role of independent observer, and seeks to remove their own bias and to standardise methods. It is assumed that the researcher is capable of investigating the object of study without influencing it or being influenced by it. This differs from our subjective understanding, which constitutes a different form of knowledge from knowledge made up of scientific facts.

Positivist methodology is usually deductive and the aim is often concerned with the prediction and control of phenomena, and involves testing hypotheses to support or disprove a theory. Research procedures need to be followed rigorously to prevent values and biases from affecting the data. Methods are reported in detail to enable others to repeat the study and show that the results are replicable. The methods used typically involve experimental or manipulative research designs that produce quantitative data. The aim is to generalise findings to a larger population than the study sample.

Quality is assessed by internal validity (findings are congruent with sound research methods which have minimised the effects of confounding variables), external validity (findings are generalisable to other settings, other people, and over time), reliability (findings are stable), and objectivity (the researcher has not influenced findings).

Knowledge, Values, and Ethics

Knowledge from positivist research is built up incrementally, like building blocks, by adding new knowledge to old and determining where it fits with existing knowledge. This research frequently aims to form rules and laws such as cause and effect. While the reporting of scientific knowledge is acceptable, criticisms focus on claiming that scientific knowledge is the only valid form of knowledge and that it is completely objective and accurate.

Ethics and values are important for all types of research, although treated differently by them [17]. Values are excluded in positivism, which claims to be value-free as a result of its epistemological position that research can be objective if rigour is applied. Values are viewed as confounding variables that need to be controlled and excluded, as are subjectivity and bias (e.g. using standardised instructions and double-blind experiments). Research ethics, although of importance in positivism, is largely viewed as something external to the research itself. Ethics is seen as something that would be applied to the research, possibly by an external research ethics body or a professional body that may advise on the professional conduct of researchers.

Is there Conflict with other Perspectives?

Proponents of positivism take a *reductionist* stance, in that it is assumed that at some point in the future a structure will be identified on which questions of difference can be considered and explained. There is much disagreement about this from proponents of critical theory and constructivism. Positivists would see action research as a contamination of both the research process and research findings.

Post-positivism

Post-positivism emerged following a realisation that the scientific method could not be applied to all scientific theory and much of what was accepted as 'fact' was theory and had not been observed or the act of observation changed the subject. Popper (1902–1994) introduced the principle of *falsification*, where the emphasis moved from proving a theory correct to being unable, through repeated testing, to prove it was wrong. Popper maintained that no theory could ever be proven, only disproved, and if a theory or hypothesis was not open to refutation from experimentation or observation, then the claims or theories made were not truly scientific.

Kuhn (1922–1996) questioned the objectivity and value neutrality of the scientific method and highlighted findings that could not be explained within the positivist paradigm. This led him to question the adequacy of the paradigm and called for a 'paradigm shift' and a shift in the way scientists viewed reality. The post-positivist perspective is less absolute; probability has replaced certainty; a level of objectivity has replaced absolute objectivity; and approximate truth has replaced absolute truth.

Ontology, Epistemology, and Methodology

The ontology of post-positivism is critical realism. Like positivism, reality is assumed to exist, but unlike positivism, reality cannot be truly 'known'. Access to reality is imperfect due to weaknesses in the human researcher and the complexity of the inquiry. Post-positivist epistemology is objectivist; objectivity is the ideal, but the data are subject to critical review. The post-positivist perspective acknowledges that no matter how much rigour is applied to the scientific method, research outcomes are never totally objective or certain, and claims are tempered. Emphasis is placed on collecting more than one type of data (triangulation) and on the falsification of hypotheses rather than confirmation. Research typically aims to provide an explanation and, when possible, predict and control phenomena. Again the researcher takes on the role of independent enquirer, who is impartial to the study findings and reports them objectively. Post-positivism aims to address some of the problems of positivist research by collecting data in natural settings and collecting the insider views. Like positivism, quality is assessed by internal validity, external validity, reliability, and objectivity.

Knowledge, Values, and Ethics

Knowledge consists of hypotheses that thus far have not been falsified and is made up of facts and laws that are probably 'true'. As with positivism, knowledge is built by adding new knowledge to old, in order to fit into existing patterns and form generalisations or rules such as cause and effect. As in positivism, post-positivist values are excluded. Values are perceived as confounding variables that need to be controlled. Research ethics is again viewed as something largely external to the research itself.

Is there Conflict with other Perspectives?

Proponents of this perspective take the same reductionist stance as positivism. It is assumed that at some point in the future a structure will be identified upon which questions of difference can be considered and explained. There is much disagreement about this from proponents of critical theory and constructivism.

Critical Theory and Related Ideological Positions

In contrast to positivist or post-positivist perspectives oriented to understanding or explaining the world, critical theory is oriented towards critiquing and changing society. Critical theory is used here as a blanket term, which includes, among others, the feminist and Marxist perspectives that are used here as illustrative examples.

Feminist research starts with criticism of science, stating that it is incomplete and reflects a male distortion of the social world. Although there are multiple forms of feminism, there is agreement that society has marginalised women and that this is reflected in research practice. Science perpetuates the myth of the superiority of men to women. Gender, as a significant issue in dealing with explanations of social phenomena, has largely been absent. The feminist perspective maintains that perpetuating a male view of science narrows ideas and limits understanding of the social world, and that if the male viewpoint were not dominant, a different research model would be dominant. Positivist research has stressed the importance of emotional separateness of researchers from their research participants to maintain objectivity. The feminist perspective maintains that research is a two-way process: detachment and objectivity are impossible, the researcher is affected by the research, and the researcher's own biography becomes a fundamental part of the research process and informs the analysis [21].

The Marxist perspective, like the feminist, seeks to challenge the status quo, to recognise conflict and oppression, and to bring about change. Marx perceived a basic conflict between capital and labour, between the bourgeoisie and the proletariat, and believed similar class struggles were part of earlier society. Marx maintained that economic forces determine how we think. Thoughts and consciousness come from our social being, itself the result of economic forces. Marx maintained that those who held economic power also held the intellectual power. The ruling classes ruled as thinkers, producers of ideas, and regulated the production and distribution of ideas [22].

Ontology, Epistemology, and Methodology

The ontology of critical theory is historical realism. Reality is assumed to be capturable but is shaped over time by social, cultural, gender, ethnic, political, and economic factors such that reality 'has set' over time. The epistemology is transactional and subjectivist: the researcher and the object of the research are assumed to be linked by the values of the researcher and relevant others who influence the study. Findings or knowledge are value dependent; they are mediated by the values of the researcher and the relevant others. It is the epistemological position that sets it apart from positivism and post-positivism. Methods require a dialogue between investigator and the subjects of inquiry. In critical theory, the researcher takes on the role of facilitator, raising not only their own level of consciousness about the object of study but also that of others. The researcher may facilitate change in the study group by providing greater insight into their situation and provide a stimulus for members of the community to take control of their future and initiate action and change. The aim of the research is to critique and change factors that constrain and exploit individuals. Quality is assessed by the historical context of the study; that is, whether it takes account of the social factors of the studied situation, and the extent to which the study acts to remove a lack of knowledge, and acts as a stimulus for action in the sense of bringing about a change in the existing structure.

Knowledge, Values, and Ethics

Knowledge is made up of historical or structural insights that will transform with time. Transformations occur following informed insight. Knowledge grows and changes with historical revision as ignorance is eroded. Values play a central role in critical theory and are important in shaping research outcomes. Excluding values would go against the interests of any minority or powerless group who were part of the study. The aim is to give the weak and powerless groups a platform to let their voices be heard along with any others who may be more dominant. Unlike the positivist and post-positivist perspectives, ethics is more internal than external to the research study. The critical theorist takes more of a moral standpoint in revealing full details about the study to ensure the study participant can be fully informed prior to consent and with no deception.

Is there Conflict with other Perspectives?

Critical theory and constructivism (see below) agree that they are in conflict with positivist and post-positivist perspectives. The epistemological position of critical theory sets it apart from the positivist and post-positivist perspective; research can be value-free or it cannot; and a single model cannot support both tenets.

Constructivism

Guba and Lincoln's [18] constructivism is a broad eclectic framework that embraces interpretive, phenomenological, and hermeneutic perspectives (for more detail see other publications [18, 20, 23]). Constructivism is the view that knowledge, and therefore all meaning, is not discovered (as in positivism) but socially constructed. Crotty [20] states that constructivism mirrors intentionality (meaning we intentionally create understanding) in that consciousness is directed towards an object that is shaped by our consciousness and what comes to the fore is the interaction between subject and object. From this, meaning is born.

It is accepted, even by the positivists, that social realities are socially constructed. The difference between constructivists and positivists is that the former maintain that all meaningful reality is socially constructed. A table may have a real existence irrespective of whether anyone is consciously aware of it. However, it exists as a table only if it is recognised as a table by our consciousness. The table is also constructed through social life, and our culture informs how we see these objects and in some cases whether we see them at all. Throughout our lives we learn about the social and natural worlds and interpret them, not as separate worlds but as one human world.

Ontology, Epistemology, and Methodology

The ontology of constructivism is relativism; this assumes multiple and sometimes conflicting realities that are socially and experientially based and dependent on individuals for their form and content. There is no 'real' world that pre-exists and is independent of human consciousness. People could therefore inhabit very different worlds based on different sets of meaning. The ontological position of constructivism is crucial in terms of separating it from other perspectives. The answer to the epistemological question of 'How do I know what I know?' is that reality is subjective. The researcher and the research object are assumed to be related, such that the research findings or knowledge are created from the relationship between the researcher and the subject of study. It is the epistemological position of constructivism that sets it apart from positivism and post-positivism. Guba and Lincoln [24] maintain that the inquiry methodology is a two-way process of listening to the constructions of both the researcher and the research participant, the researcher compares and contrasts different constructions to achieve a consensus. For Guba and Lincoln, the researcher cannot and should not be separated from the research participant, and hence the research outcomes are a joint construction of the research process. The aim of the research is understanding. Two sets of criteria are used to assess quality: *trustworthiness* (parallels internal validity), *transferability* (parallels external validity), *dependability* (parallels reliability), and *conformability* (parallels objectivity) make up the first set. These criteria are analogous to those used to judge quality in positivist research. The second set consists of authenticity criteria of fairness: *ontological authenticity* (develops and enhances personal constructions), *educative authenticity* (leads to improved understanding of others), *catalytic authenticity* (provides the stimulus to action), and *tactical authenticity* (the research empowers action) (see [24]). The second set of criteria share some common ground with critical theory.

Knowledge, Values, and Ethics

Knowledge consists of constructions about which there is relative consensus. Multiple constructions can coexist and be of equal weight, depending on interpretation and factors that influence interpretation such as social, political, and gender issues. For constructivism, values play a central role in creating and shaping the research outcomes. Constructivism views the role of researcher as the producer and facilitator of the research and acknowledges their central role in the research process. The researcher uses different approaches to the analysis of data by synthesising and coding data into themes and identifying meaning in them. Increasingly, constructivists aim to involve research participants in the study, by suggesting questions and outlets for research findings. The role of ethics, like values, is central to constructivism. The researcher's role is to recognise his/her own constructs and values and, as in critical theory, inform the study participants fully about the research prior to requesting consent, work towards uncovering the constructs of the study participants, and work towards understanding constructs. The methodology involves close personal interactions and, as a result, may raise some difficulties with confidentiality and anonymity [25].

Is there Conflict with other Perspectives?

According to Guba and Lincoln [24], the ontological stances of constructivism and critical theory are in conflict with the positivist and post-positivist perspectives. Either there is a 'real' reality or there is not; it is either value-free or it is not. The concept of reconciling both of these positions in one system seems impossible. However, mixed method approaches attempt to do just that. They ask different research questions and generate different types of knowledge, which can help to provide a more complete answer to the research aim by offering different types of knowledge.

Participatory Action Research

Participatory action research is a form of action research that involves research participants as both subjects and co-researchers. It is based on the proposition put forward by Kurt Lewin (1890–1947) that causal inferences about human behaviour are more likely to be valid if the relevant humans participate in building and testing them. Participatory action research arose partly out of recognition that a gap often exists between the completion and publication of high-quality research and the implementation of findings. To address this, researchers and participants engage in a collaborative cycle of planning, acting, observing, reflecting, feedback, and re-planning [26, 27]. Participatory action research involves research participants working alongside researchers throughout the research process, from the first steps of designing the study through to research outcomes [25]. The participatory perspective underpins forms of action research and is often seen in studies involving patient and public participation.

Ontology, Epistemology, and Methodology

The ontology of participatory action research is subjective–objective. Participatory action researchers recognise the multiple realities experienced by individuals, but state that the natural world has an objective reality. Perception represents an interplay between the 'real' objective world and our subjective experience of it. Epistemology is also closer to constructivism with critical subjectivity, where knowledge is seen as an interplay between researcher and participant.

The methodology is a collaborative form of action research and is explained in terms of knowing: people collaborate to define both the questions they wish to explore and the methodology for that exploration (propositional knowing); together or separately they apply this methodology in the world of their practice (practical knowing); which leads to new forms of encounter with their world (experiential knowing); and they find ways to represent this experience in significant patterns (presentational knowing) which feeds into a revised propositional understanding of the originating questions [19].

Heron and Reason [19] argue that cooperative inquiry has two participatory principles: first, that the research outcome is grounded in the researcher's own experiential knowledge, and second, that research participants have a right to participate in research that is about them. They argue that researchers are also research participants and that the co-researchers are also the co-subjects. These two principles do not apply within constructivism (where there is no identified epistemological role for experiential knowing); researchers are not also subjects and the findings are grounded in the experiential knowing of others. Heron and Reason argue that participatory research differs from other forms of qualitative research in that research participants inform the research design and inform how knowledge is generated about them. They also argue that the purpose of research within the participatory perspective is closer to the purposes of critical theory – 'the critique and transformation of social, political, economic, ethnic, and gender structures that constrain and exploit humankind' – than constructivist, where the aim is about 'understanding and reconstruction' [19, p. 285]. The aim therefore is to create a situation in which participants give and receive valid information and are committed to the outputs of the study.

Social scientists are frequently faced with the dilemma of rigour or relevance. From the participatory action research perspective the aim is to define the standards of appropriate rigour and then meet them without loss to the relevance of the study.

Knowledge, Values, and Ethics

Knowledge is the result of collaboration and is built up from this collaborative relationship. Participatory action research emphasises the importance of a 'living knowledge' that is linked to the practical knowing (how to do something) that comes from being grounded in the situation within which an action occurs. Participatory action research maintains that research subjects have a basic human right to be engaged in research that intends to gather knowledge about them. The roles of values and ethics are embedded into the study; the subjects are also the researchers and the researchers also the subjects.

Is there Conflict with other Perspectives?

Participatory action research relates closely to both critical theory and constructivism, but uses the same type of measurement and standards as positivism and post-positivism. Arguably, the movement towards action research has come about as a result of non-utilisation of research findings and a desire to conduct research that will result in recommendations being implemented.

Reconciling and Combining Research Frameworks

The type of philosophical framework underpinning a study has implications for how the research is conducted, who has control of the study, how quality is assessed, how values and ethics are viewed, and, ultimately, the type of knowledge that is produced and what is done with that knowledge. The researcher's role also differs depending on the perspective.

Guba and Lincoln [17] stated that: 'Within the last decade the borders and boundary lines between these paradigms and perspectives have begun to blur' (p. 105). Rather than philosophical perspectives working in competition, they are more often combined into one study to inform the arguments and answer a broader research question. Perspectives can be blended together into two main groups: first, the positivist and post-positivist, which share important elements; and second, the critical theory, constructivist, and participatory perspectives, which also share important elements. However, these two main groups are not easily combined into one model as their assumptions about reality and objectivity differ.

Positivism has been the dominant research perspective for many centuries. However, in more recent years the superior status of quantitative research approaches within education and the social sciences has been challenged. Criticisms of quantitative approaches have included arguments about 'context stripping' (taking data out of context and thereby removing much of the associated meaning), that by focusing on the majority or dominant view, important messages from the minority are ignored, and that even in well-controlled experiments researchers and subjects can influence each other and bias the results. In 1994, Guba and Lincoln [17] reported that the dominant perspective was the post-positivist perspective. Post-positivists tended to have the power and influenced numerous decision-making processes, namely research funding, journal publications, and committees for promotion. However, proponents of critical theory and constructivism have gained ground and recognition over the past 40 or so years, with more journals, journal articles, and qualitative research. Participatory action research is also emerging as a perspective. In 2005, Guba and Lincoln [18] acknowledged that 'the number of qualitative texts, research papers, workshops, and training materials has exploded' (p. 191) and pointed out the distinct turn towards the more recent perspectives.

Writers such as Guba and Lincoln suggest that the use of a particular method implies commitment to a particular philosophical perspective and its associated ontology and epistemology. This position assumes that a methodology is necessarily indicative of particular assumptions about knowledge creation. This position is challenged by Bryman [28] who argues that research methods are more 'free-floating' in terms of ontology and epistemology than was previously proposed. Platt also highlights that researchers are more pragmatic (Platt 1996, quoted in Bryman [28, p. 619]). Bryman continues that research that combines both qualitative and quantitative approaches in one study illustrates that these research methods can be autonomous. Patton [29] concurs with the views of Bryman, commenting first on the parallel status of qualitative to quantitative research and on the increased use of multiple methods which, when used together, can provide a fuller answer to the research question.

> Several developments seem to me to explain the withering of the methodological paradigms debate [29, p. 302].

Patton goes on to list a number of developments that explain the move towards mixed methods. For example, the importance of methodological appropriateness rather than paradigm orthodoxy, that the strengths and weaknesses of both qualitative and quantitative approaches are better understood. Maxwell [30] discusses the use of quantifying qualitative themes, moving away from the use of vague terms such as 'some' and even conducting statistical analysis on the number of themes reported. See Box 27.2 for examples of how to combine qualitative and quantitative data in mixed methods research.

The work on realist evaluation by Pawson and Tilley [31] goes a step further. The authors note that realist evaluation sits between positivism and constructivism, positing that social reality cannot be measured directly (due to the weakness of the human researcher) but can be known indirectly. This approach is close to the post-positive ontology but with a pluralist epistemology:

> One can imagine the attractions of a perspective which combines the rigour of experimentation with the practical nous on policy making of the pragmatists, with the empathy for the views of the stakeholders of the constructivist [31, p. 24].

The changing philosophical perspectives associated with *grounded theory* illustrate that research methodology does not necessarily belong to one underlying research philosophy. The grounded theory methodology was originally based on the work of Glaser and Strauss [32]. Glaser brought epistemological assumptions and methodological terms, and Strauss brought the study of process and meaning. Charmaz [33] placed Glaser and Strauss' [32] original grounded theory in the post-positivist perspective. They argued that Glaser's position came close to a traditional positivist stance with assumptions of an objective, external reality and a researcher who remains neutral and discovers data. The later work of Strauss and Corbin is considered post-positive as they proposed giving a voice to the respondents; in even later work Strauss and Corbin [34] took a constructivist stance [35]. Charmaz, who was a former student of Glaser and Strauss, has developed constructivist grounded theory by seeking the meaning of both respondents and researchers and by looking more for beliefs and values as well as acts and facts. Bryant and Charmaz [35] have repositioned grounded theory within constructivism. This example highlights that the linkage of research philosophy to methodology is not fixed; by changing the emphasis in a methodology, it can become compatible with another research philosophy.

Box 27.3 illustrates how two contrasting perspectives can illuminate the same research area, and provides a summary of two papers by O'Cathain et al. [36] and Stapleton et al. [37]. These abstracts show both quantitative and qualitative approaches being used within the same study and highlight some of the differences behind the qualitative and quantitative traditions.

It is possible to identify the post-positivist stance of O'Cathain and colleagues in the quantitative study, which attempted to control variables while manipulating others. There was concern with numbers and measurement and reporting of findings in terms of statistical differences. There was also concern about using the 'correct' measurement, and fears were expressed about contamination of the intervention by earlier exposure to the leaflets.

In the qualitative study, Stapleton and colleagues were less concerned with numbers and measurement and more

BOX 27.2 HOW TO: Combine mixed methods research using qualitative and quantitative data

- *Triangulation:* use different types of research data to cross check findings from another source. For example, a standardised survey tool (drawing from post-positivism) may seek to measure professionalism and is combined with an interview study which seeks to understand why respondents have answered questions a certain way, or why they hold certain views about professionalism and what it means to them (constructivism); both contribute a different type of knowledge and understanding.
- *Provide hypotheses:* qualitative data could be used to identify hypotheses that could be tested later using a survey gathering quantitative data.
- *Aid measurement:* qualitative data could be used to inform and develop survey items to be tested with a population.
- *Screening:* use quantitative data to screen for people with specific characteristics for an in-depth qualitative study.
- *Fill gaps:* one methodology may not provide all of the information and knowledge to fully answer a research question.
- *Snapshot versus process:* quantitative data will provide a single snapshot at a point in time, whereas qualitative data can provide detailed information about a process.
- *Where two types of data are required:* sometimes both numeric data and data about meaning and holding a certain view or having had a certain experience are required.
- *Quantification:* use qualitative data to identify problems and quantitative data to quantify the problem.
- *Explaining the relationship between variables:* quantitative data frequently need to explain the relationship between variables; this can be explored further by a follow-up qualitative study.
- *Exploring the micro and macro:* use of both approaches allows a study to explore the different levels of a problem.
- *Solving a problem:* use of a different research strategy to the one already employed to explore unexpected or puzzling outcomes.

See Bryman [28] for further discussion on the subject.

BOX 27.3 Comparison of two linked studies

Quantitative study

O'Cathain, A., Walters, S.J., Nicholl, J.P. et al. (2002). Use of evidence based leaflets to promote informed choice in maternity care: randomised controlled trial in everyday practice. *British Medical Journal* 324: 643–642. [36]

This study was a randomised controlled trial with the aim of assessing the effect of leaflets on promoting informed choice in women using maternity services. The sample was clearly defined as women reaching 28 weeks' gestation before the intervention took place. Outcomes were assessed using a postal questionnaire. Various means were used to test the validity of the questionnaire, and a power calculation was used to identify the sample size needed to detect a 10% difference between the intervention and the control groups. Results included response rates (reported in numbers and percentages) and further analysis to identify any differences that could be related to age, social class, parity, pain relief, and type of delivery. There was an attempt to examine confounding factors that would bias results, such as having been given the leaflets on another occasion prior to the start of the study.

The conclusion was that the evidence based leaflets were not effective in promoting informed choice in the women. The authors reported on the limitations of the study and expressed concerns over their measurement of informed choice and the power of the study to detect a difference. Authors referred to the qualitative findings below for further explanation.

Qualitative study

Stapleton, H., Kirkham, M., and Thomas, G. (2002). Qualitative study of evidence based leaflets in maternity care. *British Medical Journal* 324: 639–642. [37]

The stated aim was to examine the use of evidence based leaflets on informed choice in maternity services. The design involved both non-participant observation of antenatal consultations and in-depth interviews with both the expectant mothers and the health professionals. The sample was initially opportunistic (depending on which staff were doing the clinic and which women agreed to be involved), but progressed to be more selective to ensure that women from all childbearing ages, social class, minority groups, and current and past obstetric histories were represented. Observations were used to help identify how the leaflets were used, and field notes made on the setting, actions, words, and non-verbal cues. Semi-structured interviews were conducted using an interview guide. A grounded theory approach was used [32, 34] so, as the interview progressed, interviewees were selected to help confirm or refute emerging theory, until no new information was gathered (theoretical saturation). Validity and reliability were said to be ensured by using several researchers and experts, to 'guard against any researcher dominating the analytical process'. Results were reported in terms of emerging themes, and quotes were used to illustrate them. The qualitative study revealed that time pressures and competing demands within the clinical setting undermined the intervention.

The observations revealed that health professionals rarely differentiated the leaflets from other information that they offered or discussed with the women. The interviews identified that the women confused the leaflets with other information they had been given or denied having received them. The midwives reported that hierarchical power structures resulted in obstetricians defining the choices possible, resulting in informed compliance rather than informed choice.

concerned with gaining a wider range of views and identifying the issues related to the intervention from multiple viewpoints. Observed behaviour was used to identify how the intervention was implemented, and findings were generated from observer notes. Analysis was conducted by looking for common themes in the data. The quantitative study reported that the intervention was not effective, and the qualitative study explained why.

Having an understanding of what each perspective aims to achieve can increase our understanding and provide an appreciation of the different types of knowledge produced rather than viewing one approach as superior to others.

Practical Considerations when Starting Research

The Research Question

Most researchers have little problem identifying the general field in which they wish to conduct their research, but have more difficulty finding a focus and pinning down a research question or objective. Punch [38] makes a distinction between general and specific research questions. The hierarchy offered by Punch can be illustrated using the study by O'Cathain et al. [36] which can be summarised as follows (see also Box 27.3).

Research area	Maternity care
Research topic	Informed choice
General research question(s)	Does informed choice change behaviour?
Specific research question(s) or objective(s)	To assess the effects of leaflets on promoting informed choice in women using maternity services
Data collection question(s)	• Do women who receive the intervention answer 'yes' more often to the question 'Have you had enough information to make choice on …?' • Do the women who receive the intervention report greater satisfaction with antenatal information? • Do the women report being given at least one leaflet?

Novice researchers sometimes confuse data collection questions and research questions [38]. A research question is the question that the research is attempting to answer, whereas the data collection question is asked in order to collect data that will be used to answer the research question.

Coming up with a research topic is about following your interests (it is difficult to sustain interest if it is not there from the start). Looking around, listening or experiencing something, or being aware of current issues are all sources of inspiration. Think about what is known and what is not known about something. O'Leary [39] suggests the use of concept maps to help identify an area of interest. Bell [40] advises that a good first step is to simply talk over your research ideas with a colleague. Gaining another perspective early on can be very valuable. Once identified, the general area of interest needs to be narrowed down. A good research question needs to be feasible; this relates to the research expertise and resources available and, indeed, whether the question is capable of being answered at all. This last point involves checking with those who have more expertise in research and knowledge of the field of study.

A good research question not only gives the research focus and direction, but also sets boundaries. Boundaries are particularly important for novice researchers, who have more difficulty estimating how much research time is required to undertake a study and may need to limit both the size and the scope more than anticipated. Defining the terms used within the research question identifies the criteria of concern and, by exclusion, sets some boundaries on the study. Specifying a research question involves identifying the concepts or variables of interest and, where possible, identifying suitable indicators for the variables of interest. It is important to check that any assumptions made by the question are correct. Deciding on whether you need to frame your question as a hypothesis depends on the theoretical perspective the research will be framed in, and on the type of question being asked. Research within the positivist and post-positivist perspectives is more likely to contain a hypothesis, but the key question is whether the research question forms a testable statement about the relationship of one or more variables to others. Research that is exploratory or framed within the new perspectives is unlikely to start with a hypothesis. The research question should be a pointer to the methods to be used and indicate what type of data will be needed to answer the question.

The Research Proposal

All research should start with a proposal, also referred to as a *protocol*. It can be helpful for novice researchers to see another research protocol (and an application for ethical review) first to identify what is required [38]. A proposal is a plan of action, a communication on which approval to commence the study is given, and is a contract between the researcher and supervisor, university, any funding source, and ethics committee [41]. The proposal describes the research background, including relevant literature, the research question, methods, details about recruitment of the intended sample, and how the data will be analysed. The protocol starts with the relevant literature by 'setting a scene' or 'telling a story' of what is known, how the knowledge has built up to form our current understanding, and where relevant gaps in knowledge exist. The research question follows; this should extend our understanding and ideally address an identified gap.

Literature searches are mainly conducted online using databases, such as Medline, education databases, and Ovid, and key articles selected following searches on keywords or authors. Punch [38] reports that two common criticisms of literature reviews in dissertations are that they are not thematic, tending to be chronological or presented serially, and they are not properly integrated with the study. These criticisms can be addressed by creating a conceptual framework into which the literature can be organised.

The research question (or objective) should indicate appropriate methods of data collection. The two studies presented in the O'Cathain paper had the research objective: 'To assess the effects of leaflets on promoting informed choice in women using maternity services' [36, p. 643]. The objective suggests measurement in the use of the term 'assess', the leaflets were defined as '10 pairs of Informed Choice leaflets', the women were defined as 'women reaching 28 weeks' gestation', and so on. 'Effects' were measured using a questionnaire. Assessment came in the design of comparing a control group with the women who received the leaflets. For the qualitative part of the study the research objective was: 'To examine the use of evidence based leaflets on informed choice in maternity services' [37, p. 639]. The term 'examine' suggests 'look at' rather than 'measure', and again the leaflets and maternity service were defined. Outcome measures were views and responses from the expectant mothers and the staff. Exactly how the methods are arrived at will be influenced not only by the research question, but also by the interest and expertise of the researcher, supervisor, and team.

The protocol should include details about who will be recruited into the study and from where, how recruitment will take place, and the numbers involved (this may require a power calculation for research designs). This should be followed by a detailed description of the research procedure. A plan or flow diagram will be useful if the procedure is complex. Details of how the data will be collected and analysed, and any planned statistical tests, should be included, and a timeline or Gantt chart is useful to work out when each activity is planned to start and finish. A breakdown of the costs involved in the study for staff and research activities, among others, is also needed, as well as plans for the dissemination of findings [38, 39, 42] (see Box 27.4).

Ethics in Research

Ethics is concerned with rules of conduct and principles relating to moral behaviour. Researchers are responsible for ethical decisions from formulation through to the dissemination of research. As discussed above, the type of research framework influences how ethics is regarded in the study, as well as appreciating other 'realities' and empowering voices otherwise not heard. All types of study involve making ethical decisions about what is right for the research participant, as well as considering the interests of the researcher, the funding body, and the study itself.

BOX 27.4 HOW TO: Write a research protocol

Title
- Provide a clear title that articulates the aim or research question and if possible provides information about the research design or method used.

Relevant background literature
- Summarise the work that has already been done in this area.
- Search the relevant databases as well as journals, books, and policy documents, if appropriate.
- Write up thematically if meaningful, or chronologically if the topic changed and developed over time.
- Identify what is missing, and what new research should be conducted. Add any pertinent educational or clinical theory that is relevant to this area of study.
- Include all references at the end of the document.

Research question
- Provide a clearly worded research question or objective.
- Keep it feasible; set boundaries for the study, and think realistically about the resources available, e.g. time, staff, and level of expertise.
- Define what you mean by the terms used.
- Consider including a secondary research question or objective (something of a lower order which you would also like to explore).

Methods
- Study design (e.g. randomised trial, grounded theory).
- Sampling
  - Sampling strategy (e.g. opportunistic, purposive)
  - Define target sample (i.e. demographic details, how selected and recruited to study)
  - Sample size (reason for size, is it informed by a power calculation?).
- Data collection
  - Details of any instruments to be used and references to existing tools
  - Details about validity and reliability
  - Outline stepwise procedures including pre-testing and piloting of tools
  - Data collection methods (e.g. via postal questionnaire, field notes, interviews).
- Data analysis
  - Details of how data will be analysed (e.g. statistical tests, type of qualitative analysis)
  - Details of computer programs to be used in analysis.

Ethical considerations
- State if approval from ethics committee has been received or is in progress.

Plans for dissemination of findings
- Plans for dissemination at local, national and international levels; including reports, publications, conference presentations, project websites, newsletters and blogs, etc.

References

From literature review, methods, instruments, etc.

Appendices
- Costings.
- Research instruments (e.g. questionnaire, interview schedule, consent forms).
- Flow chart summarising study plan with a timeline.

Approaches to ethics follow a set of principles that guide research, including, among others, informed consent, confidentiality, and anonymity. One such principle is that of 'informed consent'. Informed consent includes providing all relevant information about the study and what taking part will involve, including risks. The research participants must be able to comprehend the information and be competent to make a decision about involvement. They should have the opportunity to ask questions and agreement to take part should be voluntary, and free of coercion or influence. Generally, consent will be obtained by asking the research participant to confirm consent by signing a consent form, by giving recorded verbal consent, or by returning a questionnaire. Gaining consent may involve gaining approval from many more people than those directly involved in the study, that is, the host care organisation, in order to access participants. The researcher must also take steps to ensure the participant is protected from any adverse consequences of being in the study and ensure that the identity of the participant is protected.

Consent to take part in research may be given on the basis that the information obtained about the participants will only be used by the researcher and only in particular ways. Confidentiality means protecting the identity of those who agree to take part in research, maintaining the data in a form such that the identity of the participant is protected. This implies keeping names and data separated by using a code that is only accessible to the researchers, and reporting data in a format that does not lead to individuals being readily identifiable. For example, it may involve removing or changing details to protect individuals who would otherwise be identifiable because of their unique characteristics or experiences.

Anonymity goes further than confidentiality, as the researchers do not collect named data at all. This means the researcher cannot identify which respondent gave the data (e.g. postal survey). This type of data allows participants to make any negative comments more freely without fears or concerns that anything they do report might be attributed to them with unknown consequences. For researchers, this might be difficult or impossible to achieve if the methods involve interviewing, and problematic if they wish to send reminders only to those who have not already agreed to participate. For a full discussion on ethics in research, see Israel and Hay [43] and Punch [44], and for ethical dilemmas in qualitative research, see Welland and Pugsley [45].

Returning to our example, ethics questions that may have been addressed before the Stapleton et al. [37] study was carried out include:

- Will the midwives and expectant mothers be given all the information they require to give their informed consent?
- Is there any pressure or coercion to take part?
- How will consent be obtained?
- How will confidentiality of the interaction of midwife and mother be assured?
- How will collected data be anonymised, particularly with reference to the use of quotes?
- Who will have access to the data?
- Have the researchers anticipated all that could go wrong? How would they respond if ethical issues emerged?

In conclusion, researchers are expected to minimise the risk of harm or discomfort to people, to conduct research in a manner that upholds certain principles such as informed consent, and to consider any consequences or harm that may result from the research. Harm from educational or social research is more likely to take the form of psychological distress than physical injury. Conversely, many researchers aim to provide benefit by conducting research that empowers participants, such as in feminist research.

Conclusions

Philosophical perspectives determine the assumptions that are made about reality and what can be known. Positivism became the dominant perspective after the Enlightenment, but following the realisation that not all research fits this paradigm, a shift in thinking occurred. This brought about a new way of thinking about social science, and new and competing philosophical perspectives emerged.

The arguments against combining qualitative and quantitative approaches centre on the acceptance that research strategies are committed to particular philosophical perspectives [17] versus the view that they are autonomous [28]. A growth in the preparedness to view research methods as techniques for data collection and a movement away from concerns about ontology and epistemology have resulted in more research using a combination of both qualitative and quantitative research methods [29].

Getting started in research involves identifying a good research question. After this, consideration needs to be given to the type of data that should be collected to answer the question. A plan or proposal needs to set out how the research will be conducted, with milestones. Consideration needs to be given to ethical questions that affect the research and how these can be addressed.

References

1 Oxford English dictionary (2013). www.oed.com (accessed 16 July 2013).
2 Twycross, A. and Shorten, A. (2014). Service evaluation, audit and research: what is the difference? *Evidence-Based Nursing* 17: 65–66.
3 Smith, R. (1992). Audit and research. *British Medical Journal* 305: 905–906.
4 Casarett, D., Karlawish, J.H.T., and Sugarman, J. (2000). Determining when quality improvement initiatives should be considered research proposed criteria and potential implications. *Journal of the American Medical Association* 283 (17): 2275–2280. doi: 10.1001/jama.283.17.2275.
5 Baily, M.A., Bottrell, M., Lynn, J., and Jennings, B. (2006). The ethics of using QI methods to improve health care quality and safety. The Hastings Center Special Report. http://www.thehastingscenter.org/wp-content/uploads/The-Ethics-of-Using-QI-Methods.pdf (accessed 15 May 2017).
6 Clarke, A. (1999). *Evaluation Research*. London: Sage.
7 Robson, C. (2000). *Small-Scale Evaluation: Principles and Practice*. Thousand Oaks, CA: Sage.

8 Cohen, L., Manion, L., and Morrison, K. (2007). *Research Methods in Education*. London: Routledge.
9 Crites, G.E., Gaines, J.K., and Cottrell, S. (2014). Medical education scholarship: an introductory guide: AMEE Guide No. 89. *Medical Teacher* 36 (8): 657–674. doi: 10.3109/0142159X.2014.916791.
10 Boyer, E.L. (1990). *Scholarship, Reconsidered: Priorities of the Professoriate*. Princeton, NJ: Princeton University of New Mexico.
11 Glassick, C.E. (2000). Boyer's expanded definitions of scholarship, the standards for assessing scholarship, and the elusiveness of the scholarship of teaching. *Academic Medicine* 75 (9): 877–880.
12 Beattie, D. (September 2000). Expanding the view of scholarship: introduction. *Academic Medicine* 75 (9): 871–876.
13 Dauphinée, D. and Martin, J.B. (September 2000). Breaking down the walls: thoughts on the scholarship of integration. *Academic Medicine* 75 (9): 881–886.
14 Fincher, R.-M.E., Simpson, D.E., Mennin, S.P. et al. (September 2000). Scholarship in teaching: an imperative for the 21st century. *Academic Medicine* 75 (9): 887–894.
15 Kneebone, R. (2002). Total internal reflection: an essay on paradigms. *Medical Education* 36: 514–518.
16 Kuhn, T.S. (1970). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
17 Guba, E.G. and Lincoln, Y.S. (1994). Competing paradigms in qualitative research. In: *Handbook of Qualitative Research* (ed. D.K. Denzin and Y.S. Lincoln), 105–117. Thousand Oaks, CA: Sage.
18 Guba, E.G. and Lincoln, Y.S. (2005). Paradigmatic controversies, contradictions and emerging confluences. In: *Handbook of Qualitative Research*, 3e (ed. D.K. Denzin and Y.S. Lincoln), 191–215. Thousand Oaks, CA: Sage.
19 Heron, J. and Reason, R. (1997). A participatory inquiry paradigm. *Qualitative Inquiry* 3: 274–294.
20 Crotty, M. (2003). *The Foundations of Social Research: Meaning and Perspective in the Research Process*. London: Sage.
21 Steier, F. (1991). Introduction: research as self-reflectivity, self-reflexivity as social process. In: *Research and Reflexivity* (ed. F. Steier), 1–11. London: Sage.
22 Marx, K. (1961). *Selected Writings in Sociology and Social Philosophy*. Edited with an Introduction and notes by T.B. Bottomore and M. Rubel, 1–11. London: Watts.
23 Schwandt, T.A. (1994). Constructivist, interpretivist approaches to human inquiry. In: *Handbook of Qualitative Research* (ed. D.K. Denzin and Y.S. Lincoln), 118–137. Thousand Oaks, CA: Sage.
24 Guba, E.G. and Lincoln, Y.S. (1989). *Fourth Generation Evaluation*. Newbury Park, CA: Sage.
25 Whyte, W.F. (1991). *Participatory Action Research*. London: Sage.
26 Glasson, J., Chang, E., Chenoweth, L. et al. (2006). Evaluation of a model of nursing care for older patients using participatory action research in an acute medical ward. *Journal of Clinical Nursing* 15: 588–598.
27 Wadsworth, Y. (1997). *Everyday Evaluation on the Run*, 2e. Melbourne: Allen & Unwin.
28 Bryman, A. (2001). *Social Research Methods*. Oxford: Oxford University Press.
29 Patton, M.Q. (2002). *Qualitative Research and Evaluation Methods*, 3e. London: Sage.
30 Maxwell, J. (2010). Using numbers in qualitative research. *Qualitative Inquiry* 16: 475–482.
31 Pawson, R. and Tilley, N. (1997). *Realistic Evaluation*. London: Sage.
32 Glaser, B. and Strauss, A. (1967). *The Discovery of Grounded Theory*. New York: Aldine.
33 Charmaz, K. (2003). Grounded theory: objectivist and constructivist methods. In: *Strategies of Qualitative Inquiry* (ed. D.K. Denzin and Y.S. Lincoln), 249–291. Thousand Oaks, CA: Sage.
34 Strauss, A.L. and Corbin, J. (1990). *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. London: Sage.
35 Bryant, A. and Charmaz, K. (2007). *The Sage Handbook of Grounded Theory*. London: Sage.
36 O'Cathain, A., Walters, S.J., Nicholl, J.P. et al. (2002). Use of evidence based leaflets to promote informed choice in maternity care: randomised controlled trial in everyday practice. *British Medical Journal* 324: 643–646.
37 Stapleton, H., Kirkham, M., and Thomas, G. (2002). Qualitative study of evidence based leaflets in maternity care. *British Medical Journal* 324: 639–642.
38 Punch, K. (2000). *Developing Effective Research Proposals*. Thousand Oaks, CA: Sage.
39 O'Leary, Z. (2004). *The Essential Guide to Doing Research*. London: Sage.
40 Bell, J. (1999). *Doing Your Research Project: A Guide for First-Time Researchers in Education and Social Science*, 3e. Buckingham: Open University Press.
41 Loche, L.F., Spirduso, W.W., and Silverman, S.J. (1993). *Proposals that Work*, 3e. Newbury Park, CA: Sage.
42 Tarling, R. (2006). *Managing Social Research: A Practical Guide*. London: Routledge.
43 Israel, M. and Hay, I. (2006). *Research Ethics for Social Scientists*. London: Sage.
44 Punch, M. (1994). Politics and ethics in qualitative research. In: *Handbook of Qualitative Research* (ed. D.K. Denzin and Y.S. Lincoln), 83–98. Thousand Oaks, CA: Sage.
45 Welland, T. and Pugsley, L. (2002). *Ethical Dilemmas in Qualitative Research*. Hampshire: Ashgate.

28 Quantitative Research Methods in Medical Education

Geoff Norman[1] and Kevin W. Eva[2]

[1]Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada
[2]Department of Medicine and Centre for Health Education Scholarship, University of British Columbia, Vancouver, British Columbia, Canada

KEY MESSAGES

- There is much more to good research than rigorous design.
- The appropriateness of a particular research design is dependent on the question to be addressed.
- The hierarchy of research methods commonly applied to clinical studies is inappropriate when judging the strength of research strategies in educational domains.
- The value of critical, synthesising, theoretically oriented, and empirically based reviews of the literature cannot be overstated.

Introduction: The Quantitative Paradigm

Quantitative research methods have been central to physical science for centuries, dating back as far as the astonishing developments in astronomy in the 1500s. Indeed, it is difficult to envision natural science without quantification, and it is even more difficult to delve into some of these accomplishments without feeling an overwhelming sense of wonderment at the ability of the scientist, whether we ponder on our understanding of the very big (e.g. cosmology) or the very small (e.g. particle physics), or closer to home, as we reflect on the rapid evolution of digital electronics in our lifetime. For those old enough to remember a computer card, a 1 GB flash drive contains as much memory as 10 000 000 computer cards – about the volume of a one-car garage. But quantitative methods are not a panacea. Many would claim that social scientists have been too quick to adopt the methods of natural science unquestioningly and have not given adequate recognition to the complexity of social situations, which are not evidently reducible to a few numbers. In Chapter 27 of this book, Illing reviews the history of the adoption of quantitative methods into social science, and the subsequent uneasy integration (or partitioning) with qualitative methods. It is not surprising that the quantitative 'dust bowl empiricism' led to a counter reaction. It is embarrassingly easy to find examples in medical education that, on a moment's reflection, exemplify the silliness of attempts to reduce the complexity of human interaction in an educational setting to a 'treatment' that half receive and half do not, and an 'outcome' such as pass or fail on an examination [1]. Although we are personally wedded to quantitative approaches to social science research, such studies, which, as Illing points out, reduce the people in the study to 'objects' that are supposed to absorb exactly the same dose of the educational drug, amount to little more than unintended '*reductio ad absurdum*'.

Little is served by identifying specific examples. The larger question is the extent to which quantitative methods have been aligned with recognised progress in the field. There is simply no dispute that the methods of the natural sciences, from the electron microscope to the wet lab and the clinical trial, have led to enormous advances in medicine with direct consequences for human longevity and welfare. Clearly, it would be specious to try to make similar claims in the small and impecunious field of medical education. Nevertheless, the past three decades of research have seen substantial advances in medical education, much of it directly related to the application of sophisticated quantitative methods [2], particularly in the area of student assessment.

In writing a chapter on quantitative methods, our goal is not to promote these methods over qualitative research strategies or even to contrast the two, but rather to provide some guidance for those trying to better understand the variety of quantitative methods available. As Bordage notes [3], the community should move on from the qualitative–quantitative debate because 'this oft-repeated debate is not productive … each approach is useful in its own right and is often most productive when complementary'. In fact, lost in the qualitative–quantitative debate is the complexity of both qualitative and quantitative methods.

Much is written about the various schools of qualitative research; Illing, as one example, cites post-positivism,

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

BOX 28.1 Quantitative research traditions

- Experimental
- Epidemiological
- Psychometric
- Correlational
- Reviews and meta-analyses

critical theory, constructivism, and participatory action as various *genres* in qualitative research. To our knowledge, no similar taxonomy exists for quantitative research. Indeed, many critics of quantitative methodology appear to presume that quantitative methods in educational research amount to testing hypotheses using randomised experiments that are proven or disproven by application of statistical methods. This is a woefully inadequate description of, for example, the psychometric methods that have led to such significant advances in assessment methods.

In this chapter, we will distinguish four research traditions – experimental, epidemiological, psychometric, and correlational – exploring some basic principles of measurement and statistical inference along the way. Finally, we will describe the methods of meta-analysis and systematic reviews, and contrast these strategies with those of reviews that are better defined as critical and theory-oriented. We begin, though, with a commentary on the importance of precisely focusing one's research question, emphasising that while good studies require good methods, the quality of a study is not completely defined by its methodological rigour (Box 28.1).

The Research Question

In the previous section, we noted that many authors equate quantitative research with hypothesis testing. In our view, not only is this association simplistic, but it gives insufficient attention to the nature and adequacy of the research question. Far too frequently, when students do discuss the research question, much effort is expended in learning how to convert a practical, common sense question into a formal research hypothesis or, even better, a 'null hypothesis' that frames it as no difference, no effect, or zero correlation. Such efforts, by promoting precision of planning, can enable one to ensure that the research question is ultimately answerable. Much of that precision, however, becomes evident in any case as the research design and methods are devised, thereby leading us to believe that much of this effort is stylistic and does little to fundamentally improve the research. It really is of little consequence if a question is framed as a question, a research hypothesis, or a null hypothesis.

In any case, the idea of a research hypothesis only applies well to some kinds of quantitative research. The development of a new evaluation instrument will proceed by a very different route, with studies of reliability and validity. To frame a reliability study as a null hypothesis would look something like: *The reliability of the new written test of reflective practice will be less than 0.5.*

This framing simply does not adequately capture the goal of the research. While our views may not represent a majority position, we believe that we should focus effort on an adequate research *question*, and forget about the niceties of null hypothesis creation. In the end, the goals of the research question are to reduce the possibility of *post-hoc* explanations, to specify or constrain the methods used to answer the question, and to enable careful analysis of whether or not the focus of the research is achievable. As human beings we are remarkably good at generating explanations for any pattern that is presented to us, so scientists try to avoid drawing conclusions without some a priori reason to have predicted the result. That is not to say that *post-hoc* speculation is not useful when unanticipated findings arise, but it is to say that further study should be engaged to confirm the result before running to the printers. Style is unimportant, but precision is invaluable to avoid wasting time and resources in pursuit of ultimately unanswerable questions.

So how do we ensure that the research question is good? Compared to the many approaches that exist to determine if the research methods are good, there is relatively little written about what constitutes a good research question. What does exist tends to focus on the technicalities of what information should be present in the question [4]. There is good reason for this. The worth of a research question cannot be judged in isolation, but can only be viewed in the context of the specific research domain in which it arises. Research is about discovery, and there is little point in discovering what is already known (Box 28.2).

'Discovery' is a useful way to think about the potential contribution of a research study, but the word implies such large leaps from current understanding that we prefer to think of the act of discovery in terms of knowledge building. Arguably, the value of a study is directly related to the extent to which it reveals some new understanding of the world – it 'discovers' some new insight or builds on existing knowledge in a meaningful way. Quantitative research is judged, in large part, by the extent to which the lessons learned can be generalised to other contexts. But what advice can we give the beginning researcher to help him or her identify research questions that are most likely to yield knowledge-building discoveries? Very simply, this is the role of the literature review.

The latter point is worth reinforcing – this emphasis on using the literature is not meant to imply that there are no practical (potentially atheoretical) questions that need to be addressed; rather our view is that grounding one's study in the existing literature is a valuable strategy for ensuring that even purely practical research projects have a decent chance of being successful and contributing to the generalisable knowledge base of the community. Too frequently, the literature review reads like a child recounting a playground fight: 'Johnny did this. Then Sally did that. Then Bob came along and said this other thing.' The literature review is not, and should not be, a chronological recounting of who did what to whom, but rather, should

BOX 28.2 HOW TO: Select an issue worthy of research

When contemplating which research ideas are worth developing into more fully fledged projects, there are a variety of criteria that should be used. Here are just a few guiding principles:

- *Novelty:* Has the study been done before? It is insufficient to say 'to our knowledge this has never been done' without a concentrated effort to determine whether or not it has been done. Talk with experts, be they local or international, and scour the literature for other studies aimed at similar issues. At this point in history it is inconceivable that no one has ever written anything of relevance to whatever topic you care to study. Only after a careful search of a variety of literatures can you make a compelling argument for how your particular study could advance understanding in some meaningful way beyond what has already been done.
- *Importance:* Medical education is an applied field. As a result, while any given study might not yield immediate practical implications, it should be possible to conceive of ways in which the research efforts might beneficially impact on the field in the future. Use your 'on-the-ground' experiences to inform your research questions just as much as you use the literature.
- *Programmatic:* Too often we in applied fields think of research in terms of projects as opposed to programmes. The latter term, put forward by the Hungarian philosopher of science Imre Lakatos, should draw attention to the fact that real advances are typically made through systematic and long-term study of a particular issue [5].
- *Guided:* What is the conceptual framework within which your study fits? [5] Which theories speak to the issue with which you are trying to grapple? Do they contradict one another in a way that you could inform through your research efforts? Is your theory/hypothesis falsifiable (i.e. could your study design yield an answer that would counter the theory you are using as a guide)?
- *Grounded:* Related to some of the previous points, we use the term 'grounded' to indicate that the study should be grounded in the context of what is already known such that the context is used to determine which of the various possible research directions is most appropriate (i.e. most likely to provide meaningful results to the community) at this point in history. One may not be able to predict this with perfect accuracy, but the issue should be considered through broad consultation and reflection.

establish a conceptual framework within which the present study will reside [5].

The literature review should clearly identify knowledge gaps, and the gaps should be substantive enough to warrant filling; statements such as 'this study has never been done in our country/city/university/discipline' are weak justification. The emerging conceptual framework should be such that it helps the researcher – and eventually his or her readers – focus on the bigger picture (the 'state of the art'), and should clearly delineate how the present study adds to this knowledge. This implies that the researcher should begin every study with a formal literature review, from which all unanswered questions will become as evident as a full moon on a clear night. Of course, research does not proceed in such a linear fashion, but one characteristic of mature research programmes (i.e. long-term and systematic exploration of a domain [6]) is that situations where a study is created de novo from a literature review are the exception. When a programme of research is ongoing, new study questions arise from existing study findings.

So what can be done to ensure the quality of the research question? One way to answer this question is to draw on the notion of 'theory'. In contrast to a research hypothesis, which ultimately leads to only two conclusions: (i) it worked or (ii) it did not, a research theory involves an understanding of the interaction of multiple variables. Such theory-based research is conspicuously in the minority within our field [7]. Only about half the articles in a recent review were identified as having a conceptual framework [8] let alone a testable theory. Yet another 'study' showing that students gave your new course in anatomy an average rating of 4.5/5 is unlikely to provide any new insights into teaching and learning. But application of a novel theoretical perspective may alter the way the community thinks about the issue and reveal insights that are relevant to various curricular strategies. Although educational theories typically do not make quantitative predictions, they nevertheless often involve the interplay of several variables and may get people thinking about the problem in a more refined manner. That is the importance of theory; by the time the theory has been subjected to a critical test, inevitably involving a few to a few dozen studies, we can gain good insight into the limits and generality of the findings (thanks to an accumulation of evidence) rather than being stuck with a series of only superficially related results.

It is worth noting here that a defining property of scientific theory is, to use Popper's words [9], that it is falsifiable, i.e. it can be proven wrong. As such, studies that use theory as a basis for knowledge building have a more dynamic quality than typical invocations of theory as context. Scientific theories are not permanent and immutable; we expect them to change and evolve (and increase in explanatory power) as new evidence arises. This stands in stark contrast to the use of theory as justification implied by statements such as: 'The curriculum was designed to be consistent with theory Y.' Such statements are not terribly useful, as many theories can be implemented in countless ways and other theoretical positions might also promote the same types of learning activities [10]. More critically, many theories are framed in such sweeping generalisations

		Considerations of Use	
		Low	*High*
Quest for Understanding	*Yes*	Pure basic research *e.g. Bohr*	Use-inspired research *e.g. Pasteur*
	No	-	Pure applied research *e.g. Edison*

Figure 28.1 Pasteur's quadrant (*Source*: From Stokes [12]).

as to be virtually unfalsifiable. Just because a set of data 'is consistent with' one theory does not in any way amount to an accumulation of knowledge unless the data can be shown to be inconsistent with some alternative theory or unless they lead to revision and refinement of the theory being utilised.

It might also be noted that, while some notions of theory building remain firmly rooted in a positivist tradition, theory is probably best used when it moves some towards the recognition that the world is not adequately described by a single 'it worked/it did not' reductionistic package [11]. Further, as we alluded to earlier, the fact that research is theory based does not mean it is irrelevant to practice. On the contrary, as Stokes [12] has convincingly described, theory-based basic science and practically oriented research efforts should not be construed as lying at separate ends of a continuum. Rather, he argues, the two agendas should be considered orthogonal continua in their own rights, with the best research lying in what he called 'Pasteur's quadrant', to reflect the great strides Louis Pasteur made in advancing fundamental knowledge about bacteriology, while simultaneously having practical, real-world, impact in the wine and silk industries, and, of course, medicine (Figure 28.1).

Research Designs

When the phrase 'research design' is mentioned, many individuals who grew up in an educational research environment automatically think about experimental designs, quasi-experimental designs, and Cook and Campbell [13]. Those who are closer to clinical research are more likely to think of the epidemiological classifications of case-control study, cohort study, and randomised trial. Both are inadequate taxonomies. Within medical education, much of our quantitative research, in particular psychometric and correlational research methods, does not fit neatly into any of these pigeonholes. Furthermore, which tradition one adopts should be tailored to the specific research questions one is trying to address.

In addition, different research design traditions arise from different kinds of question. In the remainder of this chapter we will examine various issues related to selecting a particular research design. The next section deals with the experimental tradition by exploring a set of methodologies primarily aimed at testing questions of causality (e.g. 'Does increasing test frequency cause better retention of studied material?'). A related tradition that we will then examine briefly is the epidemiological approach. Because many methodological reviews address research design on the continuum of case-control study, cohort study, and randomised trial, we will define these terms and show some (limited) applications in medical education.

The subsequent section will focus on the psychometric tradition, a method of study primarily directed (at least within educational circles) at the development of better measures of various aspects of competence or other outcomes of the educational process. These studies do not speak of interventions, control groups, outcomes, and the like. Instead, the preoccupation is with issues of reliability and validity, which are indices of the ability of the instrument to differentiate between individuals in a defensible manner.

Finally, we will look at correlational research designs that tend to be used when the desire is to seek understanding by examining relationships among measured variables. As these measurements are frequently gathered from surveys and rating forms, we will also examine some basic principles of questionnaire design.

As can be seen from this introduction alone, each tradition has different aims, and the design conditions necessary for one may be exclusionary of another. As one example, which we will elaborate on later, correlational research requires individual variation in order to see relationships, while experimental research does its best to eliminate individual differences in order to detect treatment effects. There are other differences: experimental research is almost by definition prospective, whereas correlational research is often conducted on existing databases. Use of existing data, while often necessary, can encourage an attitude where the questions are driven by the available data, rather than the reverse, adding little in the way of advancing understanding. That danger noted, it is certainly true that prosaic questions are not proprietary to any one research approach, and conversely, some of the most interesting research has arisen from retrospective analysis of large institutional databases [14, 15].

The Experimental Tradition

The act of discovery, which is central to science as we discussed earlier, is often, although not entirely, directed at identifying causal relationships among things (variables). The experimental tradition exemplifies this agenda. The basic notion of an experiment is that there is a relationship between the independent variable, which is usually under the control of the experimenter, and the dependent variable, which is observed to change as a consequence of the intervention. Many methodological discussions focus on devising studies that can allow one to unambiguously infer causal relationships between experimenter-controlled independent variables and observed dependent variables.

Statements like the following exemplify the causal goal of the experimenter:

- the absorption of a neutron by the ^{32}P nucleus decreases its stability, making it radioactive and leading to decay to ^{32}S by emission of a β particle
- excess sodium chloride in the diet leads to hypertension that results in increased risk of stroke
- a half-day nutrition workshop given to patients with transient ischaemic attacks increased compliance with a restricted salt diet.

Yet while all these statements imply causation – an independent variable that 'causes' a change in a dependent variable – none mentioned the word 'cause'. Further, the meaning of causation is very different as we proceed from top to bottom, and the steps one must take to ensure a valid test of the inference are correspondingly more and more complex. For the neutron, there is no ambiguity. Everyone in atomic physics knows what a neutron is, how to 'make' one, and how to get a phosphorus nucleus to absorb it. It is relatively easy to create a neutron target that is 100% phosphorus. The methods to detect β particles are clear and well understood. Further, the relationship is absolutely causal – if the P nucleus absorbs a neutron, it will eventually emit a β particle (with a known half-life of 14.28 days); if it does not, it will not. No control group of other phosphorus atoms that do not receive neutrons is necessary. While philosophers of science may challenge the reality of a neutron or a β particle, physics practitioners are unlikely to share their concern. However, there is much more uncertainty in the second statement. 'Excess' is not defined, nor is stroke, although there are probably fairly unambiguous criteria for the latter. 'Hypertension' has a definition, but this has drifted lower over the years and is somewhat cohort-dependent [16]. Nonetheless, it is not simply a definitional problem. The causal relationship in this example is far more probabilistic; reduction of salt intake has a fairly small effect on blood pressure, and hypertension is only one contributor to stroke, so excess salt may only 'cause' a small proportion of strokes.

The final causal statement is even more vague. It is difficult to unambiguously define compliance with a diet, and it is more difficult to attempt to define a cut-off point that unambiguously separates 'compliant' from 'non-compliant' patients. Further, it is well-nigh impossible to identify what aspect of the workshop was the causal variable in inducing change, nor, for that matter, is it even likely that any single variable was causal for everyone in the workshop. If the 'causal' relationship is confirmed, this may simply be a stimulus for more research to establish the 'active ingredient' (or combination of ingredients) that led to the change.

Of course, much of educational research resembles the final example rather than the first example. This has two critical implications for our understanding of the role of experimental research in education. First, in contrast to physical sciences, the relationships we seek are inevitably probabilistic, and the signal of a causal relation is almost always swimming in a sea of noise. It is for this reason that we must impose such strategies as control groups, randomisation, and inclusion criteria. Second, the complexity of the relationships may well stifle any serious attempt at understanding processes and mechanisms. To the extent that experiments are directed at, and useful for, discovering lawful causal relationships, it may well be the case that experimental methods like randomised controlled trials (RCTs), which tend to focus on curriculum-level interventions, are over-used rather than under-used in education [17]. That is, although many recent reviews have decried the paucity of good randomised trials in education [8, 18], it makes little sense to conduct elegant studies of interventions that are so complex in their idiosyncracies as to be unreplicable [19]. Still, much can be learnt from good experimentation in appropriate settings.

Study Designs

The essence of the experimental approach is a comparison – between one group of individuals who received an intervention and another who did not. In an ideal situation, the participants in the two groups are as alike as possible before the intervention (which is why randomisation is used in an effort to achieve equivalence), so that any differences observed later can be unambiguously attributed to the intervention and nothing else.

However, although the two-group, intervention–control study design is ubiquitous, it is far from unique. Books on research design dating back over several decades have described many increasingly complex designs [20], and we will discuss some of the more common variants.

One Group: Pre-test–Post-test and Post-test only

A recent review in medical education showed that a single-group pre-test–post-test design was the most commonly reported experimental methodology (32% of 105 studies) followed by a one-group post-test only design (26%) [8]. It is easy to see why this is the case. These designs can easily be incorporated into an ongoing curriculum or course change. All one needs to do is teach something to students taking the course and measure them at the beginning and the end. By contrast, comparisons with a control group require identification of a comparable control group of participants who will volunteer to be tested but will only receive a sham intervention or none at all. Regrettably, one-group designs, labelled 'pre-experimental' by Campbell and Stanley [20], have very limited scientific value. The problems are myriad. Logically, there is no way that whatever changes are observed from beginning to end can be ascribed to the intervention as opposed to competing hypotheses such as maturation, co-intervention, or any number of other plausible explanations.

Further, while the logical flaws may appear parochial, there is a more fundamental educational problem. If one shows a change in performance before and after an intervention, the comparison is against zero change, which implies a comparison against no education at all. While it may well be useful to determine whether, for example, a homeopathic remedy has any effect whatsoever [21], we can pretty well assume that an hour or two of education is going to result in more learning than none – although not always [22]. In the end, a demonstration that students learnt something after a course reveals nothing about the contribution of any specific aspect of the course.

Two Groups: Randomised Controlled Trials and Cohort Studies

This difficulty with identifying a causal agent in one-group designs naturally leads to RCTs. The standard RCT involves randomising participants to two groups, so that at the end of the study the only difference between the two groups (except for chance variation) should be that one group received one intervention and the other did not (or received a second intervention). Randomisation is intended to ensure equivalence at inception; standardisation of interventions facilitates interpretation; blinding avoids bias as does complete follow-up, and so on. In this manner, if a difference is observed, it can unequivocally be attributed to the intervention alone short of the omnipresent influence of chance.

The criteria in Box 28.3 are easy to understand but much more difficult to put into practice. Some aspects, such as random assignment, are not too difficult. Blinding of participants, however, so they do not know what educational intervention they received, is effectively impossible. Indeed, if a student does not know whether he or she received problem-based learning (PBL) or lectures, one would worry about the student, the intervention, or both. Standardisation is much easier with drugs than curricula – what does 300 mg of PBL t.i.d look like? We might recall that teacher differences typically account for twice as much variance in learning as curriculum differences, and it is not clear how one standardises teachers [23].

Let us critically examine these aspects in more detail, and in doing so, identify the art of the possible.

Randomisation, Quasi-randomisation, and Intact Groups

One methodological sine qua non of the experimental approach is randomisation, assignment to groups using a random process. But randomisation is difficult to achieve at times – a student may sign up for the Tuesday tutorial because he or she plays piano on Monday, and may not take kindly to a Monday tutorial assigned by a random number. We forget that randomisation is a means to an end; if students select a tutorial or a hospital rotation in some manner that is highly unlikely to have an impact on their ultimate performance, what we might call 'quasi-randomisation', that may well be good enough. Further, safeguard against bias must scale against the likely size of the treatment effect; if the treatment effect is large, concern about bias can be reduced. Lipsey and Wilson [24] analysed 319 systematic reviews of educational and psychological interventions and showed: (i) an average effect size of 0.45 (effect sizes of clinical interventions are much smaller; one study of aspirin in preventing myocardial infarction had a computed effect size of 0.02) [25] and (ii) no influence of randomisation on effect size; the effects were of equal magnitude (on average) whether randomisation took place or not. Under such circumstances, the potential bias from 'quasi-randomisation' is negligible.

Sometimes randomisation of individuals is just not possible. Students are in one section or another of the course; they are assigned to one hospital or another. A variant of randomisation to deal with this situation is called 'cluster randomisation' – where clusters (e.g. classes) are assigned to one treatment or another. Note that the analysis must account for clusters, and this may have an impact on sample size.

On the other hand, many studies use *intact* groups – turning them into what epidemiologists would call *cohort* studies (which we will discuss later). As one example, many studies have looked at PBL versus lecture-based curricula. Most of these studies involved between-school comparisons. A few studies, usually from the 1970s, involved within-school comparisons, where the school ran a parallel track. Fewer still randomised students to the two tracks. In considering any differences that emerge, between-school comparisons must be viewed with caution, because different schools differ on myriad variables, from the admissions criteria to the cost of tuition. Within-school comparisons may be better, although often students were selected using different criteria in the two tracks or may have self-selected one track or another. Thus, the finding that PBL students have better interpersonal skills [26, 27] must be tempered by the likelihood that the PBL school may well select students for interpersonal skills, or that students with good interpersonal skills may prefer the small group focus of the PBL track [28]. More recently, Schmidt et al. have provided empirical data suggesting that the PBL intervention might alter drop-out rate, leaving groups at the end of the programme that are no longer comparable even were randomisation at the start perfectly effective [29].

The conclusion about allocation is a conditional one. In some circumstances there is a good likelihood that non-random assignment can be viewed as equivalent. In others, this may lead to serious confounding. To make an informed decision about which is most likely to be the case, researchers should gather as much information as possible on dimensions relevant to the question of interest from both groups and judge whether or not differences exist that are strong enough to account for differences observed in the outcome.

BOX 28.3 FOCUS ON: The randomised controlled trial

The randomised controlled trial has only a few critical elements:

- it usually has two groups (occasionally more)
- participants are randomly assigned to each group
- the study is conducted prospectively
- all study participants are blinded to what group they are in
- the intervention(s) are standardised and under experimenter control
- outcome assessment is performed blind on all participants (i.e. the person collecting the data does not know to which group the participants were assigned)
- complete follow-up of participants is achieved.

Placebo or Usual Care

The choice of control group in educational research is rarely given sufficient attention. This is hardly surprising. Programme evaluation is often initiated by someone who has put time and energy into a new curriculum, course, or learning module. It hardly seems worth the effort to now put equal time into a second intervention that is only there for comparison. Consequently, it is often the case that students receiving the innovation are compared with students at another hospital, say, who receive the regular instruction – what epidemiologists may refer to as 'usual care'. A variant is where the intervention is made available to some students but not others.

Such comparisons may be of limited value, regardless of how well other methodological criteria are accomplished. If usual care consists of, for example, lectures that are examined to ensure that the same content is covered, this is fair. If, however, the intervention amounts to a greater amount of time spent studying the to-be-learnt material relative to the control group (a situation that often arises when one simply adds the innovation, such as a high-fidelity simulation, to the curriculum in the treatment arm), then we again find ourselves in the awkward situation of concluding simply that the more they study, the more they learn. From a scientific perspective, a 'usual care' group is about as valuable as none at all, unless the specific aspects of the control intervention can be described as accurately as the experimental arm. Similarly, comparing two groups where one had access to additional resources and the other did not amounts to comparing (A+B) to A; again it amounts to a 'no treatment' comparison.

It is far more informative to compare two experimental interventions where it is possible to standardise for total time of instruction, quality of instruction, or other confounders. As one excellent example of how this strategy would work, Cook [30] has discussed the many studies of e-learning and argued for studies that make comparisons within medium (e.g. both arms of the study use the computer) so that pedagogical variables can be systematically manipulated (i.e. controlled) and the specific medium is not confounded. In deciding on a comparison, one must be careful also to avoid over-controlling the study by equating the two groups with respect to the very variables that are likely to make a difference. The literature on class size, in trying simply to test the impact of size, is a case in point; many studies control the very features of small class discussions (e.g. the opportunity to interact with the professor) that may yield benefits [31].

Blinding

As we said before, one criterion of a good RCT is that all participants – teachers, students, and researchers – are 'blind' as to who is in which group. It may be possible to have outcome measurement performed with blinded assessors or with objective tests; however, it seems highly unlikely that students and teachers will be blinded. But the issue is broader than that. Implicit in the experimental method is that the participant is an 'object', whose motivation and ability is under experimental control. Orwell's vision of 1984 was never achieved, fortunately, and we are left with students who are unlikely (in the extreme) to knuckle under to a researcher's whims. Does this negate the experimental agenda? Not necessarily. But it does ring a cautionary note. To ensure the validity of the study we must make some calculated guesses about the effect of the inevitable unblinding. Failure to do so may lead to false interpretation. As one example, all medical students in North America are highly motivated to pass the licensing examination, for obvious reasons. Consequently, it makes little sense, in our view, to use a licensing examination as a criterion to evaluate a PBL curriculum because student performance on the licensing examination is likely to reflect many hours of study activity unrelated to the curriculum. The outcome may be of interest if it shows a difference, but the numerous studies showing no difference add little to our understanding, and are certainly no basis for any claim of equivalent curricula.

The Perils of Pre-tests

One variant on the RCT is a two-group, pre-test–post-test design. The usual reason for considering a pre-test, to correct for baseline differences, turns out not to be logically defensible, and the potential side effects of a pre-test often go unrecognised. The issues surrounding use of change scores are quite complex, and we can only highlight some [32].

The problem with baseline differences is this. If the two groups were created by random assignment, then any difference between groups arose by chance, and to some extent can be adequately dealt with by statistical procedures, which explicitly examine the role of chance in any observed difference. Pre-tests *may* serve a useful role in identifying whether or not there are baseline differences that should factor into one's interpretation; however, if this is a consequence of non-random allocation, no amount of pre-test correction can control for such differences, simply because any correction involves strong assumptions about the relation between pre-test and post-test. In education, pre-tests have one further serious liability. There is no better way to inform students about what the final test will look like than to give them a parallel pre-test. The pre-test becomes part of the curriculum and has the potential to completely eliminate curriculum differences. In fact, the pedagogical value of testing has become a topic of considerable research in medical education in recent years. Using exactly the same test, both pre- and post-intervention simply magnifies the concern, as highlighted by Larsen et al.'s findings [33], which suggest that the material one is tested on becomes particularly memorable (a phenomenon known as test-enhanced learning or testing effects). One solution that explicitly recognises this issue is called the Solomon Four Group Design. In this design there are four groups:

- pre-test, intervention, post-test
- pre-test–post-test
- intervention–post-test
- post-test.

It is then theoretically possible to disentangle the effect of pre-test from the intervention itself.

Outcomes: Self-assessed versus Performance-based, and Short-term versus Long-term

The choice of the appropriate outcome is perhaps the most difficult part of study design. It almost inevitably represents a compromise between what would be assessed in an ideal world and what can reasonably be assessed with the inevitable constraints of time, money, and acceptability. Moreover, the simple fact is that many outcomes of interest in education (like the CanMEDS roles that have been broadly adopted) are theoretical constructs rather than absolute objective entities [34]. Of course, we would like to show that ultimately the educational innovations we are studying have an impact on patient outcomes, and editors of four journals in the field have argued that this is a goal to be seriously entertained [35–38]. But realistically, with rare exceptions [14], few studies will last long enough to examine patient outcomes. In any case, we agree with Gruppen [39] that such a quest is ill-advised for the more fundamental reason that there are so many intervening variables between educational treatment and patient outcome that it is unlikely any educational intervention will lead to detectable differences.

There is, however, another reason to seek more immediate measures, which aligns with the philosophical commitment to theory-based, programmatic research. While demonstrating that an intervention leads to a (small) increment in performance in, say, a final exam or a licensing examination may be of some practical value, these outcomes are subject to so many confounders that they are unlikely to reveal a cause–effect relationship with the intervention. It is helpful to think of a causal chain where an intervention at each level will have maximal effect on proximal outcomes, and less impact as the chain lengthens. (It is possible that the effects are additive or even multiplicative – the rich get richer – but this appears unlikely in the situations we have examined.) For example, if an intervention can be shown to improve knowledge levels in the first year of medical school and first-year performance is shown to be predictive of clerkship performance, clerkship performance is found to relate to scores received on a national licensing exam, and so on, then we can develop a richer picture of what variables have an impact at each stage, and make more informed decisions regarding the educational activities that should be undertaken. Our recent work on testing the validity of various admissions procedures provides an example of this approach [40]. As an aside, while it is difficult to show enduring effects of curricula, some recent studies have shown that individual differences in students, either in performance or ethical behaviour [14, 41], may have long-term effects that are consequential.

A second issue to consider when deciding on outcome measures is the source. Satisfaction scales, completed by learners, are ubiquitously used as measures of programme effectiveness, probably because of their ease of administration. However, it is difficult to imagine how someone who has spent the time and money to take a course would perceive that they had learnt nothing and it was all a waste – even though some highly touted courses are exactly that [42]. Satisfaction with teaching is moderately related to performance gains [43]; however, this may be a chicken–egg phenomenon as the strongest relations result when students know their scores. Worse still, self-reported judgements of competence have been shown repeatedly to have minimal relationship with observed competence [44–46], thereby making it important that self-assessments are not used as surrogate markers of an individual's performance. That said, recent data do suggest that consideration of self-assessments in the aggregate (i.e. averaged across many individuals) can offer reliable information regarding which aspects of an educational intervention (i.e. a curriculum) have been particularly effective in yielding performance improvement [47].

The optimal choice must be a measure that, on the one hand, is sufficiently close in time and context to be sensitive to intervention effects and to permit causal inferences, but is sufficiently relevant, in some absolute sense, to be viewed as a valid and important outcome. This latter point requires careful consideration, consultation, and pilot testing to ensure that one is looking at outcomes that are likely to represent the changes that could conceivably be occurring. Again, one should use both theory and experience to interpret whether the intervention is likely to impact upon measures of affect, behavioural outcomes, or cognitive indicators of ability (the ABCs of outcomes). Finally, note for now that the measures must be psychometrically sound, with proven evidence of reliability and validity. We will have more to say on this topic in the next section.

Three or More Groups and Factorial Designs

Apart from feasibility, there is rarely any reason to restrict a study to only two groups. Certainly, when one abandons the simplistic approach of a placebo or usual care control and seeks better explanation through systematic manipulation of a number of independent variables, there is good reason to consider multiple groups. Analytical methods are straightforward: analysis of variance (ANOVA) followed by some *post-hoc* procedure. The primary disadvantage is that each additional group requires an additional sample of participants. However, an alternative design strategy, using 'factorial designs', has the remarkable property that one can address multiple hypotheses with very small penalty in sample size. An example is a study of e-learning in brain anatomy by Levinson et al. [48]. There were two variables: key views (front, back, top, section) versus multiple views (where the visualised brain could assume multiple positions) and active versus passive control over the presentation (one group had control over the time the computerised image was presented in each of the various orientations, the other did not). This made four groups:

- active–key
- active–multiple
- passive–key
- passive–multiple.

These four groups can be thought of as lying within a 2×2 table (Figure 28.2).

Now, suppose that 25 students are in each group. The data would be analysed with a two-way ANOVA, giving a significance test for active versus passive based on two groups of 50, key versus multiple based on two groups of 50, and the interaction between them (i.e. whether or not

	Key views	Multiple views
Active control	(a)	(b)
Passive control	(c)	(d)

Figure 28.2 Example of a two-group comparison: e-learning in brain anatomy [44].

the effect of active versus passive control is the same for both levels of the views variable), based on four groups of 25. In terms of the two main hypotheses, the comparison is virtually as powerful as the test using only two groups with 50 subjects per group.

Further, one cannot assess an interaction without including both variables in a single research study, and the interaction is often where the most interesting findings lie. Such was the case in the Levinson study in that the best group was the passive–key view participants, the worst was the passive–multiple view participants, and the two active groups were intermediate with respect to their performance at test.

This is the simplest of factorial designs. In our view, these designs, with their capability to examine multiple hypotheses simultaneously and to explore interactions among variables at almost no cost in sample size, are underutilised in education research.

Sample and Effect Sizes

Given that the previous section argued for the value of multifactorial designs, in part on the basis that one can get more information with very small penalty in terms of sample size, it would be remiss of us not to include an answer to the ubiquitous question of 'How many people do I need?' Of course, questions are usually ubiquitous when the answer is 'it depends', but we will provide an indication of what it depends on by highlighting issues that should be considered when engaging with experimental research paradigms.

There are two central issues to take into account in determining the required sample size. The first is statistical and is related to the concept of 'power' alluded to in the last section. In educational research, there is always variability related to differences among learners. Some students will learn more than others. Regardless of whether one is learning via large-group lecture or small-group tutorial (to name one example of a comparison that has been made ad nauseum) there will be variability in the amount learnt, and the distribution of one group is likely to overlap (often substantially) with the distribution of the other group. As such, statistics are required to determine the probability that the differences observed in the mean scores achieved by both groups are unlikely to have arisen simply by chance. The standard of '$p = 0.05$' means that the likelihood that an observed difference (or association, as we will discuss later) arose due to chance is less than 5%, so the odds are good that the intervention, and not random variation, resulted in the differences between the groups. Because it is based on probabilities, the conclusion is fallible – one can falsely conclude that there is a difference when there is none.

'Power' reflects the opposite concern; concluding that no difference exists when in fact there is an underlying effect of the intervention. It is the probability that a study has a large enough sample to detect an educationally important effect. Specific calculations for sample size are beyond the scope of this chapter as the formulae differ depending on the statistical test one needs to perform, but it should be noted that in all cases sample size calculations are dependent on predictions of how large a difference and how much variability in the sample one would expect. The predictions should be based on the best information available, but inevitably the calculations will be guesstimates to some degree. If one is able to find a statistically significant difference, however, then by definition the study was sufficiently powered (i.e. had a large enough sample size). One could debate the representativeness of the sample and whether or not the effect might disappear with greater sampling, making replication an invaluable strategy for confirming the accuracy of the results, but whether or not the study had sufficient power (i.e. was statistically 'big enough') is not debatable. Power calculations are only relevant before a study has ended or if significance was not obtained, but, like sample size calculations, these require an assumption of the magnitude of difference one wanted to detect. Big differences require smaller samples than small differences.

The second issue to consider is implicit in the preceding discussion. Very large samples can yield the opposite problem to small samples. An intervention can have a statistically significant effect even if it is of no practical importance, simply because the study had a very large sample size. For this reason, it is wise to consider not only the statistical significance, but also the size of the effect, typically defined as the difference between group means divided by the standard deviation (see any statistical text for more detail). The larger the effect size, the easier it is to argue that the findings are 'clinically important'. By convention, effect sizes of 0.2, 0.5, and 0.8 (i.e. differences that amount to 20, 50, and 80% of the standard deviation) are considered small, medium, and large [49].

Summary

For many, experimental research is held up as the gold standard for which we should strive when addressing all research questions. Clearly this is simplistic, but it is true that experimental designs, when applied appropriately, can have great impact on our ability to understand the extent to which causal relationships exist between variables (i.e. if one variable changes, does it cause another to change?). But such inferences rarely arise from curriculum-level interventions, which contain many variables and covariates. It is often more informative to design a series of small-scale experiments that tease apart the active ingredients, thus building knowledge of which elements are critical for learning. However, when one is interested in testing the relationship between naturally occurring variables that cannot easily be manipulated, then methodologies drawn from the epidemiological or correlational traditions may be more appropriate, as discussed in the following sections.

The Epidemiological Tradition

The term 'RCT' has been used repeatedly in this chapter, referring to what is commonly viewed as the optimal experiment, in which subjects are randomly assigned to different groups that receive different treatments, and are then compared on some outcome measure. The RCT has a position of honour at the top of a hierarchy of research designs derived from epidemiology. While other designs in this hierarchy are rarely explicitly used in educational research, they may occasionally have a very useful role, as we shall point out.

Because many epidemiological investigations are based on a dichotomous outcome (dead or alive, improved or worsened, diseased or disease-free) as well as a dichotomous classification at inception (drug/placebo, risk factor present/absent, e.g. smoker/non-smoker), the easiest way to think about these designs is as a 2 × 2 table.

We have already discussed the RCT, the design of which is shown in Box 28.4. Respondents are randomised to rows (i.e. to the drug or placebo group) and the outcome (the columns) is tabulated. A *cohort* study looks the same, except that participants are not randomised to the two groups; rather, they are members of each cohort as a result of processes beyond the experimenter's control and, as such, the word 'intervention' should be replaced by 'exposure' or some other descriptor appropriate to the particular focus of the study.

BOX 28.4 Randomised controlled trial study design

A *case–control* study can be illustrated in the same way, but the method of allocation runs in the opposite direction. Cases are selected by the outcome – they had the disease, or they failed the examination – and controls are selected by who did not have that outcome. The study then looks back to determine whether the cases were more likely to be exposed to some risk (e.g. smoking or PBL). A case–control study, therefore, looks the same as the RCT with the exceptions that:

- the rows are better labelled as risk factors (present/absent)
- subjects are assigned to the columns rather than by the rows as the researcher looks for different rates of risk factor across the columns.

	Outcome	
	Present	*Absent*
Intervention		
Control		

Many studies of PBL versus lecture-based curricula can be classified as cohort studies, because students are in one cohort or the other through non-experimental factors such as self-selection or differential admissions policies. They may then be followed forward to determine, for example, success rates on licensing examinations or acceptance into primary care residency programmes. One example is the study by Woodward et al. [26] of billing patterns of McMaster graduates matched to graduates from the rest of Ontario.

The important study by Papadakis et al. [14] of predictors of disciplinary action by a medical board is a good example of a case–control study. They identified a group of cases: 235 graduates from three medical schools who had some kind of disciplinary action as physicians and matched them to 469 controls, who were comparable demographically (i.e. were matched), but who had no record of disciplinary action. They then looked back at both groups' undergraduate records to determine whether they had episodes of unprofessional behaviour in medical school, and found that 92/235 cases (39%) revealed problems as a student, whereas only 90/469 controls (19.2%) revealed similar issues. One must consider base rates to determine the consequences of acting on such differences [50], but these compelling findings draw appropriate attention to the types of behaviour that should be considered in the context of professionalism.

Summary

Perhaps ironically, the most applicable study design from this hierarchy is the last. There are many examples of studies that resemble cohort studies or RCTs, and it is not particularly helpful to single these out with a new label. However, the case–control study is uniquely useful in situations analogous to its application in clinical medicine, where the outcome is categorical (disciplinary action: yes/no), the prevalence of the outcome is low, and the time delay until it occurs is long. Because it is usually retrospective, the case–control study may well be the only, or at least the most efficient, response to the concern to link educational interventions to patient outcomes; any alternative is likely to be too large, too costly, and too inefficient to show any yield (Box 28.5) [41].

The Psychometric Tradition

As alluded to in the last section, the RCT is frequently held up as the 'best' research design with accompanying consternation regarding how few studies in medical education use an RCT. Such an attitude is, in our view, amazingly myopic [17]. One need not leave the quantitative domain to recognise that many of the most important questions and issues in the field cannot, *and should not*, be addressed with an RCT or, for that matter, any other experimental or quasi-experimental design.

Arguably, the most significant advances in our field have been in assessment, where medical education has, quite literally, led the world [51] in developing authentic, yet psychometrically defensible, measures like the Objective

BOX 28.5 FOCUS ON: Quasi-experimental designs

Quasi-experimental designs, including cohort and case–control studies, have the following characteristics:

- they usually have two groups (occasionally more)
- they enrol participants who are assigned to each group based on some predetermined characteristic on which it is not possible to randomise (e.g. presence of disease or risk factor)
- they are conducted retrospectively (usually)
- participants are not blinded to the group they are in
- complete follow-up of participants is rarely achieved.

Structured Clinical Examination (OSCE) [52]. Assessment methods are analogous to diagnostic tests, designed to identify those who have a lot or a little of the characteristic of interest, and perhaps to eventually create a cut-off to label a disease, in this case, 'incompetence'. Just as the starting point in developing a diagnostic test (e.g. a new radiographic procedure) is to assemble a group of patients, administer the test, and look at (i) inter-rater agreement and (ii) relation to some other measure of the same characteristic, we would set about testing a new assessment method by assembling a group of students, administering the test, and examining reliability and then validity. Any radiologist who introduced a new diagnostic procedure by conducting a multi-centre trial to see whether patients who had the test lived longer before they first proved that radiologists had adequate inter-rater agreement and that the test results converged on other measures of disease would be at risk of psychiatric referral. Similarly, while we might like to eventually use an experimental method to show that students who get a new assessment method eventually perform better than those who are assessed using some conventional method, this is hardly the first step.

Because so much research in medical education is directed towards the development and testing of assessment methods, this section is devoted to a discussion of basic issues in psychometric methods. The discussion is necessarily brief; for a substantial elaboration, we refer you to Streiner and Norman [53].

Basic Concepts

Psychometric methods are designed to ensure that data are sufficiently trustworthy to enable appropriate interpretations and accurate decision-making. Medical educators and the lay public are bombarded with 'data-based' claims every day, but it goes without saying that not all such claims should be treated equally. Most threats to validity require us to formalise this intuitive scepticism and devise ways to test whether or not our measurement instruments are indeed measuring what we intend them to be measuring.

These strategies can be applied to any number of domains, including but not limited to assessments of the quality of admissions decisions [54], research into the relation between personality and professionalism [55], or use of a questionnaire to measure the extent to which learners demonstrate self-directed learning [56]. In each case it is easy to collect information and make decisions, but much more is involved in determining whether or not those conclusions stand up to proper scrutiny. To understand this statement it is of course necessary to define 'proper scrutiny'. For any measurement instrument to provide useful information, be it an objective indication of some physical state (as would be measured with thermometers) or a subjective claim about a more ethereal construct (like one's *perceptions* of one's own abilities), it is necessary to ensure that the tool satisfies the four '*-ities*' of good measurement:

- feasibility
- acceptability
- reliability
- validity.

The first two need no explanation in that it seems fairly straightforward to suggest that a tool should only be used to the extent that it can be used (feasibility) and to the extent that people will use it (acceptability). How to assess these '-ities' requires some thought; part of acceptability, for example, entails demonstrating that the measure does not show undue bias against particular subgroups of the population. However, it is the latter two -ities that likely require elaboration. Before we begin, a disclaimer: while we will discuss *the* psychometric properties of measurement instruments, we do so simply for the sake of useful shorthand, as it is inaccurate to make acontextual claims about such properties in relation to any instrument. That is, the utility (see Chapter 20) of an instrument is based entirely on the population and context within which the instrument is to be used. While a grading system based on sad and smiling faces might be appropriate for school-age children, it is unlikely to be accepted in the context of national licensing exams in medicine; a survey that asks people to answer questions about their sexual activity levels may elicit accurate responses in some respondent cohorts, but may not be answered (and may cause offence) in others; a clinical skills examination that requires blood pressure to be taken may discriminate among beginner medical students but may be useless with medical residents. As we will see, these contextual variables can have significant impact on the assessment of reliability and validity.

Reliability

Reliability may be the most misused word in all of medical education. It does not mean agreement (although agreement is relevant), it does not mean variability (although variability is relevant), and it is not indicated by consistency of mean scores calculated for a group of individuals (although one would expect such consistency if the tool is reliable). Reliability is a statistical term indicating the extent to which a measurement instrument *consistently differentiates* between *individual* subjects of interest. The subjects may be learners, teachers, courses, schools, survey respondents, or any other group of individual entities. As attempts to differentiate between students are most common within our community, we will use that domain as an example.

Were one interested in developing a tool to assess knowledge of professional responsibilities, it would not be hard to generate items that result in variability of responses. That, in fact, would be a predominant goal as, presumably, if everyone were to provide the same responses there would be little reason to administer the test. That variability, however, could be attributable to any number of factors. Our hope is that scores on the test reflect true or consistent differences between students, with respect to their knowledge of professional responsibilities. Some portion of the variability, however, will be attributable to error of measurement because systematic biases and random forces can be expected to impact on the scores assigned to students.

While there are an infinite number of sources of error, the primary question is how much of the variability in scores can be attributed to error in relation to actual differences between the students? In other words, if we were to re-administer the test (or have different examiners rate the responses or use a parallel test), how consistent would individuals' scores be from one administration to the next? Mathematically, the simplest way to represent this concept is with the following equation:

$$\text{Reliability} = \frac{\sigma^2_{\text{subjects}}}{\sigma^2_{\text{subjects}} + \dfrac{\sigma^2_{\text{error}}}{n}}$$

σ^2 is the conventional symbol used to express variance, so the numerator (the top line of the fractions) indicates the amount of variance attributable to differences between the students themselves and the denominator (the bottom line of the fractions) represents the total variability observed in the scores. This equation is not presented to scare away the innumerate, and we will not elaborate here on how one would calculate reliability. Rather, the formula is presented because it enables us to illustrate some fundamental points about reliability and, in turn, research in the psychometric tradition (see Box 28.6).

BOX 28.6 FOCUS ON: Reliability

- *Reliability is not a fixed property of the measurement instrument*. If a test of professionalism knowledge, as described earlier in this chapter, is designed to provide an assessment of second-year residents, then its reliability (i.e. its ability to consistently differentiate between subjects) must be tested on a sample of second-year residents. To recruit a more heterogeneous sample (e.g. by enrolling first-year undergraduates and practising ethicists) will result in artificial inflation of the numerator, and as a result, artificially inflated estimates of the reliability of the tool. Researchers must make a concerted effort to specify the context within which they want to use their instruments and test by recruiting a sample of respondents representative of those working in that context.
- *Repeated measurement across the variables of interest is required to estimate the reliability of a tool*. If raters are liable to disagree about the strength of a student's performance, then multiple raters should be asked to rate the student's performance. If performance varies across the cases (content specificity), then students should be assessed on multiple cases. Simply administering a test and revealing that the scores are normally distributed tells us absolutely nothing about the extent to which the tool consistently differentiates between subjects because the variation can result from true differences between students or measurement error.
- *The more observations per individual one is able to average across, the more reliable the instrument will tend to be*. The *n* under the error term represents the number of observations collected (be they from multiple test questions, multiple raters, multiple administrations of the exercise or some other source of error variance). An average over multiple observations provides a better estimate of the amount of the construct held by the individual than any one score because random positive sources can cancel out random negative forces. Of course if a particular source of variance does not contribute error to a particular measurement, averaging across multiple observations collected across that source will have no impact (dividing zero by anything still leaves one with zero). An important aspect of psychometric analysis, therefore, is to determine how many observations one must collect for the total to achieve reasonable levels of reliability – if the answer is too many to be feasible, it suggests that the tool should be modified or abandoned.
- *A tool that does not discriminate is useless for assessment*. There may be other aspects of utility (specifically, motivating individuals to engage in desired study behaviour – an educational impact) that warrants use of a particular measurement instrument, but generally, claims of utility rely on evidence of reliability and, if everyone receives the same score, from an assessment perspective one may as well assume the result and do something better with one's time than administer the test.
- *Claims that the mean score of a group did not change over time (or across raters) provide no evidence of reliability*. One would find a stable mean even if the rank ordering of individuals within the sample perfectly reversed from one test administration to the next (i.e. if there was absolutely no consistency in the scores assigned and, as a result, all variance could be attributed to error). A random number generator can be expected to result in equivalent means on different occasions, but random number generators can hardly claim to provide reliable measures of performance.
- *Occasionally, the claim is made that reliability of a measure is irrelevant because validity is more important*. Such a statement is simply illogical. One can view reliability as the correlation between a measure and itself (on repeated occasions). One aspect of validity expresses the correlation between a measure and some external (preferably 'gold') standard. It is axiomatic that a measure cannot correlate with something better than it correlates with itself. Hence, reliability is not dissociated from validity; instead, it sets an upper limit on possible claims to validity. And, in fact, modern models of psychometrics consider reliability to simply be one aspect of validity [57].

Within the psychometric tradition, readers of educational journals will also encounter the notion of *generalisability*, a close cousin of the concept of reliability. Generalisability theory is a way of expressing the extent to which the scores assigned to individual subjects generalise to the scores assigned in another context (with another rater, at another time, etc.) [52]. If that sounds familiar, it is because generalisability theory is simply an extension of classic test reliability theory that provides the mathematical infrastructure to enable multiple sources of error variance to be considered simultaneously. The fundamental advantages are that one need not complete multiple studies to assess the relative error contributions of multiple variables and that, as a result, one can determine the relative benefits of increasing the number of observations collected across one variable relative to the benefits that can be gained by increasing the number of observations collected across another.

Validity

Historically, most descriptions of validity have used one taxonomy or another to differentiate various ways in which one can consider the trustworthiness of a set of ratings [58]. *Content validity* is considered to be the extent to which the items in a tool adequately sample the domain of interest without extending beyond it (i.e. are the questions sufficient and relevant?). *Criterion validity* refers to the extent to which the measure correlates well with another measure of the same underlying construct. *Construct validity* indicates the extent to which the scores derived from the instrument align with expectations based on understanding of the underlying construct that the tool was intended to measure (e.g. a new measure of height should result in higher scores for basketball players relative to jockeys). Other taxonomies have been used, but in our minds it is all just validity (i.e. an indication of whether or not the scores derived from the use of the instrument vary in conjunction with the extent to which the amount of construct in the individual being measured varies). In fact, some have argued that reliability is simply one aspect of validity rather than a separate concept, the argument being that if the amount of underlying construct is not expected to have changed across administrations of the instrument, then the scores should not change either [59]. The various taxonomies that have been published may be useful to generate ideas as to how the validity of an instrument can be tested, but one should not allow the taxonomy to distract from awareness that proper validity testing requires systematic study, preferably with a variety of methodologies.

That said, one aspect of what Messick calls '*consequential validity*' is worth highlighting [58]. Any time we put in place an instrument to assess students, we must worry about the extent to which the measurement instrument has an impact on behaviour. Assessment has long been known to have a steering effect on the learning activities of students [60]. As a result, to ensure the utility of an assessment instrument it is necessary to engender a match between the learning activities one hopes to promote and the learning activities stimulated by the tool [60]. As these five principles of good measurement (the four '-*ities*' and educational impact) that form a tool's utility do not always align (and, in fact, often run counter to one another), it is almost inevitably necessary to decide on an appropriate compromise, the balance of which should be determined by the specifics of the situation.

Most will start their study of the validity of an instrument by testing its reliability as described above, for the simple reason that if a tool is not reliable, it cannot be valid. For example, in studying the consistency of ratings provided to medical school applicants during panel-based interviews, Harasym et al. [61] noted that over 50% of the variance in scores could be attributed to the person doing the interview, thereby fundamentally calling into question both the reliability and the validity of the panel-based interview process, as that process is intended to provide information on the quality of applicants, not the stringency of the interviewers.

However, reliability is insufficient. Just because something can be measured consistently does not mean that the measurements are valid. It is easy to measure the circumference of an individual's head in a consistent and reliable manner. Those data, however, are completely useless if one is trying to assess the empathy levels of the subjects given that phrenology was discredited a century ago [62]. A more direct example comes from the literature on OSCE testing formats. As most readers of this chapter will be familiar with the OSCE, we will note simply that it is a 'bell ringer' type of examination in which examinees interact with multiple patients in sequence while striving to demonstrate their clinical skills. The 'O' in OSCE stands for 'objective', to indicate the initial idea that one could evaluate performance by generating a checklist of appropriate behaviours and noting which were undertaken by the examinee. Indeed, such checklists have been shown to robustly provide very reliable measurements of individuals' performance [63]. In various studies, however, they have been seen to bear no relation to experience, an important variable if one wants to make claims about measuring ability. In contrast, global ratings of performance do tend to correlate with experience levels, suggesting that while checklists may provide a valid measure of comprehensiveness, subjective judgement provides a more valid measurement of clinical expertise in many domains [64, 65].

One could go on ad nauseum about the variety of methodologies that can be used to study validity. Comparing average scores received by different groups that can be anticipated to differ in amount of the construct, correlating individual scores with other continuous variables that are expected to be related to the underlying construct, and examining the change in scores that takes place after an intervention expected to change levels of the underlying construct provide three broad classes of approach that might be adopted. Interesting examples include the work of Tamblyn et al., which revealed a relationship between performance on the Canadian licensing examination and professional behaviours during practice as a physician [41]; the work of Ramsey et al. [66], which revealed that specialist certification based on multiple-choice testing is predictive of peer ratings 10 years into practice; and the work of Davis et al. [46], which has continued a line of inquiry that casts doubt on the validity of self-assessments as indicators of performance. In fact, many of the methodologies included in other sections of this chapter (as well as those not included) could be deemed strategies for testing validity.

Instead of trying to generate a comprehensive list of strategies, we will close this section by simply noting that one can rarely expect to prove validity in any absolute sense. It is important to test validity because claims that a measure provides information that should be allowed to guide decision-making rest on the balance of evidence. In essence, validity testing is theory testing; each new test that reveals a positive result supports both the theory and the validity of the instrument, but each new test that reveals a negative result should lead one to question (and study) whether the theory is incorrect or the tool provides an inadequate measure of the construct.

Summary

The overarching theme of research undertaken within the psychometric tradition is that researchers, educators, and clinicians, and the lay public for that matter, need to strive to be sure that the data that guide our thinking and decision-making are sufficiently trustworthy to warrant using them to draw conclusions. This is not simply an academic issue, as too often people's lives are altered (by admittance to/rejection from/advancement within their chosen profession, by decision-making within the legal system, or by personal life/marriage counsellors) on the basis of 'data' of dubious validity. Ensuring the validity of one's research instruments and assessment strategies is an ethical imperative [67]. It is not easy and there are certainly factors that need to be considered beyond psychometric properties, but the methods and concepts outlined here hopefully provide a good start and, if nothing else, should provide a basis for reasoning about these types of problem.

The Correlational Tradition

A significant proportion of research in medical education is derived from survey questionnaires. These may cover a potentially vast array of topics, from intrapersonal issues such as learning styles or emotional intelligence, to observer ratings of achievement or other aspects of observable behaviour, to satisfaction measures. It would be unrealistic to attempt to cover this vast, heterogeneous, and complex field; instead this section will be devoted to a number of common issues related to scoring, research design, and analysis. Questionnaire 'design' is addressed by Lovato and Peterson in Chapter 30 of this book, but we would remind the reader that proper questionnaire-based research is not easy, and reliability and validity must still be ensured. In general, questionnaires are useful for systematically determining a large group of individuals' perceptions and attitudes towards a particular issue they have experienced. Beyond that, however, one must always be aware of the limitation that people are notoriously bad at accurately judging the cause of their behaviour [68] or the adequacy of their own performance [45, 46].

Scoring

Quite commonly, responses from individual items are to be summed into a score. Much effort is sometimes expended to decide what weight should be given to each item comprising the score. As it turns out, an extensive literature dating back to at least 1976 is absolutely consistent – an equal weighting model, where all items are simply summed together, is as reliable and valid as *any* alternative [69]. There is one cautionary note: a simple sum assumes that the items are similar in means and standard deviations. It would be no more appropriate to add together items, some of which are binary (0 or 1) and some of which are on seven-point scales, than it would be to add an interview score based on seven-point scales to grades out of 100, or for that matter, to add weight in kilograms to height in metres as a measure of overall size. When the individual items are on different scales, the correct approach is to convert to Z scores

$$Z\,\text{score} = \frac{\text{score} - \text{mean}}{\text{standard deviation}}$$

before combining scales, but the combination should still retain equal weights.

One other point about scoring is that there is extensive debate about whether the scores assigned to the sorts of scale described here should be summed to calculate a mean given that, technically, the data are ordinal in nature (i.e. they indicate rank order without any guarantee of there being equal intervals between all pairs of sequential points) [70]. In practice, the parametric statistical tests that require interval-level data tend to be quite robust to deviations from normality [71–73] and their ease of application provides great advantage in most situations.

Validation

The validation methods described in the section on psychometric methods are appropriate for survey instruments and should be considered carefully to avoid making decisions based on survey data that might not be trustworthy for the purpose for which they were intended.

Analysis

The correlational approach is based on a search for relationships among variables, and analysis typically begins (and all too often ends) with every variable being correlated with every other, and then *post-hoc* stories being constructed around the few 'significant' correlations. The problem with the strategy is researchers appear to forget the meaning of '$p = 0.05$' – the likelihood that an observed relationship of this magnitude could have arisen by chance *if there was in fact no relationship*. In other words, for every 20 correlations that are calculated, one will be significant by chance at the 0.05 level (actually, there is a 64.2% chance that at least one will be significant). It is worth noting here that this applies to any statistical analyses (see Box 28.7), including ANOVAs, *t*-tests, and other strategies that rely on *p*-values to determine whether or not the data can be accounted for by chance alone. An obvious solution to this 'data dredging' is to begin with a substantive theory about what

BOX 28.7 FOCUS ON: Common statistical tests

Statistical tests are based on two broad classes.

Parametric tests

Applied to data on which it makes sense to calculate means and standard deviations.

Tests used for comparing means:

- t-*test:* one independent variable with two groups, or two related observations (such as before–after).
- *ANOVA*: one or more independent variables, each containing two or more groups ('levels').
- *Repeated measures ANOVA*: a special case of ANOVA in which repeated observations are taken within an independent variable on the same subjects. Also used for reliability and generalisability studies.

Tests used for examining relationships:

- *Pearson's correlation:* provides relation between two measured continuous variables.
- *Multiple regression:* provides relation between multiple predictor variables and a single *continuous* dependent variable.
- *Factor analysis:* provides relations (underlying factors) for a large number of related variables.

Non-parametric tests

Tests used for frequency counts:

- *Chi-squared:* compares proportions in two or more related categories (e.g. 2×2 tables).
- *Logistic regression:* provides relation between multiple predictor variables and a single *dichotomous* independent variable.

relationships are expected. At a minimum, this can direct attention to specific correlations rather than using a 'shotgun' approach. Further, as it is likely that the researcher will still be interested in more than one correlation, the critical p-value should be set at $0.05/n$ where 'n' is the total number of statistical tests – a 'Bonferroni correction' [74].

A more sophisticated approach than correlations involves multivariate methods such as multiple regression, factor analysis, and structural equation modelling. Strictly speaking, the term 'multivariate' should only apply to a situation with multiple *dependent* variables. So multiple regression is a univariate procedure, factor analysis and structural equation modelling are multivariate. Multiple regression involves predicting a single dependent variable with multiple independent variables, e.g. predicting licensing exam performance with a combination of variables like undergraduate grades, gender, and Medical College Admission Test (MCAT) score. Factor analysis seeks underlying associations among clusters of variables, which are called 'factors'. More sophisticated is the family of multivariate methods, including confirmatory factor analysis, hierarchical linear models, and structural equation models. In all these methods, the researcher begins with a theory about the relationship among variables (e.g. *good tutors* succeed by increasing *motivation* of students, and this, in combination with their *prior achievement*, predicts their *final exam performance*). Different causal models are fitted to the data set and the degree of fit computed. Challenges for these approaches include: (i) all of these methods are sample intensive, and the rule of thumb is that the sample size should be at least 5–10 times the number of variables; (ii) as the complexity of the model increases, it becomes less and less clear what it actually means to say that one model fits the data but that another model does not; and (iii) because it is unlikely that any two studies will use the same combination of variables, the concern remains that the causal theory, whatever it may be, is unique to the data set on which it is based.

Nevertheless, these approaches do represent a considerable advance over the mindless cranking out of dozens of correlation coefficients that is all too frequently the norm in correlational research. As in the discussion on effect size presented in the experimental studies section, we urge a focus on the correlation coefficient, not the associated p-value. With large samples, even small correlations (e.g. $r = 0.1$) can be statistically significant. The coefficient of variation (r^2), however, reveals that $r = 0.1$ describes a relationship that accounts for only 1% of the variance in the data. As a result, r^2 should always be used to judge the 'clinical' importance of a correlation.

Cronbach's 'Two Disciplines'

At the outset of this chapter, we pointed out that many research questions cannot *and should not* be answered with experimental designs. The methodology that is most appropriate, whether choosing between quantitative and qualitative designs or between experimental and correlational methods, is dependent on the question the researchers want to address. In promoting theory-based and programmatic research efforts, we advocate using a variety of methods to enable triangulation on a problem, thereby developing a richer understanding of the underlying relationships than any one methodology would allow. However, the choice is not quite as value-free as might be imagined. Lee Cronbach first recognised a fundamental duality in a classic paper published in 1957 [75] called 'The two disciplines of scientific psychology'.

The essence of the dichotomy is as follows. Correlational methods, including psychometrics, are critically dependent on individual differences. It begins with the reliability coefficient, which is zero if everyone is alike (i.e. if there is no subject variance). If we want to examine the relation between some individual attribute, such as intelligence quotient or premedical grades, and some outcome, such as licensing examination performance, unless some students are high or low on each measure there can be no correlation. By contrast, if we were to do an experiment to see whether a supplementary course can help students achieve higher scores on a standardised admissions test, ideally we would like to begin with a cohort of students whose abilities, as measured by undergraduate grades, are exactly the same. To the extent that some students are already very good at biology, physics, etc. and others in the course have little knowledge or aptitude, this will lead to large variability in the scores of students in the experimental and control groups. This variability will, in turn, end up in the denominator of any statistical test designed to show that the treatment was statistically significant (i.e. will add 'noise' to the data).

To the experimentalist, any variation between people will dilute the chances of finding a treatment effect. To the correlationalist, the goal is explicitly to understand the differences between people. Thus it is literally the case that one person's signal is the other's noise.

Given this dichotomy, it makes no more sense to argue which is 'better' methodologically than to try to find evidence that red is better than blue, irrespective of the use to which the colours are to be put. They are not better or worse, except in relation to what one is trying to achieve; they are just different. The situation was nicely summarised recently by a wag who declared: 'Randomised controlled trials are the best design of all to find out if a treatment works, and the worst design to find out who it works for.'

Reviews

As mentioned earlier, any primary research study will have flaws; even if the perfect study could be designed, there is no way to completely control for the powerful forces of random variation. As a result, it is important to consider the balance of evidence available when deciding how to use the information that has accumulated in the literature. This is the key reason that scholarly reviews of a field are so valuable – when done well, they synthesise the available evidence in a way that can refine readers' understanding of the focal problem and help them better understand the implications of the literature for their own practices or their own knowledge-building research efforts. We have already noted that every research effort should be informed by a review of existing literature. Here we provide some insights into the creation and interpretation of more formal efforts, starting with systematic reviews, as they represent a form of review on which emphasis has been growing in recent years. Chapter 31 provides a more detailed description and discussion of various approaches to knowledge synthesis.

Systematic Reviews and Meta-analysis

In part as a consequence of the Best Evidence Medical Education (BEME) movement, initiated by Harden et al. [76], systematic reviews have become increasingly popular in the medical education literature. To some degree they epitomise the reductionistic approach to quantitative research – the goal is to determine the one number that best specifies how well 'it' works. Although many might assume that medical education has adopted the technology of systematic reviews from clinical research, where systematic reviews of, for example, the risk reduction for stroke from beta-blocker therapy are commonplace, in fact the path is more tortuous than that. The first proponent of systematic reviews and the accompanying meta-analyses were educational statisticians, Smith and Glass, whose 1977 article in *American Psychologist* [77] is usually cited as the first publication of the type.

What is a systematic review? The goal is, more or less, to identify *all* of the empirical literature on a particular question, and to then use statistical methods to best estimate the effect (or non-effect) of a particular intervention in a way that is relatively free of bias compared with less comprehensive strategies. There are, therefore, three aspects of 'systematic':

- a systematic search for all the literature relevant to a topic
- a systematic review to select the subset of articles achieving at least minimal quality and relevance
- a systematic summary using specific statistical methods to arrive at the best estimate of the effect in question.

It is clear that these aspects, while equally important, are separable. First, careful computer algorithms to search electronic databases must be devised and then supplemented with manual searches. Once the key articles have been located they must be reviewed in detail to ensure methodological rigour, often using a detailed reporting form that enables a quality score for each study. Finally, each study is typically analysed to estimate an 'effect size', indicating the strength of the intervention in each instance in which it was used.

In meta-analyses these effect sizes are then combined, using a weighting by sample size, to arrive at an overall (i.e. average) effect size and a statistical test of significance. It is at this point that the second 'systematic' emerges – a systematic statistical averaging of all the individual effects into an overall unbiased estimate. This is the whole point of the exercise: to determine whether a particular intervention affected a particular outcome. One example might be the effect of PBL on national licensing examinations [78]. Another is the recent BEME review of the predictive validity of undergraduate assessment instruments predicting licensing examination performance [79].

Problems with Systematic Reviews

While wonderful in theory, there are at least three problems with trying to put these methods into practice.

Quality of the Evidence

It seems that an inevitable consequence of the systematic review is a note of despair about the poor quality of the studies, based on the number of criteria that were not fulfilled. It has almost reached the point of there being a standard disclaimer in systematic reviews: 'The authors take no responsibility for any personal damage resulting from the quality of the studies that went into this review.'

This disparaging of the quality of published papers seems a bit strange because, for the most part, the articles had all satisfied peer reviewers. Either the peer reviewers are not very good at their jobs, or they are basing their judgements on different criteria altogether than the systematic reviewers. We sense the latter; as editors, we rarely judge the worth of a paper by the number of methodological criteria it fulfilled [80, 81] and, indeed, Bordage [82] has shown the same to be the case on the part of peer reviewers. Just as the OSCE literature has found that global judgements are superior to checklists, Bordage's examination of the peer-review process suggests that the methodological components of a paper provide a poor indication of its overall value. Further, a preoccupation with quality might be tempered by the finding of Lipsey and Wilson [24], mentioned earlier, that there was no relation between judged study quality and treatment effect, nor did randomised trials yield systematically different treatment effects than non-randomised studies.

Heterogeneity of the Outcome

While the use of effect sizes enables putting various measures *of the same construct* on a common metric, it appears that systematic reviews are rarely able to reduce the outcome to one measure such as examination performance. In fact, in the recent review of learning portfolios [83] there was commonality among studies only at the broad level of classification (e.g. learning versus assessment) and no attempt was made to try to average outcomes across studies. Instead, the review reported findings along the lines of 'two studies reported that portfolios contributed to reflective learning'.

Undoubtedly, the most careful and comprehensive approach to systematic reviews in medical education has been mounted by the BEME group. Very careful quality control is exercised at every step of the process, and the group of collaborators goes to enormous lengths to ensure consistent quality. The first review, of high-fidelity simulation, by Issenberg et al. [84] began with over 600 abstracts that were then reduced to 109 studies for detailed analysis. The review concluded: 'Heterogeneity of research designs, educational interventions, outcome measures, and time frame precluded data synthesis using meta-analysis.' The authors then went on to describe the conditions that led to effective use of simulation. This approach, where a systematic search on many abstracts yields a small number of suitable studies, which in turn are combined with too many potential outcomes to permit any quantitative synthesis or meta-analysis, emerges as the norm for these reviews.

Low Yield of Studies

Systematic reviews can be enormously labour intensive, primarily because the yield of useful articles is so low. Examples drawn from the BEME monographs can be given. The study of early community experience [85] catalogued 23 outcomes from 73 studies (out of 6832 abstracts) and no quantitative synthesis was possible. Perhaps the worst example of 'needle in a haystack' was the synthesis of interprofessional education [86] that began with 10 495 abstracts and ended with 12 worthy of detailed review. Interestingly, a hand search added a further nine, despite beginning with only 46. Again, the results consisted of counts of what kind of intervention led to what kind of result. To avoid some of these problems, while admittedly creating others, many scholars choose to engage in generating critical rather than systematic reviews. In the following section we will strive to compare the two strategies, highlighting strengths and weaknesses of both approaches.

Critical Reviews

In our initial discussion of the research question, we described the characteristics of a good literature review, namely that it represents a critical synthesis of a literature, identifying what is well established, what is only poorly understood, and what remains to be understood. It may, when done well, bring together several disparate literatures and, as a result, offer a new perspective. It should *not* end up as a chronological 'blow-by-blow' account, with one paragraph per study and no real synthesis. There is no pretence in a critical review that the cited literature represents all the relevant literature in the field, so there may be less of a tendency to provide a one-paragraph summary of every related study; the author is bound by an unwritten moral code to represent the various perspectives fairly, but that is all.

To our knowledge little is written about how to go about such reviews, which is somewhat strange, as there is little doubt that some of these papers become, over time, the 'citation classics' of the field. Far and away the most cited papers in the area of PBL are three old chestnuts: Albanese and Mitchell [87], Vernon and Blake [78], and Norman and Schmidt [88]; two of the three are critical reviews. What distinguishes a good critical review from a poor one? One expects that it has little to do with comprehensiveness or systematicity. Instead, the cited reviews appear to be those that present unique perspectives and marshal evidence convincingly to support the claims. Rather than scouring the nooks and crannies of the literature for every paper that is relevant to a narrow question, successful critical reviewers explore a variety of literature, mining for gold nuggets that often alter the way the community fundamentally defines the question.

In practice, critical and systematic reviews in education have often led to similar conclusions, in part because, while no one would debate that the goal of systematicity – to eliminate bias in the data one draws on – is laudable, to some extent the mantle of systematicity is just a guise of credibility. If one cannot combine the findings in some systematic way as a result of heterogeneity of outcomes to the point of having to describe each study independently, then the only thing separating systematic reviews from critical narrative reviews is the amount of time and resources spent searching for information. Given the typical low yield of studies, it is questionable whether or not that effort proves a worthy use of resources or serves as a key arbiter of quality. Again, this is not meant to imply that systematicity is bad or that all systematic reviews have been thoughtless – the examples used in this chapter make it clear that is not the case. It is to say though, that too often the claim of systematicity is applied thoughtlessly as a criterion by which quality is judged, when in fact true advances in the field are as often gained from critical syntheses of diverse ideas rather than systematicity itself [89]. That said, we in no way mean to imply that systematic reviews are not useful when done well, as there are many exemplary examples in the literature [90, 91].

Problems with Critical Reviews

Despite the arguments expounded above, critical reviews too are not without their own problems.

Author Bias

The strengths of critical reviews are also their weaknesses. When literature is marshalled to support a unique perspective, there is the vague disquiet that the selected literature may be, consciously or unconsciously, biased in favour of the claim. The author is under no explicit mandate to present all the evidence for and against, only to be unbiased in his or her conclusions. But such a stricture may not be realistic; if we were aware of our biases, we may well not be biased, so it is not uncommon for two critical reviews to come to diametrically opposite conclusions. Such is the fodder for academic debate.

Biased Sampling of Literature

A second problem is that if the purpose of the review is really to obtain a best estimate of the value of something like the predictive value of standardised aptitude tests or the effect of PBL on outcomes such as standardised examinations, or the effectiveness of faculty development programmes in changing faculty teaching competence, the synthesis methods used in critical reviews, if used at all, are primitive at best. They often reduce to a summary like '22/30 studies showed a positive effect'. That is precisely what systematic reviews do best – sometimes.

Finally, critical reviews can also assume a mantle of academic dithering. Such reviews rarely conclude with a final 'it works/it does not work', instead providing far more nuanced discussion than purely systematic reviews, with an inevitable self-fulfilling call for more research. Again, however, we would emphasise that where one might call this a weakness we consider it a strength, as this sort of academic dithering can enable much more refined appreciation for issues than was available in the field before the review was generated.

Summary

In the examples cited, the distinction between systematic review and critical review becomes vanishingly small. While each type of review may be stimulated by differing goals ('Does it work?' versus 'How does it work?'), inevitably, as the systematic review identifies subgroups and subgoals, the additional knowledge is more of the form of revealing how different circumstances may influence the results. And while the critical review may be directed at advancing a theory, the reality is that there are very few theories in this field, so it is more likely that it will be focusing on the various things that may influence the effect under review. So it would seem that the ecology of the domain may be forcing a convergence between the two approaches.

Discussion

For about three decades, educational research has been embroiled in the 'qualitative–quantitative' debate, to the detriment of both. A careful read of this chapter and Chapter 29 on qualitative methods, reveals that there is probably as much divergence in goals, design, and methods within each tradition as there are differences between the two camps.

Is there any way to resolve the differences? The insight was, we believe, again provided by Lee Cronbach. In his 'Two disciplines' paper that we described earlier [75], he advocated a search for aptitude–treatment interactions – using more complex quantitative methods such as analysis of covariance to relate aspects of the individual learner (aptitudes) to curriculum design factors (treatments). As one recent example, a series of studies of learning anatomy from a computer showed that students with high spatial ability had a small benefit from being presented views of an animated specimen at multiple angles, but that students with poor spatial ability were seriously handicapped by multiple views [92]. The current use of methods such as confirmatory factor analysis and structural equation models is a logical extension of this aptitude–treatment interaction approach. However, after two decades, failures have far exceeded successes, and in our view, the more powerful methods now in vogue follow in this tradition, yielding little in the way of substantive theoretical explanation.

Cronbach's resolution in a later paper [93] was to abandon attempts at greater experimental control in favour of more careful observation:

> [This paper will] explore the consequences of attempting to establish in psychological experimentation, empirical generalisations in a world in which most effects are interactive. While the two scientific disciplines of experimental control and systematic correlation are designed to answer pre-stated formal questions, the time has come for more open-ended, inquisitive investigation that will more fully explore the richness of scientific reality.

Another way to think about the intent of an individual study is a categorisation developed by Schmidt [94], who described three goals for studies.

- *Description*, which focuses on the first step of the scientific method – observation. An approach is described, but no comparison is performed.
- *Justification* represents the opposite extreme, where the goal is to justify a particular approach by a careful experimental study showing it is superior to some alternative. The problem, as identified by Cook et al. [95], is that without sufficient theory specification the results may have limited application.

- *Clarification* studies are modelled on the scientific method, beginning with observation, proceeding to careful theory building, testing, and elaboration.

In an initial study, Schmidt found that 64% of 850 medical education studies reviewed were description studies, a further 29% were justification studies, and only 7% were directed at clarification [94]. Cook et al. looked at a different database of experimental studies only and found that 16% were description, 72% justification, and 12% clarification [95]. However desirable theory-based research is, it is a small minority of educational research studies. We hope that by drawing attention to these issues, this chapter may, in some small way, help to redress the balance.

References

1 Dolmans, D. (2003). The effectiveness of PBL: the debate continues. Some concerns about the BEME movement. *Medical Education* 37: 1129–1130.
2 Norman, G. (2002). Research in medical education: three decades of progress. *British Medical Journal* 324: 1560–1562.
3 Bordage, G. (2007). Moving the field forward: going beyond quantitative-qualitative. *Academic Medicine* 82: S126–S128.
4 Meltzoff, J. (1998). Research questions and hypotheses. In: *Critical Thinking about Research*, 13–30. Washington, DC: American Psychological Association.
5 Bordage, G. (2009). Conceptual frameworks to illuminate and magnify. *Medical Education* 43: 312–319.
6 Eva, K.W. and Lingard, L. (2008). What's next? A guiding question for educators engaged in educational research. *Medical Education* 42 (8): 752–754.
7 Albert, M., Hodges, B., and Regehr, G. (2007). Research in medical education: balancing service and science. *Advances in Health Sciences Education: Theory and Practice* 12: 103–115.
8 Cook, D., Beckman, T.J., and Bordage, G.E. (2007). Quality of reporting of experimental studies in medical education: a systematic review. *Medical Education* 41: 737–745.
9 Popper, K. (1959). *The Logic of Scientific Discovery*. London: Hutchinson and Co.
10 Norman, G.R. (1999). The adult learner: a mythical species. *Academic Medicine* 74: 886–889.
11 Regehr, G. (2010). It's NOT rocket science: rethinking our metaphors for research in health professions education. *Medical Education* 44 (1): 31–39.
12 Stokes, D.E. (1997). *Pasteur's Quadrant: Basic Science and Technological Innovation*. Washington, DC: Brookings Institution Press.
13 Cook, T.D. and Campbell, D.T. (1979). *Quasi-Experimentation: Design and Analysis for Field Settings*. Chicago, IL: Rand McNally.
14 Papadakis, M.A., Teherani, A., Banach, M.A. et al. (2005). Disciplinary action by medical boards and prior behavior in medical school. *New England Journal of Medicine* 353: 2673–2682.
15 Hojat, M., Paskin, D.L., Callahan, C.A. et al. (2007). Components of postgraduate competence: analyses of thirty years of longitudinal data. *Medical Education* 41: 982–989.
16 Rose, G. (2001). Sick individuals and sick populations. *International Journal of Epidemiology* 30: 427–432.
17 Norman, G.R. (2003). RCT = results confounded and trivial. *Medical Education* 37: 582–584.
18 Schuwirth, L.W. and van der Vleuten, C.P. (2006). Challenges for educationalists. *British Medical Journal* 333: 544–546.
19 Norman, G.R., Eva, K.W., and Schmidt, H.G. (2005). Implications of psychology-type theories for full curriculum interventions. *Medical Education* 39: 247–249.
20 Campbell, D.T. and Stanley, J.C. (1963). *Experimental and Quasi-Experimental Designs for Research on Teaching*. Chicago, IL: Rand McNally.
21 Goldacre, B. (2007). A kind of magic? www.guardian.co.uk/science/2007/nov/16/sciencenews.g2 (accessed 28 June 2018).
22 Norman, G.R. and Shannon, S.I. (1998). Effectiveness of instruction in critical appraisal (evidence-based medicine) skills: a critical appraisal. *Canadian Medical Association Journal* 158: 177–181.
23 Schmidt, H.G. and Moust, J.H.C. (1995). What makes a tutor effective? A structural equations modeling approach to learning in problem-based curricula. *Academic Medicine* 70: 708–714.
24 Lipsey, M.W. and Wilson, D.B. (1993). The efficacy of psychological, educational, and behavioral treatment. Confirmation from meta-analysis. *American Psychologist* 48: 1181–1209.
25 Meyer, G.J., Finn, S.E., Eyde, L.D. et al. (2001). Psychological testing and psychological assessment: a review of evidence and issues. *American Psychologist* 56: 128–165.
26 Woodward, C.A., Cohen, M., Ferrier, B.M. et al. (1989). Correlates of certification in family medicine in the billing patterns of Ontario general practitioners. *Canadian Medical Association Journal* 141: 897–904.
27 Koh, G.C.-H., Khoo, H.E., Wong, M.L., and Koh, D. (2008). The effects of problem-based learning during medical school on physician competency: a systematic review. *Canadian Medical Association Journal* 178: 34–41.
28 Colliver, J.A. and Markwell, S.J. (2007). Research on problem-based learning: the need for critical analysis of methods and findings. *Medical Education* 41: 533–535.
29 Schmidt, H.G., Rotgans, J.I., and Yew, E.H. (2011). The process of problem-based learning: what works and why. *Medical Education* 45: 792–806.
30 Cook, D.A. (2005). The research we still are not doing: an agenda for the study of computer-based learning. *Academic Medicine* 80: 541–548.
31 Ehrenberg, R.G., Brewer, D.J., Gamoran, A., and Willms, J.D. (2001). Does class size matter? *Scientific American* 285: 79–84.
32 Norman, G.R. and Streiner, D.L. (2007). *Biostatistics: The Bare Essentials*, 3e. Hamilton, ON: BC Decker.
33 Larsen, D.P., Butler, A.C., and Roediger, H.L. (2008). Test-enhanced learning in medical education. *Medical Education* 42: 959–966.
34 Lurie, S.J., Mooney, C.J., and Lyness, J.M. (2009). Measurement of the general competencies of the accreditation council for graduate medical education: a systematic review. *Academic Medicine* 84: 301–309.
35 Whitcomb, M.E. (2002). Research in medical education: what do we know about the link between what doctors are taught and what they do? *Academic Medicine* 77: 1067–1068.
36 Bligh, J. (2003). Editorial. *Medical Education* 37: 184–185.
37 Colliver, J.A. (2003). The research enterprise in medical education. *Teaching and Learning in Medicine* 15: 154–155.
38 Norman, G.R. (2006). Outcomes, objectives and the seductive appeal of simple solutions. *Advances in Health Sciences Education* 11: 217–220.
39 Gruppen, L. (2007). Improving medical education research. *Teaching and Learning in Medicine* 19: 331–335.
40 Eva, K.W., Reiter, H.I., Trinh, K. et al. (2009). Predictive validity of the multiple mini-interview for selecting medical trainees. *Medical Education* 43 (8): 767–775.
41 Tamblyn, R., Abrahamowicz, M., Dauphinee, D. et al. (2007). Physician scores on a national clinical skills examination as predictors of complaints to medical regulatory authorities. *Journal of the American Medical Association* 298: 993–1001.
42 Werner, L.S. and Bull, B.S. (2003). The effect of three commercial coaching courses on Step One USMLE performance. *Medical Education* 37: 527–531.
43 Cohen, P.A. (1981). Student rating of instruction and student achievement: a meta-analysis of multisection validity studies. *Review of Educational Research* 51: 281–309.

44 Eva, K.W. and Regehr, G. (2005). Self-assessment in the health professions: a reformulation and research agenda. *Academic Medicine* 80: S46–S54.
45 Gordon, M.J. (1991). A review of the validity and accuracy of self-assessments in health professions training. *Academic Medicine* 66: 762–769.
46 Davis, D.A., Mazmanian, P.E., Fordis, M. et al. (2006). Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *Journal of the American Medical Association* 296: 1094–1102.
47 Peterson, L.N., Eva, K.W., Rusticus, S.A., and Lovato, C.Y. (2012). The readiness for clerkship survey: can self-assessment data be used to evaluate program effectiveness? *Academic Medicine* 87: 1355–1360.
48 Levinson, A.J., Weaver, B., Garside, S. et al. (2007). Virtual reality and brain anatomy: a randomised trial of e-learning instructional designs. *Medical Education* 41: 495–501.
49 Cohen, J.J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Erlbaum.
50 Colliver, J.A., Markwell, S.J., Verhulst, S.J., and Robbs, R.S. (2007). The prognostic value of documented unprofessional behavior in medical school records for predicting and preventing subsequent medical board disciplinary action: the Papadakis studies revisited. *Teaching and Learning in Medicine* 19: 213–215.
51 Swanson, D.B., Norman, G.R., and Linn, R.L. (1995). Performance based assessment: lessons from the professions. *Educational Research* 24: 5–11.
52 Harden, R.M. and Gleeson, F.A. (1979). Assessment of clinical competence using an objective structured clinical examination (OSCE). *Medical Education* 13: 41–54.
53 Streiner, D.L. and Norman, G.R. (2003). *Health Measurement Scales: A Practical Guide to Their Development and Use*, 3e. Oxford: Oxford University Press.
54 Eva, K.W., Rosenfeld, J., Reiter, H.I., and Norman, G.R. (2004). An admissions OSCE: the multiple mini-interview. *Medical Education* 38: 314–326.
55 Hodgson, C.S., Teherani, A., Gough, H.G. et al. (2007). The relationship between measures of unprofessional behavior during medical school and indices on the California Psychological Inventory. *Academic Medicine* 82: S4–S7.
56 Mamede, S. and Schmidt, H.G. (2004). The structure of reflective practice in medicine. *Medical Education* 38: 1302–1308.
57 Kane, M.T. (2006). Validation. In: *Educational Measurement* (ed. R.L. Brennan). Westport, CT: Praeger Publishers.
58 Messick, S. (1989). Validity. In: *Educational Measurement* (ed. R.L. Linn), 103–104. New York: Macmillan.
59 Downing, S.M. (2003). Validity on the meaningful interpretation of assessment data. *Medical Education* 37: 830–837.
60 Newble, D.I. and Jaeger, K. (1983). The effect of assessments and examinations on the learning of medical students. *Medical Education* 17: 165–171.
61 Harasym, P.H., Woloschuk, W., Mandin, H., and Brundin-Mather, R. (1996). Reliability and validity of interviewers' judgements of medical school candidates. *Academic Medicine* 71: S40–S42.
62 Paul, A. (2004). *The Cult of Personality: How Personality Tests are Leading Us to Miseducate Our Children, Mismanage Our Companies, and Misunderstand Ourselves*. New York: Free Press.
63 Hodges, B., Turnbull, J., Cohen, R. et al. (1996). Evaluating communication skills in the OSCE format: reliability and generalizability. *Medical Education* 30: 38–43.
64 Hodges, B., Regehr, G., McNaughton, N. et al. (1999). OSCE checklists do not capture increasing levels of expertise. *Academic Medicine* 74: 1129–1134.
65 Eva, K.W. and Hodges, B.D. (2012). Scylla or Charybdis? Can we navigate between objectification and judgement in assessment? *Medical Education* 46: 914–919.
66 Ramsey, P.G., Carline, J.D., Inui, T.S. et al. (1989). Predictive validity of certification by the American Board of Internal Medicine. *Annals of Internal Medicine* 110: 719–726.
67 Norman, G.R. (2004). The morality of medical school admissions. *Advances in Health Sciences Education* 9: 79–82.
68 Bargh, J.A. and Chartrand, T.L. (1999). The unbearable automaticity of being. *American Psychologist* 54: 462–479.
69 Wainer, H. (1976). Estimating coefficients in linear models: it don't make no nevermind. *Psychological Bulletin* 83: 213–217.
70 Jamieson, S. (2004). Likert scales: how to (ab)use them. *Medical Education* 38: 1217–1218.
71 Glass, G.V. and Stanley, J.C. (1970). *Statistical Methods in Education and Psychology*. Englewood Cliffs, NJ: Prentice Hall.
72 Carifio, J. and Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. *Medical Education* 42: 1150–1152.
73 Norman, G. (2010). Likert scales, levels of measurement, and the "laws" of statistics. *Advances in Health Sciences Education* 15: 625–632.
74 Wikipedia (2010). Bonferroni correction. http://en.wikipedia.org/wiki/Bonferroni_correction (accessed 28 June 2018).
75 Cronbach, L.J. (1957). The two disciplines of scientific psychology. *American Psychologist* 12: 671–684.
76 Harden, R., Grant, J., Buckley, G., and Hart, I. (2000). Best evidence medical education. *Advances in Health Sciences Education* 5: 71–90.
77 Smith, M.L. and Glass, G.V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist* 32: 752–760.
78 Vernon, D.T. and Blake, R.L. (1993). Does problem-based learning work? A meta-analysis of evaluative research. *Academic Medicine* 68: 550–563.
79 Hamdy, H., Prasad, K., Anderson, M.B. et al. (2006). BEME guide no. 5: predictive values of assessment measurements obtained in medical schools and future performance in practice. *Medical Teacher* 28: 103–116.
80 Eva, K.W. (2009). Broadening the debate about quality in medical education research. *Medical Education* 43: 294–296.
81 Eva, K.W. (2012). The state of science 2012: building blocks for the future. *Medical Education* 46: 1–2.
82 Bordage, G. (2001). Reasons reviewers reject and accept manuscripts: the strengths and weaknesses in medical education reports. *Academic Medicine* 76: 889–896.
83 Driessen, E., van Tartwijk, J., van der Vleuten, C., and Wass, V. (2007). Portfolios in medical education: why do they meet with mixed success? A systematic review. *Medical Education* 41: 1224–1233.
84 Issenberg, S.B., McGaghie, W.C., Petrusa, E.R. et al. (2005). Features and uses of high-fidelity simulation that lead to effective learning: a BEME systematic review. *Medical Teacher* 27: 10–28.
85 Dornan, T., Littlewood, S., Margolis, S. et al. (2006). How can experience in clinical and community settings contribute to early medical education? A BEME systematic review. *Medical Teacher* 28: 3–18.
86 Hammick, M., Freeth, D., Koppel, I. et al. (2007). BEME guide no. 9: a best evidence systematic review of interprofessional education. *Medical Teacher* 29: 735–751.
87 Albanese, M.A. and Mitchell, S. (1993). Problem-based learning: a review of literature on its outcomes and implementation issues. *Academic Medicine* 68: 52–81.
88 Norman, G.R. and Schmidt, H.G. (1992). The psychological basis of problem-based learning. *Academic Medicine* 67: 557–565.
89 Eva, K.W. (2008). On the limits of systematicity. *Medical Education* 42: 852–853.
90 Cook, D.A., Hatala, R., Brydges, R. et al. (2011). Technology–enhanced simulation for health professions education: a systematic review and meta-analysis. *JAMA: The Journal of the American Medical Association* 306: 978–988.

91 Cook, D.A. and West, C.P. (2012). Conducting systematic reviews in medical education: a stepwise approach. *Medical Education* 46: 943–952.
92 Garg, A., Norman, G.R., Spero, L., and Maheshwari, P. (1999). Do virtual computer models hinder anatomy learning? *Academic Medicine* 74: S87–S89.
93 Cronbach, L.J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist* 30: 116–127.
94 Schmidt, H.G. (2005). Influence of research on practices in medical education: the case of problem-based learning. Paper presented at the Association for Medical Education in Europe, Amsterdam, The Netherlands (September 2005).
95 Cook, D.A., Bordage, G., and Schmidt, H.G. (2008). Description, justification and clarification: a framework for classifying the purposes of research in medical education. *Medical Education* 42: 128–133.

Further Reading

Bowling, A. (2014). *Research Methods in Health: Investigating Health and Health Services*, 4e. New York: Open University Press.

Chalmers, A. (2014). *What Is This Thing Called Science?* 4e. Buckingham: Open University Press.

Norman, G.R., van der Vleuten, C.P.M., and Newble, D.I. ed. (2002). *International Handbook of Research in Medical Education*, vol. 7. New York: Springer International Handbooks of Education.

Norman, G.R. and Streiner, D.L. (2008). *Biostatistics: The Bare Essentials*, 3e. Hamilton, ON: BC Decker.

Streiner, D., Norman, G.R., and Cairney, J. (2015). *Health Measurement Scales: A Practical Guide to Their Development and Use*, 5e. Oxford: Oxford University Press.

29 Qualitative Research in Medical Education: Methodologies and Methods

Stella L. Ng[1,2,3], *Lindsay Baker*[1,2,4], *Sayra Cristancho*[5,6], *Tara J. Kennedy*[7], *and Lorelei Lingard*[6,8]

[1]Centre for Faculty Development, St Michael's Hospital, University of Toronto, Toronto, Ontario, Canada
[2]Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada
[3]The Wilson Centre, University of Toronto, Toronto, Ontario, Canada
[4]Li K Shing Knowledge Institute, University of Toronto, Toronto, Ontario, Canada
[5]Department of Surgery, Schulich School of Medicine & Dentistry, Western University, London, Ontario, Canada
[6]Centre for Education Research & Innovation, Schulich School of Medicine & Dentistry, Western University, London, Ontario, Canada
[7]Pediatric Autism Rehabilitation Services, Stan Cassidy Centre for Rehabilitation, Fredericton, New Brunswick, Canada
[8]Department of Medicine, Schulich School of Medicine & Dentistry, Western University, London, Ontario, Canada

KEY MESSAGES

- Qualitative research explores social issues and processes, and people's experiences and perspectives in relation to social issues and processes.
- Qualitative research can contribute to theory and accumulating knowledge of complex social issues in medical education.
- Qualitative research encompasses multiple research methodologies, including ethnography, grounded theory, case study, phenomenology, narrative inquiry, action research, and discourse analysis.
- Qualitative sampling, data collection, and analysis methods must be selected for their suitability to a particular research question and methodology.
- Principles of ethics, rigour, and reflexivity should be considered when ensuring and appraising the quality of qualitative research.

Introduction: Paradigms and Purposes of Qualitative Research

What is Qualitative Research?

Qualitative researchers study social, relational, and experiential phenomena, situated within everyday contexts. For questions of how people make meaning of an experience, what is happening in a specific context, what a social process looks like, how a social construct or norm has come to dominate, or how people experience a particular phenomenon, a qualitative approach is appropriate. *How* and *what* questions are particularly suited for exploration through qualitative research (see Box 29.1).

The term 'qualitative research' encompasses a broad range of philosophical and theoretical traditions, methodologies, and methods, which the following sections will take up in detail. But most qualitative approaches share some basic principles. Qualitative research tends not to control or intervene, instead proceeding naturalistically.

BOX 29.1 When would I use qualitative research?

You would use qualitative research when you are interested in studying social, relational, or experiential phenomenon in naturalistic (rather than experimental) settings.

Qualitative research helps you:

- *Explore meanings*, e.g. What does 'being an advocate' mean to community-based physicians?
- *Derive understanding*, e.g. What do interprofessional relations look like in an ambulatory internal medicine unit?
- *Explain processes*, e.g. How do medical students learn to respond to professional dilemmas?
- *Provoke new ways of thinking*, e.g. How and why did competency-based education come to dominate medical education?
- *Describe experiences*, e.g. How did residents at two hospitals experience the duty hour reform?

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

Qualitative research seeks to understand and represent complexity, producing context-rich accounts of social phenomena. Qualitative researchers are often found observing and interacting with people in their everyday contexts in order to shed light on the phenomenon of interest. Or, they may be found gathering and analysing textual or visual representations of societal norms and structures (e.g. policy documents). The assumption underpinning these principles and tendencies is that the social phenomena of interest to qualitative researchers are inextricable from their contexts. Aligned with this assumption, the goal of qualitative research is the careful understanding of instances, instead of making claims to generalisability. Qualitative research does not aim to produce general truths but rather aims for in-depth explanations. Done well, qualitative research can offer transferable lessons and, over time, rich theory. It can identify injustices and inequities embedded in education systems and the subsequent opportunities for change. It can also describe the processes involved in producing particular learning outcomes, and can reveal new ways of seeing where entrenched and restrictive beliefs about health professions education otherwise occlude [1].

Origins of Qualitative Research in Medical Education

Qualitative research comes to medical education from the social sciences and humanities, from disciplines such as anthropology, sociology, education, and history. At various points, each of these disciplines used medical education as a site for research shaped by their own disciplinary questions and theories. Now, medical education researchers use tools from these disciplines to explore questions arising in the domain of medical education.

According to Harris [2], the importation of methods from these disciplines into medical education began in the 1980s amid calls for more theory building to complement the dominant paradigm of controlled experiments. Interestingly, these calls continue [3, 4].

Research Paradigms

Discussions of qualitative research often begin with discussions of research paradigms, which can help explain how quantitative and qualitative research differ. As Denzin and Lincoln [5] explain, paradigms are basic sets of beliefs that guide action; Harris describes them as 'cognitive road maps, taken-for-granted assumptions within communities of scholars' [2] that orient researchers towards meaning and the research endeavour. Paradigms encompass ontology and epistemology. Ontology refers to the study of being and nature of existence; it is linked to epistemology. Epistemology is the study of knowledge, including what counts as knowledge and how one can come to know. Ontology can be thought of as questions of 'what is' and epistemology as questions of 'what it means to know' [6]. For example, the ontology most commonly associated with medical research is realism, which assumes one true reality exists. Realism implies an epistemology of objectivism, which asserts that we can accurately and directly attain knowledge of the one true reality through perception. The most congruent paradigm for a realist–objectivist position, then, would be positivism, which attempts to empirically measure reality and asserts that it can. Pure positivism has given way to today's post-positivism [6].

Post-positivism, a common paradigm in medical education research, shares with positivism the belief that there is an objective reality that can be discovered if the correct research procedures are in place. What distinguishes post-positivism from positivism is the acknowledgement that complex human behaviour is shaped by individual motivations and cultural environments, and research must represent these complexities rather than elide them in search of a contextual 'essence' or truth. Most quantitative research aligns with a post-positivist paradigm, although, because it is the dominant research paradigm in medicine, rarely would a quantitative researcher make mention of their paradigm in a journal article; the dominant paradigm is assumed. Some qualitative research also derives from the post-positivist paradigm. Irby's account of how clinical teachers make decisions about what to prioritise in their round exchanges with students represents the post-positivist paradigm in his search for the essence of teachers' decision-making while paying attention to the contextual and individual features that shape this process [7].

More common to qualitative research is constructivism, which departs from post-positivism in its acceptance of reality and meaning as relative, produced through the interaction between researcher and researched. Research in the constructivism paradigm acknowledges the subjectivity of the researcher, producing accounts of a social phenomenon that reflect the researcher's interaction with the phenomenon. Lingard's accounts of tension, collaboration, and professional socialisation within operating room teams provide an example of this approach, as she views team communication through her training as a rhetorician and blends this perspective with those of study participants and 'insider informants' engaged in the collaborative analysis process [8–11].

Also prevalent in medical education research is work within the critical inquiry paradigm, which is identifiable by its goal of revealing power dynamics in studied phenomena and fostering empowerment through the careful description and analysis of these dynamics. Albert's account of tensions within the medical education research community uses Bourdieu's theoretical notion of field to explore the configuration of power relations in this research community [12].

Two additional elements of paradigm worth mentioning are axiology and rhetorical structure [13, 14]. Axiology refers to the place or role of values, and rhetorical structure to the use of language in 'writing up' the research. For example, a study drawing on feminist theory might be written in the first person and include explication of the researcher's own experience and position relative to the research. In contrast, a study informed by a post-positivist paradigm may adopt language that is more in line with the objectivist scientific tradition.

Although contrasting examples are useful to highlight the nuances of each paradigm, the researcher's paradigm does not inflexibly dictate methodological choice. A thoughtful consideration of research question and paradigmatic position will guide the qualitative researcher in selecting the most appropriate methodology for inquiry. The best methodological approach for a particular study depends on the research question, the research setting, and the objective of the research. These factors also determine what the best methods are. The selection of a well-aligned and well-justified paradigmatic position, methodological approach, and methods relative to the research question is a primary marker of methodological rigour in qualitative research. This notion of 'best fit' and alignment is important when considering the quality of qualitative research [15].

Together, the components of a research paradigm should be congruent with the methodology. So, one's ontology informs one's epistemology, which together form one's paradigm, which guides selection of methodology, which guides the use of particular methods. See Figure 29.1 for a visual depiction of paradigmatic alignment. For a helpful summary of research frameworks and paradigms, see Chapter 27 of this book and for a succinct summary, see the Bergman et al. resource listed in the *Further Reading* section of this chapter.

Methodology and methods are described next. As in quantitative research, in qualitative research there is a distinction and synergistic relationship between methodology and method. Methodology refers to the theory of how inquiry should proceed and what the research aims to produce. It entails assumptions, principles, and procedures governing the use of particular methods [16]. Methods are the specific investigative tools or procedures used to gather and analyse data [16].

Your research paradigm will influence the types of questions you ask and the methodologies you use to answer these questions, which in turn influences the methods you will adopt and how you will apply them. Alignment of paradigmatic position, methodological approach, and methods relative to your research question is a prerequisite for rigour and a determinant of quality in qualitative research.

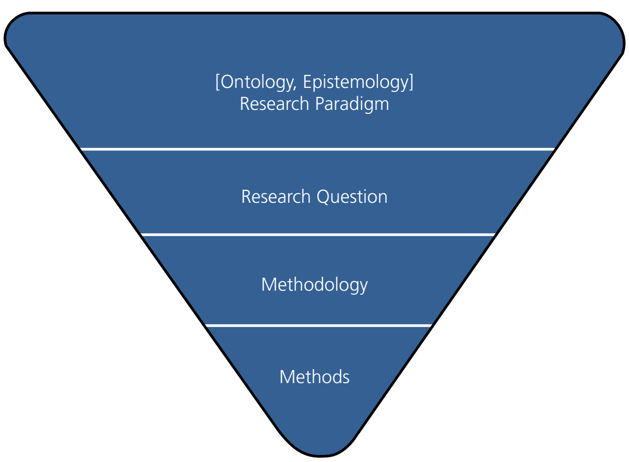


Figure 29.1 Paradigmatic alignment.

Qualitative Research Methodologies

Qualitative research encompasses an eclectic group of research methodologies, which are linked by their common aim to explore social processes through interpretation or representation of qualitative data. These methodologies, or 'systems of inquiry', are, according to Denzin and Lincoln [5], 'a bundle of skills, assumptions, and practices' that the researcher employs to generate and address their research questions. The various qualitative methodologies stem from different philosophical and/or theoretical perspectives, with resultant implications for the research process. Although there can be a significant overlap between them, and some qualitative methodologists may creatively and effectively employ combinations of methodologies, the following section provides a brief overview of seven major qualitative research approaches (summarised in Box 29.2), with examples of their contribution to medical education research.

Ethnography

The tradition of ethnography originates in the field of anthropology, in which a researcher would travel to study an 'exotic' tribe [17, 18]. Current-day ethnography often rejects the traditional notion of a privileged researcher, and ethnographic studies are now more likely to occur in local subcultures (such as a medical school or an operating room) than in far-flung locations [19]. However, ethnographic studies carry on the practice of long-term engagement in a study setting, and the collection, through observation and conversational interviews, of data that are analysed to

BOX 29.2 Common qualitative research methodologies

Once you've identified your research question, you want to choose the appropriate methodology to answer that question. Ask yourself: what is the goal or purpose for your research?

Methodology	Goal/purpose
Ethnography	Describing and interpreting a cultural and social group
Grounded theory	Developing a theory grounded in practical experiences of participants
Case study	Developing in-depth understanding of a 'bounded system' (programme, event, activity, process, group, etc.)
Phenomenology	Understanding the essence of a phenomenon through those who experience it
Narrative inquiry	Exploring, in depth, one or more individuals' experience(s) through story
Action research	Producing social change through the research process and the direct engagement of research participants in the research process
Discourse analysis	Studying the uses and effects of language in society

understand the meaning inherent in the everyday activities of a particular social group. There are a number of classic ethnographies in the domain of medical education, including Becker's *The Boys in White* [20], a study of the nature of student culture in medical school, and Bosk's *Forgive and Remember* [21], a study of the treatment of medical error in postgraduate surgical education. An example of a variant, critical approach to ethnography within medical education research is Mykhalovskiy's investigation of the social organisation of evidence-based medicine using institutional ethnography [22].

Grounded Theory

Grounded theory research explores social phenomena through the development of theoretical explanations that are 'grounded' in (i.e. derived from) the practical experience of study participants [23]. Grounded theory was developed by two sociologists, Glaser and Strauss, in the 1960s, to provide a systematic approach to the analysis of qualitative data that would live up to the standards of 'rigour' imposed by the quantitative paradigm and that would focus on theory generation rather than theory testing [23]. Since its inception, four main methodological schools of grounded theory have gained popularity [24–26]: post-positivist or classicist approach [23], pragmatist [27], constructivist [28], and post-structuralist [29]. Key elements across all 'schools' of grounded theory include:

- *iterative* study design (cycles of simultaneous data collection and analysis, in which the results of the ongoing data analysis inform the subsequent data collection)
- *purposeful* and *theoretical* sampling (purposeful selection of data sources for their ability to provide data that would confirm, challenge or expand a developing theory)
- a *constant comparative* analytic approach to data analysis (through which incidents or issues of interest in the data are compared against other examples for similarities and differences) [30]
- *theoretical saturation or sufficiency*, the end point of data collection for a particular study, which occurs when no new codes or concepts are found in newly collected data; sufficiency has been proposed as the more appropriate term by some grounded theorists, particularly constructivists, who assert that saturation may be a misnomer given the interpretive approach to the grounded theory research process [31, 32].

In the domain of medical education research, Ginsburg has used grounded theory to develop a behavioural theory of professionalism [33–38]. Watling has used grounded theory to theorise feedback in medical education [39–43].

Case Study

Case study research involves an in-depth analysis of a phenomenon within context. Not to be confused with case presentations common to medical education, the case in a case study research project is defined as the 'bounded system' to be researched. The bounded system is defined from the outset; it could be a programme, an event, an activity, a process, a group, or even an individual [44]. The case study can have intrinsic value – that is, one wishes to understand the case itself – or instrumental value – that is, the case can be used as one situated instance of a larger phenomenon, and its study enables understanding of that larger phenomenon [44]. One hallmark of case study methods is triangulation, which is the use of multiple data collection tools or data sources to gain rich insight into the study phenomenon from multiple perspectives (see the section 'Ethics, Rigour, and Reflexivity', later in this chapter).

Combining multiple cases can be a useful way to understand what is common and what is different about the same phenomenon across two different cases [44]. For example, one could study the impact of a new resident duty hour policy in three different types of hospitals, in order to unpack the common effects and the contextually situated differences of the policy reform. In this example, each hospital site serves as a 'case'. Studying multiple cases in this manner is called multiple or collective case study.

An example of a qualitative case study in medical education is Perley's study of a group of primary care physicians to explore their use of the 'curbside consultation' with colleagues as a continuing education tool [45]. In a study of patient-engaged teamwork in a rural setting, Casimiro et al. used case study methodology and within it borrowed and modified grounded theory analysis techniques, demonstrating the flexible use of case study design in combination with other aligned approaches [46].

Phenomenology

Phenomenology arose early in the twentieth century from philosophical reflections on consciousness and perception. Phenomenological research aims to understand the essence of a social phenomenon from the perspective of those who have experienced it [47]. Phenomenology with a descriptive intent involves the 'bracketing' (or putting aside) of the researcher's own preconceptions and perspectives in order to understand the 'lived experience' of the research participants [47]. Phenomenological studies often involve an in-depth exploration of the experiences of a relatively small number of individuals. Bearman has used phenomenology to explore the experiences of medical students during interactions with virtual patients [48].

Narrative Inquiry

Narrative inquiry stems from the ancient practice of storytelling as a method of communicating, arranging, and interpreting human experience. Narrative inquiry is a qualitative approach that 'solicits and analyzes personal accounts as stories' [49], using these stories as a means of understanding or making sense of a particular experience or situation. Narrative analysis seeks meaning in the content, structure, context, and relational aspects of a story [50]. Narrative methods have been promoted as an educational tool for teaching empathy and communication skills to medical students [51], but are also used to address research questions in medical education. Ventres has used narrative case reports of patient interviews conducted by residents to compare differences between patients' and physicians' perspectives [52]. Bennett et al. have incorporated narrative inquiry into their study of becoming a 'good doctor' [53].

Action Research

Action research has its roots in the social activism movements of the mid-twentieth century. Key principles of action research are the explicit aim of producing social change through the research process and the direct engagement of research participants in the research process [54]. Action research classically occurs through sequential cycles of planning a change, implementing the change while observing the process, and reflecting on the consequences of the change [55]. Participants collaborate with researchers to construct the results of the research and implement social change. An action research approach was employed successfully in the design and implementation of a new general practice curriculum in Dundee, Scotland [56]. Action research principles also inform integrated knowledge translation approaches – the processes of bridging research and practice throughout the research process – by emphasising engagement of research subjects and/or the enactors of research findings [57].

Discourse Analysis

Discourse analysis is an approach to qualitative research that analyses data at the level of language. It is an umbrella term that references a number of different approaches to the analysis of socially situated language use. Discourse is a term meaning 'socially situated language' [58]. Generally, the aim of discourse analysis is to make explicit what is normally taken for granted about language use or to show what talking accomplishes in a particular social context. Some discourse analysts, often in the domains of linguistics or conversation analysis, work to understand the complex mechanisms and structures of social language. Others, in fields like sociolinguistics or critical discourse analysis, use talk as a source of evidence about social processes. In critical discourse analysis, a central concern is the explication of power relations, with analysis focusing on identification of that which is constructed as 'truth' within a particular discourse and how those truths, from a socio-historical perspective, came to be [59]. Discourse analysis has an extensive history in the study of physician–patient communication [60], but has been more recently applied to the domain of medical education [61]. Hekelman et al. conducted a discourse analysis to investigate the changes in language use in the teaching encounters of a physician–teacher who was enrolled in a peer-coaching programme intended to improve clinical teaching skills [62].

The Ongoing Evolution of Methodologies

Methodologies are not static. For example, the development of various schools of grounded theory, over time, exemplifies the shifting nature of methodology [26]. Methodologies also have permeable boundaries; they may be blended together, borrowing concepts as needed to best address a research question [63]. For example, researchers may use a collective case study design to deeply understand the experience of resident fatigue, defining their cases as two different residency programmes. The researchers may have chosen collective case study out of an interest to deeply understand the phenomenon and policy change in context, with attention to differences and similarities across two contexts experiencing the same phenomenon. In analysing the data, the researchers may use grounded theory analytic techniques, such as coding techniques and the constant comparative method. If the study is designed thoughtfully and with attention to paradigmatic alignment, this borrowing of grounded theory analytic methods for use in a case study methodology would not only be acceptable but beneficial.

Qualitative Research Methods

Qualitative research studies are carried out through a set of tools for data collection and analysis. In the following section, methods for data collection and approaches to data analysis will be reviewed separately. This separation of data collection from analysis is somewhat artificial in qualitative research for two reasons. First, many qualitative studies employ an iterative study design [2] in which cycles of data collection and analysis occur simultaneously, influencing one another. For example, analysis of an early set of interviews in a study may inform subsequent interview questions [23]. Second, the choice of data collection methods necessarily informs the choice of the analytical approach and vice versa. For the sake of clarity, in this chapter, data collection and data analysis methods are considered separately.

Data Collection Methods

The various qualitative research methodologies have in common a set of data collection tools. Although certain qualitative approaches are classically associated with particular data collection or analysis methods (e.g. ethnography with participant observation or critical discourse analysis with document analysis), contemporary qualitative researchers commonly choose from the available methods the one(s) that is (are) best suited to address the research question at hand (see Box 29.3). Methods must be enacted thoughtfully, not prescriptively [64].

Interviews

Individual interviews are probably the most familiar and the most often used form of data collection in qualitative medical education research [2, 65]. Interviews provide access to participants' personal perspectives and relevant experiences on a number of topics [66]. The qualitative interview typically goes 'in-depth' [66] to provide a rich and detailed exploration of a research topic and generally lasts between 45 minutes and a few hours [67]. Qualitative research interviews usually follow a 'semi-structured' format. The semi-structured interview is guided by a predetermined set of open-ended questions (in an interview/topic guide), but the researcher and participant are free to pursue additional relevant topics as they arise. On either end of the spectrum of interview types, interviews can be unstructured, wherein the researcher may name a topic and have the participant speak to it freely, or structured, whereby the interview guide is followed strictly. However, while a structured interview would indeed gather qualitative data, such a structured approach would be better

BOX 29.3 Common qualitative data collection methods

Choosing the right qualitative research method is like choosing the right tool for the job. Ask yourself: which method(s) will best help me address my research question?

Method	What are they?	When would you use them?	What should you consider?
Interviews	A purposeful conversation, not an oral survey Typically 45–90 minutes Interviewer is there to listen, to observe with sensitivity, and to encourage the participant to respond Types: semi-structured; unstructured; informal	When you want in-depth exploration of a particular topic When you want a flexible, iterative, responsive approach When you want to access participants' understandings, attitudes, perceptions	Power differentials between interviewer and participant Costly (interviewer time, transcription etc.) Skill-level of interviewer
Focus groups	Roughly 4–12 individuals interviewed together, but not a 'group interview' Participants usually similar in some way; or otherwise purposively chosen NB: The focus group is the unit of analysis, not the individual participants (n = number of focus groups)	When the interaction among group members is *desirable* When the perspective of a *group* is desirable When group 'consensus', interaction, or capturing divergent views, is a goal	Dominant people may take over Challenging to moderate May fall victim to social desirability biases
Observations	Naturalistically observing people in their setting Recording detailed jottings, which become detailed fieldnotes Capturing anecdotes, reflections, physical space; relations among people and objects; atmosphere or tone Types: participant observation; non-participant observation; walking interview	When you want to capture behaviours and actions When you want to study the cultural and relational aspects of social phenomena When you want to build a thick description of a particular context	Researcher interpretations might be inconsistent with the meanings participants ascribe to their experiences Participant reactivity: researcher may impact participant behaviour Time consuming Challenges accessing (being able to spend time within) the field
Extant texts and visuals	Texts and visuals are gathered Examples include policies, curricula, art, film, news media Types: archival; historical; purposive	When you want to understand how a social construct or institutional practice is represented and shaped	Setting bounds for data collection (e.g. time period) Access (e.g. internal policies) Privacy
Elicited/ generated texts and visuals	Texts and visuals are created by participants and/or researcher Examples include written reflections, photographs, drawings Types: participant generated; researcher generated	When you want to access researcher or participant understandings or representations of experiences, phenomena or processes When you want to elicit different insights than an interview or focus group alone would provide	Time consuming Rapport-building Fit with research question

suited to supplement a quantitative study and would be unlikely to provide sufficiently rich data for the purposes of an interpretive, qualitative research study. Qualitative interviews are usually audiotaped and later transcribed to facilitate analysis, but recent advances in analysis software allow analysis directly from a digital audio or video recording. Researchers should budget for the transcription process if transcription will be required.

Interviews are not meant to accurately recount events or to detail the outcomes of an intervention; rather they focus on the participants' interpretations and meaning-makings about events or experiences [68]. Nonetheless, interview researchers must be careful to avoid leading or closed-ended (yes/no) questions. Consider, for example, the potential difference in responses to the questions 'Did you experience any barriers to mentorship?', 'What barriers to the mentorship process have you encountered?', and 'Could you tell me about your experiences of mentorship?'. The first question may be more suitable for a structured interview or oral survey; the second question may elicit more robust responses yet it primes the participant to think of barriers; the third question invites the participant to share what they perceive as relevant and meaningful, and to make meaning of it as they talk. Interviewers shape the interview through their setting of the tone and building of rapport, how they engage with the participant, and what additional probes they put forth. Therefore, it is important for researchers to attend to the power dynamics of the interview. For example, candid opinions from medical students about their experiences during clerkship are unlikely in an interview conducted by the clerkship director. These considerations also make the chosen interviewer's skill particularly important to the overall quality of the study.

Focus Groups

Focus groups have recently become well known as a marketing research tool, but they have a long history in the domain of social sciences research. Focus groups are sessions involving 4–12 participants and a moderator or facilitator who guides the group discussion of a topic relevant to the research question [69]. Focus groups provide access to multiple stories and diverse experiences in an efficient manner. But they are not merely group interviews. Instead, focus groups provide a dynamic and socially interactive exchange among participants that can stimulate exploration of contrary opinions, reflection on group norms and common practices, and exposure of taken-for-granted values [70]. Like individual interviews, focus group discussions often follow a semi-structured format and are audio-recorded and transcribed for analysis. The focus group moderator also records notes on group dynamics and interactions.

Researchers using focus groups must consider whether their topic would benefit from exploration in the synergistic and dynamic focus group format. Some deeply personal topics might be more safely or productively explored in an individual interview. Attending to power dynamics and social desirability tendencies are also critical in focus group methods; one influential, opinionated group member can monopolise the discussion, or participants may try to appease one another or the researchers. It is the facilitator's job to mitigate these concerns [69]. Researchers should also note that each focus group is the unit of analysis, not each individual participant.

Observation

Observation of study participants as they go about their regular activities can provide powerful insights into social processes. Researchers conducting observations have access to data on what participants do and not just on what they recall or say they do [71]. Qualitative researchers conducting observations make jottings in the field, which become more elaborated records called 'field notes'. Field notes can be structured to capture details such as the content of conversations, the context of discussions, the participants and intended audience for relevant comments, and the nonverbal nuances that accompany these interchanges [72]. Observations are sometimes accompanied by audio recording of 'naturalistic' conversations, which are later transcribed for analysis. Informal interviews may also occur in the midst of observations, as the researcher seeks clarification or further insight from participants.

Beginning researchers may worry about how they will accurately capture everything happening in the field, from conversations to environmental features. Keep in mind the research question and the notion that over time in the field, an observational focus will become clearer and narrower, and the jottings and elaborated field notes richer. In terms of accuracy, appropriate representation of the subject(s) and attempts to check in with participants about the researcher's understandings should be made. However, a constructivist or critical ethnography will involve the researcher's theory-informed interpretations of what they are observing. The process of the researcher making meaning of what s/he is seeing should involve clarifications with participants and connections with extant theory, and should not be considered inaccurate if done with attention to rigour and reflexivity (see the section on Ethics, Rigour, and Reflexivity, later in this chapter).

Observational researchers must also deal with 'participant reactivity' [73]; that is, the manner in which the observer's presence may shape participants' behaviours. The Hawthorne effect asserts that the observer influences the behaviour of study participants, but the effect has been questioned in terms of its pervasiveness [73]. In some medical education research contexts, the Hawthorne effect may not play out due to the acute, fast-paced, and highly observed nature of the environment [73]. Nonetheless, various techniques may help address participants' reactivity to the presence of researchers. Assuming adequate access to the field has been well-established, at times a challenge in and of itself, some researchers will spend long periods of time in the field to allow participants to become accustomed to their presence. Others will not reveal the specific focus of their observations to prevent participants from altering specific behaviours (e.g. a researcher might obtain consent to observe all clinical teaching in an intensive care unit without revealing to participants that the research question related specifically to the teaching of technical skills) [74]. Still others will take care to document evidence of the impact of their presence and then reflect on and write about the significance of this impact on their results.

Recently, medical education researchers have used in situ or walking interviews as a form of observational and interview-based data collection [75]. In these methods, the researcher journeys with the participants as they go about their lives. For example, one might travel with a clinical trainee as they make their journey to work in a remote community. Or one might shadow a resident throughout their call shift. Along the way, jottings, drawings, and formal and informal interviews may be used to capture insights relevant to the research question. While time-consuming, this method may offer useful knowledge that other forms of observation and interviewing may not, including a sense of fatigue of a participant, insights triggered by environmental cues, and geographical and spatial features [75].

Texts and Visuals

Extant

In the domain of medical education, a myriad of texts are used and created on a daily basis, many of which can yield important insights into educational processes. Sources of text for analysis include course curricula, assignments and examinations, student and faculty evaluations, clinical notes, and policy documents. More recently, texts from websites, email correspondence, and even digital images and video have been included in qualitative analyses [76]. Analysis of pre-existing documents can be an inexpensive data collection method, and because they were created for purposes other than research, the content of these data is not influenced by the research process [77].

Extant texts can be gathered in numerous ways, for example, as a comprehensive archive of historical records, or as a purposively collected corpus to address a particular question. Researchers must consider the bounds (e.g. temporal or topical) of the texts they wish to gather, how they will access the texts, and whether relevant privacy issues have been addressed.

A common use of textual documents for qualitative analysis in medical education is the analysis of documents produced as course assignments by students. For example, Olney analysed written 'experience summaries' created by medical student participants in a community service project to explore learning outcomes [78].

Elicited

A variety of texts may also be requested as part of the research approach. For example, asking medical students to write about their experiences can offer elicited texts for analysis. Increasingly, visual methods are making their way into the data collection toolbox of qualitative medical education researchers. When used as elicitation tools, visual methods add to the exploration of personal perspectives during interviews, particularly of situations that are difficult to verbalise or that contain multiple interacting dimensions. A variety of visual methods can be used to elicit interviews, including drawings, photos, videos, maps, etc. They can be generated by participants or by researchers [79].

When participant-generated, the goal of visual methods as elicitation tools is to help participants take time to reflect deeply about the situation/experience and be free to explore different aspects impacting the situation/experience while creating the visual. When researcher-generated, the main goal is to create common ground between researcher and participant. By sharing in the researcher's experiences, participants may feel more comfortable to explore difficult or complex situations.

Before using visual methods, consider the following: they can make the data collection process more time consuming and they require concerted effort from the researcher during the rapport building stage as they may be perceived as intimidating. Researchers must be careful while choosing when and when not to use visuals during interviews. Certain research questions are more amenable to visual methods than others. Research questions that intend to explore the multiple dimensions affecting an experience and their interrelationship are usually more suitable, as visuals may allow participants 'play' (via metaphors) with different ways of framing, representing, and re-telling their stories. Similar to interviews, researchers must attend to power dynamics when using visuals. In this case, researcher-generated visuals might be the option to choose as they may help participants offer their perspective through another person's experience.

Data Analysis Methods

Qualitative data analysis is the process of making sense of a qualitative data set. Qualitative data analysis is an ongoing process of reading, reflecting on, and questioning the meaning of the data as they are collected. It can be conducted individually or as part of a research team that analyses as a group or meets to compare and discuss results of individual analytical work.

Although the different qualitative approaches involve somewhat different analytical procedures, there are some basic processes that are common to most qualitative analyses. The most common of these is coding. Coding is a process of sorting or organising the data, eventually forming categories representing similar trends [23].

The first step in the coding process is the selection of the unit of analysis, which can be based on topical focus or data structure. For example, analysis of medical student interviews about professionalism might involve coding for the settings in which professional lapses occurred, or for types of professional behaviour, or for specific words or phrases used by participants to describe unprofessional acts. Or, the unit of analysis can begin as small as each line of a transcript, or larger meaning units. Coding for more than one of these different units of analysis might occur over time. As the data are being sorted into categories or codes, names or labels are created for the codes that describe the essence of the category, and memos or reflective notes are written to document the process of the analysis and record reflections and analytic ideas as they arise. Qualitative software can be used as a data management tool to keep track of the coding process as it proceeds, but the cognitive, meaning-making work of categorising data, identifying trends, and interpreting meaning is done by the researcher(s).

The specific approaches to data analysis in qualitative research are wide ranging. They are illustrated below in broad clusters explaining thematic analysis, the analysis of visual data, engaging teams in analysis, and interpretation and writing.

Thematic Analysis

The most commonly used qualitative analysis approach in medical education is the organisation of data according to topics, ideas, or concepts, often called themes. Variations of thematic analysis are used in many of the qualitative approaches, and a number of different systems of thematic analysis have been developed (e.g. content analysis [80] and constant comparative analysis) [23, 30]. The basic process of thematic analysis is to identify instances in the data set that are similar in concept. As further related examples are identified, a progressively richer understanding of the concept is developed, and as other important concepts are identified in the data, the relationships between concepts or themes are explored. The set of themes can then be used for description, theory development, or interpretation (see the section on Interpretation and Writing, later in this chapter). Thematic analysis has been used to explore many complex issues in medical education, e.g. Burack's study of the process of medical students' decision-making on specialty choice [81].

It is important to note that the manner and extent to which thematic analysis is closely tied to the data or abstracted beyond depends on the methodological approach. For example, in constructivist grounded theory, initial codes should be concrete and representative of the data. As coding progresses, codes should become progressively more conceptual and abstract, with multiple initial codes being clustered or categorised together to form a broader conceptual code or theme, which will eventually be incorporated into the developed theoretical model [30]. Contrastingly, in a post-positivist action research project, the development of theory is not imperative, but rather the grassroots adoption of practices that will result in positive change in a local context. Themes may thus be useful at the level of description and identification of practical challenges and creative solutions [80].

Polytextual Thematic Analysis: A Note on the Analysis of Visual Data

Visuals are a versatile source of data. They can be analysed alone or in conjunction with interview/narrative data. If looking at the visuals alone, some researchers may be interested in categorising pictorial elements of the visuals (e.g. emotions as a theme), while others may be more interested in the meaning of the drawing as a whole (e.g. the futile story of cancer patients). Others may only consider the form as in the types of visual metaphors used (e.g. a tombstone), while some others may be interested in the content of those metaphors (e.g. dying process). Regardless, analysing visual data alone has created some controversy; hence aesthetic analysis frameworks have only recently started to appear in the literature [82].

This lack of consensus and explicit guidance on how to handle visual data alone with systematic rigour and transparency has led researchers to adapt existing analytical approaches to consider visual and interview data together [83–85]. One such adaptation is the development of polytextual thematic analysis [85]: 'It is polytextual in that it assumes that all texts (including visuals) are predicated on one another, and each can only be read by reference to others. It is thematic in that it attempts to identify the repetitive features or themes in the data that enable patterns to come into view.' For example, researchers may choose to start with a face-to-face analysis process in the form of a 'Gallery Walk' [86]. As in a gallery, drawings are hung on walls in a room and researchers begin the analysis by inspecting each drawing to capture their first impressions about repeated visual elements and the general feeling each drawing evoked for them. Once recurring themes are identified and categorised through discussions among research team members, interview transcripts are brought into the analysis to constantly compare drawings and stories highlighting similarities and differences and explore whether themes cluster together to form higher order themes. As in traditional thematic analysis, this process of moving between drawings and stories should be repeated until the research team makes a sufficiency judgement.

Team Analysis

Medical education researchers commonly conduct analysis using teams of researchers. The purpose of involving more than one individual in the analysis varies, and depends on the epistemological stance of the work. Post-positivist qualitative work uses multiple coders with an aim of achieving consensus, while constructivist qualitative work acknowledges and capitalises on different researcher perspectives on the data, treating these as a form of investigator triangulation [87]. Bringing multiple perspectives to the analysis process may be particularly valuable when studying interdisciplinary or interprofessional phenomena, such as team collaboration and communication. Researchers may resort to team analysis to handle a large, multi-site data set where iteration of data collection and analysis cannot be feasibly centralised. Or multiple analysts may arise in training situations, when graduate students or novice researchers are participating on the research team to develop their skills.

Regardless of why an analysis team is used, doing so effectively requires attention to four key issues. First, there needs to be explicit acknowledgement of each analyst's position in relation to the data – their values and assumptions, their politics and identity. These characteristics will influence what each analyst 'sees' in the data, so they should be explicitly laid out and revisited as the analysis unfolds. Second, the analytical team must thoughtfully negotiate among these analytical perspectives [88]. This is particularly true in constructivist qualitative research, when different perspectives are integrated to enrich the analytical insights rather than brought into consensus. Analytic memos should explicitly record these negotiations and analytical decisions arising from them. Third, the analysis should be informed by a comprehensive and centrally updated codebook that reflects the emerging codes and the questions that have arisen as analysts from different perspectives have applied them [89]. Finally, consistent procedures should be established for data handling. We recommend that a single researcher be responsible for updating the codebook (whether this is a paper document or housed in data analysis software). Similarly, a single researcher should coordinate the analysis process in qualitative analysis software, to ensure accuracy of versions and coherence of the final analytical product.

Interpretation and Writing

The final stage of qualitative analysis is the process of interpretation, or finding the pivotal meaning in a data set. Without interpretive work, qualitative research produces merely a catalogue of ideas or themes. Important as those ideas may be, qualitative studies that do not take the next step of exploring the meaning at an interpretive level have not fully exploited the power of qualitative research.

There are different approaches to interpretation in qualitative research. In some qualitative approaches, the production of a thick, rich description of a social phenomenon is the goal of the research process (e.g. phenomenology). In other approaches, the development or expansion of theoretical explanations is the aim (e.g. grounded theory). In still other qualitative approaches, the meaning of a data set is considered through the lens of pre-existing theory, such as feminist, rhetorical, or Marxist theory.

The process of 'writing up' has been posited as an important tool in the toolkit of methods that qualitative researchers employ [90]. In a constructivist paradigm, writing can be considered a part of the interpretive inquiry process at the stage of coding, when memos are written by researchers to document the analytical process and associated reflective thinking as it unfolds. These memos, iteratively refined, may ultimately lead to the published written form of the qualitative work, and the act of memo writing is thus an intrinsic part of the interpretive inquiry process [91]. The final manuscript or other creative representation (e.g. theatre production) of a qualitative study combines elements of reporting the study with making connections and interpretations in the discussion to ensure it contributes to an ongoing scholarly conversation in its field [92].

Ethics, Rigour, and Reflexivity

General questions relating to ethics have been dealt with in Chapter 27 of this book but particular ethics issues arise in the collection and analysis of qualitative data. These include both procedural ethics (how the research is conducted to protect research participants from harm) and practical ethics (how the researcher conducts himself or herself in what Guillemin and Gillam call 'ethically important moments' [64]). Reflexivity is important in qualitative research. It refers to the way researchers identify, articulate, and consider the influences shaping research before, during, and after the research; it thus contributes to both the ethics and rigour of qualitative research [64]. It is asserted as a sensitising concept researchers should use as they negotiate ethics tensions and quality issues that may arise in their interactions with participants in the field (see Box 29.4).

Qualitative researchers both within and outside the domain of medical education have sought to articulate criteria for judging the quality of a qualitative report. Journals have published papers with guidelines [93–97], and qualitative leaders have offered overarching concepts such as 'trust-worthiness' [98], 'utility' [99, 100], and authenticity [101]. Debate exists about the use of various criteria [102], but, fundamentally, issues of rigour and ethics in qualitative research must be attended to throughout the research process, with a reflexive approach (see Box 29.4).

Beyond general expectations for rigour in qualitative research, specific methodologies suggest their own methodology-specific quality criteria. For example, in constructivist grounded theory, Charmaz [91] asserts four main criteria for rigour: credibility, originality, resonance, and usefulness. In phenomenology, a hallmark quality criterion is termed the 'phenomenological nod', which refers to the resonance of the research findings with the reader's own

BOX 29.4 HOW TO: How to strive for ethics and rigour through reflexivity

The following guiding questions are meant to prompt reflexive considerations of ethics and rigour, and offer researchers a springboard from which to develop additional questions specific to their own research. For additional guiding questions regarding reflexive research, see Baker et al. reference in "Further reading.

Researcher responsibilities and team development

- What are the boundaries in your role as a researcher?
- What principles, processes, or guidelines will you draw upon to ensure the quality of your research?
- How will you attend to and manage power differentials between participants and researchers?
- Does the research team have the appropriate experience and qualifications?
- How will you ensure you obtain informed consent in a valid manner?

Sampling and recruitment

- Are the right people/activities being sampled?
- Is the sample size likely to yield sufficient insight?
- Does disconfirming data need to be sought?
- Does theoretical or purposive sampling need to be conducted to further explore a developing or emerging concept or theme?

Data collection

- How will you ensure participants don't feel coerced?
- How will you protect the privacy of your participants?
- How do you build a rapport and make participants feel comfortable?
- Is the researcher's relationship to the participants/setting considered and explicated?
- What will you do if your participant becomes distressed during an interview?

Data analysis

- Is it appropriate to share analysis with participants in the study? What purpose does this serve? If participants disagree with your analysis, what will you do?
- Can your analysis be audited?
- How will you select your representative data excerpts (e.g. quotations)?

Writing and sharing the work

- Will you present your data within or without context?
- Does anonymising your data remove participant voice?
- Have you made thoughtful plans to share research findings, with the right audience(s), in a meaningful way?
- Is it appropriate to involve participants in dissemination (e.g. visual methods)?

experience [103]. This section outlines and illustrates some general principles of rigour to assist the newcomer in their appreciation of 'quality' in qualitative research.

Researcher Responsibilities and Team Development

Researchers should consider their role relative to the research. In the case that the researcher is an insider to the research context, additional considerations regarding roles and power relations include whether the researcher can ethically consent participants to participate, conduct interviews, and if yes, how power relations would be handled. The researcher is also responsible for considering who else is needed on the team to balance or strengthen perspectives, and for ensuring quality is considered from start to finish of the research process.

Sampling and Recruitment: Adequacy and Appropriateness

Sampling in qualitative research is not just about 'how many' subjects to include in the study. Because qualitative research explores social and experiential phenomena, deciding whom to include and exclude is a critical step in the sampling logic. A social phenomenon often engages a wide variety of participants, and the researcher must justify their decisions about who best to observe/interview and who to leave out of their study boundaries. In some qualitative research methodologies, sampling refers not only to individuals but also to groups, concepts, or documents [91]. In case studies, one must sample for cases first, then sample within cases (e.g. three hospitals implementing resident duty hour reforms serve as the cases sampled, and then residents and faculty at each hospital serve as the within-case sample) [44]. In other qualitative research methodologies, e.g. in institutional ethnography, the term 'sampling' is a slight misnomer, because the goal of selecting participants or informants for the research is not to report on a particular population's perceptions and experiences, but rather to learn from the informants about the actualities, work processes, and social coordination of a particular phenomenon [104]. However, overall, qualitative research seeks to sample with the aim of achieving a thorough exploration of the study questions.

Often, such thoroughness is referred to as 'saturation', which means that data collection was considered complete when dominant themes/trends were recurrent and no new issues arose from subsequent data collection. For example, if after 10 interviews, including probing and sampling for discrepant instances, the researcher is not hearing anything new on the topic and recurrent themes are similar across interviews, saturation (or as some prefer, sufficiency) is said to have been reached, and data collection may be stopped using this rationale.

Sample estimations may be justified by reference to method-based estimates (e.g. in-depth interviews) [66], sampling strategy (e.g. theoretical sampling) [105], past research findings, or sufficient information power [106]. Given the debate about what constitutes sufficient sample size, Malterud's outlining of sufficient information power offers useful guidance. Malterud suggests five considerations in determining sufficient sample sizes: the aim of the study, sample specificity, use of established theory, quality of dialogue, and analysis strategy. If the study aim is broader, the sample is diverse, theory is not used, the dialogue quality is weak, and analysis spans across cases, then a larger sample is needed because the information richness, or 'power', is lower [106].

Data Collection: Authenticity and Reflexivity

Because the qualitative researcher engages with their research participants in the collection of data, the researcher's role in the construction of meaning must be considered. As part of this, their relation to the participants, and the ways in which that relationship may shape the data that are being collected, requires careful thought both when deciding how to collect the data and when considering constraints on the researcher's interpretation. In educational settings, hierarchical relationships between researchers, who may be medical faculty members, and participants, who may be trainees, can have a distorting effect on the authenticity of the data collection. Participants in vulnerable positions may feel the need to safeguard themselves, to please the researcher, or to advertise their membership in a group. Data collection processes must take such participant motives and actions into account, and researchers must use strategies to maximise the authenticity of their data and reflect on the ways in which the data are a construction of a research relationship in a hierarchical situation. Participant reactivity during data collection was previously discussed, including methods to ameliorate its effects [73, 107]. Another previously discussed strategy used by qualitative researchers to maximise the quality of their data set is 'triangulation' [87]. Triangulation requires the selection of multiple relevant data sources and their integrated analysis, exploring how they confirm or disconfirm one another. It is important to respond to disconfirming instances identified through triangulation, whether this means representing the discrepant instances in one's findings, or seeking out more data and conducting more analysis to better understand and explain the difference.

Data Analysis: Clarity and Audit Trails

Although a challenging task, given its iterative nature, the analysis process in qualitative research should be described such that there is little or no 'mystique' surrounding how the researchers went from numerous transcripts to a list of conceptual or thematic categories. This is not to suggest that there is no 'art' to qualitative analysis; there is, of course, and it includes serendipitous links, just as it may in the analysis of experimental data. However, on the whole, the steps involved in the analysis process can and should be made explicit, both in a published manuscript and in the researchers' own journals. These journals can form the basis of an 'audit trail' to review the analytical journey. Reflexivity can be usefully engaged to articulate what otherwise seem like 'conceptual leaps' [108]. Examples of reflexive questions are shared in Box 29.4.

As suggested earlier, some qualitative researchers would argue that the broad application of proceduralist principles is a sub-optimal way of measuring quality. As Eakin argues, this approach can oversimplify and distort the complex and non-formulaic nature of qualitative inquiry [109]. Instead, the notion of a paper's 'so what' factor – its ability to contribute to the understanding of a social phenomenon – is offered as the most important criterion [109]. Similarly, Sandelowski [100] has proposed a study's 'utility', its power to 'be of use' in the world, as another holistic principle for consideration when evaluating qualitative research. The utility of a study is related to how it is 'written up'. Charmaz [91] and Richardson [90] encourage researchers to attend closely to the aesthetics of the written product of qualitative research in order to maximise understanding and potential impact. These more holistic approaches build upon other principles, such as sampling and authenticity, while trying to avoid the pitfalls of a naïve, checklist approach to quality in qualitative research.

Writing and Sharing the Work

How work is shared also brings with it considerations of rigour and ethics. How much contextual information to share depends on the methodology chosen as well as the protection of participant anonymity and privacy. One could argue that how the work is presented to various audiences (e.g. vulnerable patient populations, medical students, etc.) entails ethical sensitivity, and also integrity to the research findings. Finally, how much participant voice will come through in the writing and sharing of the work also relies on the methodology and methods chosen (if using elicited visual representations, will these be published and have participants consented fully to this?).

The Role of Theory

A final note regarding theory. A beginner qualitative researcher should be aware that theory may play a differing role in the various stages of a qualitative study depending on the methodological approach. That is, depending on the underlying assumptions of a given paradigm and methodology, theory may be more or less involved from the initial phases of developing a research question and designing a study, to the final stages of analysing data and writing up findings. So, how does one know if theory has been appropriately employed? Generally, when assessing the rigour of a qualitative study with regard to the use (too much, or not enough?) of extant theory, the principle of 'best fit' or alignment applies again. Have the authors justified their approach in a cogent manner? If breaking from a methodological tradition or trend with regard to the use of theory, has the break been convincingly explained? If drawing from extant theory, have the authors reproduced more of the same, missing the opportunity to develop new knowledge? A sound understanding of the paradigmatic theories that underlie a chosen methodological approach is beneficial, and most qualitative researchers would argue, obligatory. Debates about the role of theory relate to two main considerations: (i) the interpretation of data through a theoretical lens or frame and (ii) the production of theory through qualitative research.

On the one hand, qualitative research can effectively use theory to inform analysis and interpretation; when this is done well, the research ultimately moves beyond extant theories to produce new ways of thinking. For example, in constructivist approaches to grounded theory, which of course aim to produce theory, 'sensitising concepts' have been proposed to provide a theoretical lens for data analysis [110]. Grounded theory is often cited as an ideal methodology for process-based questions for which there is little extant theory. Forcing data into pre-conceived categories is strongly opposed by classical grounded theorists, who suggest that constructivist approaches legitimate such forcing [111]. Yet, despite such resistance, a methodology is evolving to engage existing theory in the analysis process [112]. The use of 'sensitising concepts' in grounded theory may make way, if controversially, for grounded theorists to expand existing theory or make use of extant theory to understand similar processes in different contexts.

On the other hand, some methodologies aim neither to use theory to guide analysis nor to produce theory as an outcome of the research. For example, descriptive phenomenology aims to remain true to a rich description of the 'essence' of the lived experience of a particular phenomenon [103], and institutional ethnography aims to explicate the 'actualities' of every day work without imposing theory to explain this work, and without producing theory (but rather, enabling social change) from the explication [113].

As an interdisciplinary field, medical education draws from myriad disciplines, which offer countless social theories that need not be completely reinvented. So, at times, drawing from or building upon extant theory can be the 'best fit' for a research purpose. At the same time, a value for theory does not hierarchise theory over description; such a hierarchy may create implicit pressure to claim theory when one has produced description, which in turn may undermine the rigour of some qualitative research [114]. Researchers need to be both thoughtful and transparent about their purposes and procedures with regard to theory building and theory use, in order to advance understanding of medical education through rigorous qualitative research.

Conclusions

Qualitative research has made important contributions to medical education research in the past few decades. This form of inquiry is situated within a particular set of paradigms and draws on recognisable approaches and methodological tools to build knowledge regarding the experiences and activities of teachers, trainees, patients, and team members in medical education settings. Particular rigour and ethics issues must be considered in a qualitative project, with reflexivity engaged throughout to attend to these issues. Used properly, qualitative research promises to offer profound insights into the complex social and human aspects of how health professionals develop their identity, expertise and practice.

References

1 Timmermans, S. (2013). Seven warrants for qualitative health sociology. *Social Science & Medicine* 77: 1–8.

2 Harris, I. (2002). Qualitative methods. In: *International Handbook of Research in Medical Education* (ed. G. Norman, C. van der Vleuten and D. Newble), 45–95. Dordrecht, The Netherlands: Kluwer Academic Publishers.

3 Shea, J.A., Arnold, L., Mann, K.V. et al. (2004). A RIME perspective on the quality and relevance of current and future medical education research. *Academic Medicine* 79: 931–938.

4 Regehr, G. (2010). It's NOT rocket science: rethinking our metaphors for research in health professions education. *Medical Education* 44: 31–39.

5 Denzin, N. and Lincoln, Y. (2000). Introduction: the discipline and practice of qualitative research. In: *Handbook of Qualitative Research* (ed. N. Denzin and Y. Lincoln), 1–28. Thousand Oaks, CA: Sage Publications.

6 Crotty, M. (1998). *Foundations of Social Research: Meaning and Perspective in the Research Process*. Thousand Oaks, CA: Sage Publications.

7 Irby, D.M. (1992). How attending physicians make instructional decisions when conducting teaching rounds. *Academic Medicine* 67: 630–638.

8 Lingard, L., Espin, S., Whyte, S. et al. (2004). Communication failures in the operating room: an observational classification of recurrent types and effects. *Quality & Safety in Health Care* 13: 330–334.

9 Lingard, L., Garwood, S., and Poenaru, D. (2004). Tensions influencing operating room team function: does institutional context make a difference? *Medical Education* 38: 691–699.

10 Lingard, L., Espin, S., Rubin, B. et al. (2005). Getting teams to talk: development and pilot implementation of a checklist to promote safer operating room communication. *Quality & Safety in Health Care* 14: 340–346.

11 Lingard, L., Reznick, R., DeVito, I., and Espin, S. (2002). Forming professional identities on the health care team: discursive constructions of the 'other' in the operating room. *Medical Education* 36: 728–734.

12 Albert, M. (2004). Understanding the debate on medical education research: a sociological perspective. *Academic Medicine* 79: 948–954.

13 Ponterotto, J.G. (2005). Qualitative research in counseling psychology: a primer on research para-digms and philosophy of science. *Journal of Counseling Psychology* 52: 126–136.

14 Creswell, J.W. and Plano Clark, V.L. (2007). *Mixed Methods Research*. Thousand Oaks, CA: Sage.

15 Morse, J., Swanson, J.M., and Kuzel, A.J. (2001). *The Nature of Qualitative Evidence*. Thousand Oaks, CA: Sage Publications, Inc.

16 Schwandt, T.A. (2007). *The Sage Dictionary of Qualitative Inquiry*, 3e. Thousand Oaks, CA: Sage Publications.

17 LeCompte, M. and Schensul, J. (1999). *Designing and Conducting Ethnographic Research. An Ethnographers Toolkit*. Lanham, MD: Alta Mira Press.

18 Hammersley, M. and Atkinson, P. (1995). *What Is Ethnography? Ethnography: Principles in Practice*. London: Routledge.

19 Atkinson, P. and Pugsley, L. (2005). Making sense of ethnography and medical education. *Medical Education* 39: 228–234.

20 Becker, H., Geer, B., Hughs, E., and Strauss, A. (1961). *The Boys in White: Student Culture in Medical School*. Chicago, IL: University of Chicago Press.

21 Bosk, C.L. (2003). *Forgive and Remember: Managing Medical Failure*. Chicago, IL: University of Chicago Press.

22 Mykhalovskiy, E. (2003). Evidence-based medicine: ambivalent reading and the clinical recontextualization of science. *Health* 7: 331–352.

23 Glaser, B. and Strauss, A. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago, IL: Aldine Publishing Company.

24 Meston, C.N. and Ng, S.L. (2012). A grounded theory primer for audiology. *Seminars in Hearing* 33: 135–146.

25 Watling, C. and Lingard, L. (2012). Grounded theory in medical education research: AMEE Guide No. 70. *Medical Teacher* 34: 850–861.

26 Apramian, T., Cristancho, S., Watling, C., and Lingard, L. (2016 Oct 19). (Re) Grounding grounded theory: a close reading of theory in four schools. *Qualitative Research* 17: 359–376.

27 Corbin, J. and Strauss, A. (2008). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, 3e. Los Angeles, CA: Sage Publications Inc.

28 Charmaz, K. (2006). *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*. London: Sage Publications Inc.

29 Clarke, A. (2005). *Situational Analysis: Grounded Theory after the Postmodern Turn*. Thousand Oaks, CA: Sage.

30 Corbin, J. and Strauss, A. (1990). Grounded theory research: procedures, canons, and evaluative criteria. *Qualitative Sociology* 13: 3–21.

31 Kennedy, T.J. and Lingard, L.A. (2006). Making sense of grounded theory in medical education. *Medical Education* 40: 101–108.

32 Dey, I. (1999). *Grounding Grounded Theory: Guidelines for Qualitative Inquiry*. San Diego, CA: Academic Press.

33 Ginsburg, S., Kachan, N., and Lingard, L. (2005). Before the white coat: perceptions of professional lapses in the pre-clerkship. *Medical Education* 39: 12–19.

34 Ginsburg, S., Regehr, G., and Lingard, L. (2004). Basing the evaluation of professionalism on observable behaviors: a cautionary tale. *Academic Medicine* 79: S1–S4.

35 Ginsburg, S., Regehr, G., and Lingard, L. (2003). The disavowed curriculum: understanding students' reasoning in professionally challenging situations. *Journal of General Internal Medicine* 18: 1015–1022.

36 Ginsburg, S., Regehr, G., and Lingard, L. (2003). To be and not to be: the paradox of the emerging professional stance. *Medical Education* 37: 350–357.

37 Ginsburg, S., Regehr, G., Stern, D., and Lingard, L. (2002). The anatomy of the professional lapse: bridging the gap between traditional frameworks and students' perceptions. *Academic Medicine* 77: 516–522.

38 Ginsburg, S., Regehr, G., Hatala, R. et al. (2002). Context, conflict, and resolution: a new conceptual framework for evaluating professionalism. *Academic Medicine* 75: S6–S11.

39 Watling, C., Driessen, E., van der Vleuten, C.P., and Lingard, L. (2012). Learning from clinical work: the roles of learning cues and credibility judgements. *Medical Education* 46: 192–200.

40 Watling, C., Driessen, E., van der Vleuten, C.P. et al. (2012). Understanding responses to feedback: the potential and limitations of regulatory focus theory. *Medical Education* 46 (6): 593–603.

41 Watling, C., Driessen, E., Vleuten, C.P. et al. (2013). Beyond individualism: professional culture and its influence on feedback. *Medical Education* 47 (6): 585–594.

42 Watling, C., Driessen, E., Vleuten, C.P. et al. (2013). Music lessons: revealing medicine's learning culture through a comparison with that of music. *Medical Education* 47 (8): 842–850.

43 Watling, C., Driessen, E., Vleuten, C.P., and Lingard, L. (2014). Learning culture and feedback: an international study of medical athletes and musicians. *Medical Education* 48 (7): 713–723.

44 Stake, R. (2005). Qualitative case studies. In: *The Sage Handbook of Qualitative Research* (ed. N. Denzin and Y. Lincoln), 443–466. Thousand Oaks, CA: Sage Publications.
45 Perley, C.M. (2006). Physician use of the curbside consultation to address information needs: report on a collective case study. *Journal of the Medical Library Association* 94: 137–144.
46 Casimiro, L.M., Hall, P., Kuziemsky, C. et al. (2015). Enhancing patient-engaged teamwork in healthcare: an observational case study. *Journal of Interprofessional Care* 29 (1): 55–61.
47 Creswell, J. (1998). Chapter 4: five qualitative traditions of inquiry. In: *Qualitative Inquiry and Research Design: Choosing among Five Traditions* (ed. J. Creswell), 47–72. Thousand Oaks, CA: Sage Publications.
48 Bearman, M. (2003). Is virtual the same as real? Medical students' experiences of a virtual patient. *Academic Medicine* 78: 538–545.
49 Muller, J. (1999). Narrative approaches to qualitative research in primary care. In: *Doing Qualitative Research* (ed. B. Crabtree and W. Miller), 221–238. Thousand Oaks, CA: Sage Publications.
50 Bleakley, A. (2005). Stories as data, data as stories: making sense of narrative inquiry in clinical education. *Medical Education* 39: 534–540.
51 Pullman, D., Bethune, C., and Duke, P. (2005). Narrative means to humanistic ends. *Teaching and Learning in Medicine* 17: 279–284.
52 Ventres, W. (1994). Hearing the patient's story: exploring physician–patient communication using narrative case reports. *Family Practice Research Journal* 14: 139–147.
53 Bennett, D., Solomon, Y., Bergin, C. et al. (2017). Possibility and agency in Figured Worlds: becoming a 'good doctor'. *Medical Education* 51 (3): 248–257.
54 Thesen, J. and Kuzel, A. (1999). Participatory inquiry. In: *Doing Qualitative Research* (ed. B. Crabtree and W. Miller), 269–290. Thousand Oaks, CA: Sage Publications.
55 Kemmis, S. and McTaggart, R. (2005). Participatory action research: communicative action and the public sphere. In: *The Sage Handbook of Qualitative Research* (ed. N. Denzin and Y. Lincoln), 559–604. Thousand Oaks, CA: Sage Publications.
56 Mowat, H. and Mowat, D. (2001). The value of marginality in a medical school: general practice and curriculum change. *Medical Education* 35: 175–177.
57 Kothari, A. and Wathen, C.N. (2013). A critical second look at integrated knowledge translation. *Health Policy* 109 (2): 187–191.
58 Cameron, D. (2001). Chapter 1: what is discourse and why analyse it? In: *Working with Spoken Discourse* (ed. D. Cameron), 7–18. London: Sage Publications.
59 Mills, S. (2004). *Discourse*, 2e. New York: Routledge.
60 Maynard, D.W. and Heritage, J. (2005). Conversation analysis, doctor–patient interaction and medical communication. *Medical Education* 39: 428–435.
61 Hodges, B.D., Martimianakis, M.A., McNaughton, N., and Whitehead, C. (2014). Medical education … meet Michel Foucault. *Medical Education* 48 (6): 563–571.
62 Hekelman, F.P., Blase, J.R., and Bedinghaus, J. (1996). Discourse analysis of peer coaching in medical education: a case study. *Teaching and Learning in Medicine* 8: 41–47.
63 Varpio, L., Martimianakis, M.T., and Mylopoulos, M. (2015). Qualitative research methodologies: embracing methodological borrowing, shifting and importing. In: *Researching Medical Education*, Chapter 21 (ed. J. Cleland and S. Durning), 245. West Sussex, UK: Wiley.
64 Guillemin, M. and Gillam, L. (2004). Ethics, reflexivity and 'ethically important moments' in research. *Qualitative Inquiry* 10: 261–280.
65 Dicicco-Bloom, B. and Crabtree, B.F. (2006). The qualitative research interview. *Medical Education* 40: 314–321.
66 Miller, W. and Crabtree, B. (1999). Depth interviewing. In: *Doing Qualitative Research* (ed. B. Crabtree and W. Miller), 89–107. Thousand Oaks, CA: Sage Publications.
67 McCracken, G. (1988). *The Long Interview*. Newbury Park, CA: Sage.
68 Reeves, S., Lewin, S., and Zwarenstein, M. (2006). Using qualitative interviews within medical education research: why we must raise the 'quality bar'. *Medical Education* 40: 291–292.
69 Barbour, R.S. and Barbour, R.S. (2005). Making sense of focus groups. *Medical Education* 39: 742–750.
70 Brown, J. (1999). Focus group interviews. In: *Doing Qualitative Research* (ed. B. Crabtree and W. Miller), 109–124. Thousand Oaks, CA: Sage Publications.
71 Hammersley, M. and Atkinson, P. (1995). *Ethnography: Principles in Practice*. London: Routledge.
72 Bogdewic, S. (1999). Participant observation. In: *Doing Qualitative Research* (ed. B. Crabtree and W. Miller), 47–69. Thousand Oaks, CA: Sage Publications.
73 Paradis, E. and Sutkin, G. (2017). Beyond a good story: from Hawthorne effect to reactivity in health professions education research. *Medical Education* 51 (1): 31–39.
74 Bannister, S.L., Hilliard, R.I., Regehr, G., and Lingard, L. (2003). Technical skills in paediatrics: a qualitative study of acquisition, attitudes and assumptions in the neonatal intensive care unit. *Medical Education* 37: 1082–1090.
75 Dubé, T.V., Schinke, R.J., Strasser, R., and Lightfoot, N. (2014). Interviewing in situ: employing the guided walk as a dynamic form of qualitative inquiry. *Medical Education* 48 (11): 1092–1100.
76 Heath, C., Luff, P., and Sanchez, S.M. (2007). Video and qualitative research: analysing medical practice and interaction. *Medical Education* 41: 109–116.
77 Creswell, J. (2003). Chapter 9: qualitative procedures. In: *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (ed. J. Creswell), 179–209. Thousand Oaks, CA: Sage Publications.
78 Olney, C.A., Livingston, J.E., Fisch, S., and Talamantes, M.A. (2006). Becoming better health care providers: outcomes of a primary care service-learning project in medical school. *Journal of Prevention & Intervention in the Community* 32: 133–147.
79 Banks, M. and Zeitlyn, D. (2015). *Visual Methods in Social Research*. Thousand Oaks, California: Sage.
80 Krippendorf, K. (2004). *Content Analysis: An Introduction to Its Methodology*. Thousand Oaks, CA: Sage Publications.
81 Burack, J.H., Irby, D.M., Carline, J.D. et al. (1997). A study of medical students' specialty-choice pathways: trying on possible selves. *Academic Medicine* 72: 534–541.
82 Drew, S. and Guillemin, M. (2014). From photographs to findings: visual meaning-making and interpretive engagement in the analysis of participant-generated images. *Visual Studies* 29 (1): 54–57.
83 Cristancho, S., Bidinosti, S., Lingard, L. et al. (2015). Seeing in different ways: introducing 'rich pictures' in the study of expert judgment. *Qualitative Health Research* 25 (5): 713–725.
84 Liebenberg, L., Didkowsky, M., and Ungar, M. (2012). Analysing image-based data using grounded theory: the Negotiating Resilience Project. *Visual Studies* 27 (1): 59–74.
85 Gleeson, K. (2011). Polytextual thematic analysis for visual data. In: *Visual Methods in Psychology* (ed. P. Reavey), 314–329. Hove: Psychology Press.
86 Cristancho, S., Lingard, L., Forbes, T. et al. (2017). Putting the puzzle together: the role of 'problem definition' in complex clinical judgement. *Medical Education* 51 (2): 207–214.
87 Flick, U. (2004). Triangulation in qualitative research. In: *A Companion to Qualitative Research*, 178–183. Thousand Oaks, CA: Sage Publications.
88 Lingard, L., Schryer, C.F., Spafford, M.M., and Campbell, S.L. (2007). Negotiating the politics of identity in an interdisciplinary research team. *Qualitative Research* 7: 501.
89 Decuir-Gunby, J.T., Marshall, P.L., and McCulloch, A.W. Developing and using a codebook for the analysis of interview data: an example from a professional development research project. *Field Methods* 23 (2): 136–155.

90 Richardson, L. (2004). Writing: a method of inquiry. In: *Approaches to Qualitative Research: A Reader on Theory and Practice* (ed. S. Nagy Hesse-Biber and P. Leavy), 473–495. New York: Oxford University Press.
91 Charmaz, K. (2004). Premises, principles, and practices in qualitative research: revisiting the foundations. *Qualitative Health Research* 14: 976–993.
92 Lingard, L. and Watling, C. (2016). It's a story, not a study: writing an effective research paper. *Academic Medicine* 91 (12): e12.
93 Mays, N. and Pope, C. (2000). Qualitative research in health care. Assessing quality in qualitative research. *British Medical Journal* 320: 50–52.
94 Mays, N. and Pope, C. (1995). Rigour and qualitative research. *British Medical Journal* 311: 109–112.
95 Inui, T.S. and Frankel, R.M. (1991). Evaluating the quality of qualitative research: a proposal pro tem. *Journal of General Internal Medicine* 6: 485–486.
96 Rowan, M. and Huston, P. (1997). Qualitative research articles: information for authors and peer reviewers. *Canadian Medical Association Journal* 157: 1442–1446.
97 O'Brien, B.C., Harris, I.B., Beckman, T.J. et al. (2014). Standards for reporting qualitative research: a synthesis of recommendations. *Academic Medicine* 89 (9): 1245–1251.
98 Lincoln, Y. and Guba, E.G. (1985). *Naturalistic Inquiry*. Thousand Oaks, CA: Sage Publications.
99 Sandelowski, M. (2004). Using qualitative research. *Qualitative Health Research* 14: 1366–1386.
100 Sandelowski, M. (1997). 'To be of use': enhancing the utility of qualitative research. *Nursing Outlook* 45: 125–132.
101 Lincoln, Y. and Guba, E.G. (2000). Paradigmatic controversies, contradictions and emerging confluences. In: *Handbook of Qualitative Research* (ed. N. Denzin and Y. Lincoln), 163–188. Thousand Oaks, CA: Sage Publications.
102 Morse, J.M. (2015). Critical analysis of strategies for determining rigor in qualitative inquiry. *Qualitative Health Research* 25 (9): 1212–1222.
103 van Manen, M. (1997). *Researching Lived Experience: Human Science for an Action Sensitive Pedagogy*. London, ON: Althouse Press.
104 Campbell, M.L. and Gregor, F.M. (2002). *Mapping Social Relations: A Primer in Doing Institutional Ethnography*. Aurora, Canada: Rowman Altamira.
105 Patton, M. (1990). *Qualitative Evaluation and Research Methods*. Beverly Hills, CA: Sage.
106 Malterud, K., Siersma, V.D., and Guassora, A.D. (2016). Sample size in qualitative interview studies: guided by information power. *Qualitative Health Research* 26 (13): 1753–1760.
107 Holden, J.D. (2001). Hawthorne effects and research into professional practice. *Journal of Evaluation in Clinical Practice* 7: 65–70.
108 Klag, M. and Langley, A. (2012). Approaching the conceptual leap in qualitative research. *International Journal of Management Reviews* 15 (2): 149–166. doi: 10.1111/j.1468-2370.2012.00349.x.
109 Eakin, J.M. and Mykhalovskiy, E. (2003). Reframing the evaluation of qualitative health research: reflections on a review of appraisal guidelines in the health sciences. *Journal of Evaluation in Clinical Practice* 9: 187–194.
110 Bowen, G.A. (2006). Grounded theory and sensitizing concepts. *International Journal of Qualitative Methods* 5: Article 2. http://www.ualberta.ca/~iiqm/backissues/5_3/pdf/bowen.pdf (accessed 15 January 2012).
111 Glaser, B.G. (2002). Constructivist grounded theory? *Forum, Qualitative Sozialforschung/Forum, Qualitative Social Research* 3: Article 12. http://www.qualitative-research.net/index.php/fqs/article/view/825/1792 (accessed 15 January 2012).
112 Morse, J., Stern, P., and Corbin, J. (2008). *Developing Grounded Theory: The Second Generation*. Walnut Creek, CA: Left Coast Press.
113 Smith, D. ed. (2006). *Institutional Ethnography as Practice*. Lanham, MD: Rowman & Littlefield Publishers, Inc.
114 Ng, S.L., Bisaillon, L., and Webster, F. (2017). Blurring the boundaries: using institutional ethnography to inquire into health professions education and practice. *Medical Education* 51 (1): 51–60.

Further Reading

Baker, L., Phelan, S., Snelgrove, R. et al. (2016). Recognising and responding to ethically important moments in qualitative research. *Journal of Graduate Medical Education* 8 (4): 607–608.

Bergman, E., de Feijter, J., Frambach, J. et al. (2012). AM last page: a guide to research paradigms relevant to medical education. *Academic Medicine* 87 (4): 545.

Britten, N. (2005). Making sense of qualitative research: a new series. *Medical Education* 39: 5–6.

Crabtree, B. and Miller, W. ed. (1999). *Doing Qualitative Research*, 2e. Thousand Oaks, CA: Sage Publications.

Creswell, J. (2012). *Qualitative Inquiry and Research Design: Choosing among Five Traditions*, 3e. Thousand Oaks, CA: Sage Publications.

Crotty, M. (1998). *Foundations of Social Research: Meaning and Perspective in the Research Process*. Thousand Oaks, CA: Sage Publications.

Denzin, N. and Lincoln, Y. ed. (2017). *The Sage Handbook of Qualitative Research*, 5e. Thousand Oaks, CA: Sage Publications.

Miles, M.B. and Huberman, A.M. (2013). *Qualitative Data Analysis: A Sourcebook of New Methods*, 3e. Thousand Oaks, CA: Sage Publications.

Schwandt, T.A. (2007). *The Sage Dictionary of Qualitative Inquiry*, 3e. Thousand Oaks, CA: Sage Publications.

30 Programme Evaluation

Chris Lovato[1] *and Linda Peterson*[2]

[1]School of Population & Public Health and Evaluation Studies Unit, Faculty of Medicine, University of British Columbia, Vancouver, British Columbia, Canada

[2]Retired, Faculty of Medicine, University of British Columbia, Vancouver, British Columbia, Canada

KEY MESSAGES

- Programme evaluation focuses on questions related to whether a programme is working as intended and if there are any unintended consequences.
- Evidence from programme evaluation is essential to enhance professional practice and to achieve the best medical education for students, trainees, and doctors engaged in continuing professional development.
- There are important similarities and differences between research and evaluation.
- Evaluators have the same obligations as researchers in considering the ethical issues involved in implementing studies.
- The value of an evaluation rests on whether the information is useful. There are methods and techniques that can enhance the utility of a study.
- High-quality evaluation in medical education will ultimately contribute to delivering training that will ultimately produce quality patient care and a healthy population.

Introduction

This chapter covers the wide role of programme evaluation in medical education from micro to macro, from the evaluation of individual teaching episodes to entire curricula, for the purposes of improving pedagogy to influencing national policy.

The chapter is divided into four sections. We begin with some definitions and a discussion of the purposes of evaluation and the symbiotic relationship between programme planning and evaluation. Some key concepts are described and selected models and theories of evaluation are presented. This first section concludes by exploring the differences between evaluation and research. The second section focuses on evaluation practice. It describes ethics in evaluation, evaluation methods, sources of evidence, and promoting the use of evaluation findings. The third section addresses ways in which an evaluation can be implemented to promote its use and the role of evaluation in change management. The chapter concludes with three examples of evaluation in medical education and selected additional resources.

What is Programme Evaluation?

Programme evaluation focuses on questions related to whether a programme is working as intended and if there are any unintended consequences. There are many different definitions of evaluation in the literature. In this chapter we use an adapted version of the definition provided by Fink in her text on evaluation 'fundamentals' [1]:

> Program evaluation is the diligent investigation of a program's characteristics and merits. In the context of health care, the purpose of program evaluation is to provide information about the effectiveness of programs, so as to optimize the outcomes, efficiency and quality of health care. An evaluation may analyse a program's structure, activities, and organization as well as its political and social environment. It may also appraise the achievement of the program's goals and objectives and the extent of the program's impact and costs.

The term 'program' (or programme) can refer to any organised action such as a curriculum, a course, session, student service, event, guidelines, or a policy in medical education. The reasons for conducting an evaluation are varied and can include the desire to improve the implementation and effectiveness of a programme, manage limited resources, justify funding, support the need for increased funding, document social accountability, and meet requirements for academic standards or accreditation. Box 30.1 shows three common approaches to evaluation in medical education. This chapter will focus on the first and second approach. The third approach is normally an externally driven process such as an accreditation review process.

Evaluation is much broader than merely handing out satisfaction surveys to students and trainees at the end of teaching sessions. Evaluation is vital for curriculum development and in determining if the curriculum is operating

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

BOX 30.1 Common approaches to evaluation in medical education [2]

Decision-oriented approach: The evaluation results help programme personnel make effective decisions. The type of data included in, the research design of, and the focus of the evaluation are selected to maximise the evaluators' utilisation of results.

Outcomes-oriented approach: Objectives are solidified so that specific outcome measures can be established and tracked. The evaluation determines whether the programme objectives have been met.

Expertise-oriented approach: The evaluator relies on an external expert to determine the value of various programme criteria and data points, and the programme evaluation results are judged by an expert.

BOX 30.2 Examples of evaluation questions

- Is there comparability between regional training sites?
- What is the educational climate like for medical students in the operating theatre?
- What are the facilitators and barriers to implementing the new integrated clerkship?
- What aspects of the faculty development course had a positive impact on teaching?
- How reliable was the shortlisting and interviewing for paediatric trainees?
- What do students and residents think about career advice provided by support services?
- What is the practice location and specialty of trainees?

as intended and achieving the intended outcomes. Did the learners achieve the targeted knowledge and skills from the teaching programme? Evaluation may also be used to ensure that supporting programmes and services are meeting users' needs. It is often used to identify areas where the curriculum needs to improve. It is used to determine if an educational programme is of an acceptable standard and may be approved for training and accreditation purposes. It may be used to give feedback to instructors, administrators, managers, and faculty on a broad range of services (e.g. library services, technology, admissions, and assessment) that support medical education. It may be used as part of the information presented at annual appraisals for medical teachers, and for promotion and career development. In terms of assessments, it may be used to gather outcome measures on pass rates for qualifying and professional examinations. And it can be used to determine long-term outcomes such as specialty choice and location of practice in rural, remote, and under-served regions.

In addition, evaluation may be used to determine future educational policy in a curriculum, teaching and learning, or assessment. It may also be used as a tool to implement centrally determined policy through a number of covert and controlling processes. Box 30.2 provides examples of the kinds of questions an evaluation might focus upon.

The practice of programme evaluation involves applying theory, research findings, and the most rigorous methods possible to a real-world setting in order to address practical questions relevant to decision makers and stakeholders. The tricky part is that evaluators working in medical education programme settings often do not have the same kind of control over study conditions as researchers might.

Definitions

In everyday life the terms *evaluation*, *assessment*, and *appraisal* are often used interchangeably. This confusion is compounded by international differences in definitions. In North America, for instance, the word 'evaluation' is sometimes equated with the UK term 'assessment', to mean measurement of learners' skills [3]. For example the mini-clinical 'evaluation' exercise is actually an 'assessment' tool for testing junior doctors' history-taking and examination skills [4].

In this chapter, *assessment* is defined as 'the processes and instruments applied to measure the learner's achievements, normally after they have worked through a learning programme of one sort or another' [5]. Assessment then is about testing the learners. *Appraisal* is 'a two-way dialogue focusing on the personal, professional and educational needs of the parties, which produces agreed outcomes' [6, 7]. As noted earlier in this section, *evaluation* focuses on the design, implementation, improvement, or outcomes of a programme rather than focusing on the assessment of individual or individuals.

Programme Planning and Evaluation

Programme planning and evaluation are highly interrelated. If the programme plan is not well-developed and lacks clear goals and objectives, it is difficult and often impossible to carry out a credible evaluation. In developing a new programme it is important to identify goals and objectives that are measureable or 'evaluable'. Over the life of the programme, planning and evaluation are both part of a continuous cycle of ongoing improvement. The methods and approaches used in programme planning and evaluation occur throughout the life cycle of a programme, including assessing needs, modifying approaches, identifying indicators and measures, determining effectiveness, identifying facilitators and barriers to implementation, and making recommendations for improvement. In practice, the process and methods used for curriculum development are the same for evaluation: for example, developing a programme description, specifying a target process and outcomes, identifying or developing measures, designing and collecting data, and disseminating results. Whether you are engaging in planning or evaluating, these processes involve the use of theory, research findings, and the most rigorous methods possible in a medical education setting.

Evaluation or Research?

Evaluation and research in education are similar activities, and share many of the same methods. Programme evaluation is a systematic method for collecting, analysing, and using information to answer questions about projects, policies, and programmes, particularly about their effectiveness and efficiency. It is about providing practice-based evidence from the real world to address questions that are important to stakeholders, including funders, programme planners, implementers, decision makers, and consumers. The focus is upon stakeholder-generated questions versus questions that arise from theory, the literature, or researcher curiosity. Evaluators address study questions that facilitate evidence-based decision-making and accountability. The work is often ongoing and cyclical, focusing on continuous improvement of programmes.

The difference in perspective that evaluation brings to the study of problems has to do with intent. The intent of evaluation is to identify questions that are meaningful for making evidence-based decisions and establishing accountability. Those questions may or may not address a gap in the literature. The driving questions are always specific to the local context of the programme. This differs from research where the intent of the researcher is to undertake work that will contribute to a larger body of knowledge (i.e. the scientific literature). The questions addressed in research are curiosity driven and typically arise from previous research or theory.

Evaluation is 'methods-neutral' within the broad domain of social science methods. Evidence can be gathered based on experimental, quasi-experimental, and observational designs. Similar to research, evaluation uses qualitative, quantitative, and mixed-methods; and evaluation and research use the same principles of design, data collection, and analysis. However, when it comes to report writing and dissemination, an evaluative perspective is different from a research perspective in that evaluators use multiple forms of reporting and results that will not necessarily be published. Providing specific actionable recommendations is a focus in evaluation, and facilitating use of evaluation results is part of the role of an evaluator.

With the growing interest in knowledge translation research, there has been a greater focus on the knowledge generated from evaluation studies. For the most part, this has been driven by the commitment to evidence-based practice and the strong call for better links between research and practice. More research-funding agencies are encouraging researchers to address evaluation questions in their studies and to focus on questions related to programme implementation, adaptation of approaches, impact evaluation, and knowledge use. By publishing their work on applying theories, methods, and evidence, evaluators are in an excellent position to contribute to the knowledge base in medical education.

Theory in Evaluation

In the literature on programme evaluation, 'theory' is often used in two different ways. Theory may refer to an evaluative approach, conceptual model, or theory of practice. Alternatively, 'theory-driven' evaluation refers to an evaluation study that is based on the programme's 'theory of change', which is most often represented as a logic model. We will consider these in turn.

Evaluation Models and Approaches

There are a variety of theories, models, and approaches described in the evaluation literature. Typically, they differ based upon who is involved in the evaluation, what is evaluated, and why and how a study is conducted. In the majority of evaluation studies these approaches are blended. As many as 13 different models and approaches to evaluation have been identified [8]. Examples of those most relevant to medical education include an objectives approach, expertise-accreditation approaches, utilisation-focused approaches, participatory and collaborative approaches, and organisation learning. Although there are many different models of evaluation described in the literature, we describe four widely used models that are well suited to medical education: utilisation-focused evaluation; Kirkpatrick's hierarchy; the context-inputs-process-products (CIPP) model, and participatory, collaborative, and empowerment evaluation.

Utilisation-focused Evaluation

Utilisation-focused evaluation is an approach associated with Michael Quinn Patton. This approach is decision oriented and is based on the premise that 'an evaluation should be judged according to its utility and actual use' [9]. The evaluator focuses on developing and implementing an evaluation that places 'use' as the primary consideration in how the evaluation is planned, is implemented, and findings are reported. 'The focus in utilization-focused evaluation is on intended use by intended users' [9].

In utilisation-focused evaluation, the evaluator works with decision makers to design the evaluation. There is no particular evaluation method, approach, or model associated with utilisation-focused evaluation. The assumption is that the most appropriate method, approach, or model will be based on the needs of the primary intended users. A utilisation-based approach can be used in all types of evaluations (e.g. formative, summative, outcome, etc.) with both qualitative and quantitative data, and any type of design (e.g. experimental, quasi-experimental, and any qualitative design).

An important strength of this approach is that it increases the use of evaluation results. A potential downside to using this approach is that by focusing the evaluation primarily on the interests of the intended users of the results, the viewpoint of the programme's target population could be overlooked [10]. Vassar et al. [11] encourage the application of utilisation-focused evaluation in medical education settings because it is a flexible and pragmatic approach to answering a wide variety of programmatic questions. It also actively includes key decision makers, thus making it more likely that the results will be used.

Kirkpatrick's Hierarchy

One of the most widely applied evaluation approaches in medical education has been Kirkpatrick's hierarchy, which was first described in articles published by Donald Kirkpatrick in 1959 which were based on his dissertation in the field of training and development. He proposed a series of levels of evaluation on which to focus questions about the effectiveness of training. At the base (the lowest level) of the pyramid model is some indication of satisfaction with the teaching and learning. Next up the pyramid is a concern for what learning has taken place, followed by an indication of behavioural change. The apex of the pyramid focused on the impact of an intervention on society or a community [12]. Each level represents a legitimate area of inquiry for evaluating a programme.

Kirkpatrick's hierarchy has been adapted and widely used in medical education, but not without criticism regarding its limitations. Unfortunately, most educational evaluations have focused on the lower levels related to satisfaction and learning, with few evaluating long-term impact questions. Belfield et al. [13], for instance, found that in a study of 305 papers, only 1.6% had looked at health care outcomes. Parker [14] argued that the model does not address changes in the field of evaluation which have occurred since the model was first developed. He suggests that, while the Kirkpatrick model can be used effectively to identify achievement of objectives and how a programme can be changed to achieve its intended outcomes, it does not sufficiently address context, process, and the needs of stakeholders.

Kirkpatrick's daughter and son-in-law have recently proposed the New World Kirkpatrick model [15], which attempts to address previously cited concerns. Figure 30.1 illustrates the kinds of questions asked at each level of evaluation. While maintaining the original levels described in the pyramid, they suggest the levels of evaluation represent a chain of evidence (versus a causal chain) that can be used non-sequentially, concurrently, or in reverse order. The new model acknowledges the importance of working closely with stakeholders during programme planning in order to identify indicators and incorporate mechanisms for data collection that provide the foundation for an evaluation. They have also expanded the original model to address the context of training and the need to explore both intended and unintended outcomes. Finally, while the apex (Level 4) of the original Kirkpatrick model focused on the broad impact of a programme on society, the New World version infers that Level 4 should focus on the contribution that training makes to the organisation's goals [15, 16].

CIPP (Context, Input, Process, Products)

The CIPP model is a comprehensive practical approach that incorporates needs assessment, formative evaluation, and summative evaluation [17–19]. A basic principle underlying CIPP is that the evaluation's most important purpose is not to prove but to improve. It was developed to enable evaluators working in the 'real world', where experimental designs are difficult to employ, by guiding service providers to identify the strengths and weakness of a programme and address the need for accountability. This model has been widely applied and is frequently used in education. It is particularly suited to guiding internal evaluation. 'CIPP' refers to context, inputs, processes, and products. Box 30.3 provides examples of the different types of questions the model

> **BOX 30.3 CIPP Model: Evaluation questions [18]**
>
> - **C**ontext: What needs to be done?
> - **I**nput: How should it be done?
> - **P**rocess: Is it being done?
> - **P**roducts: Is it successful?

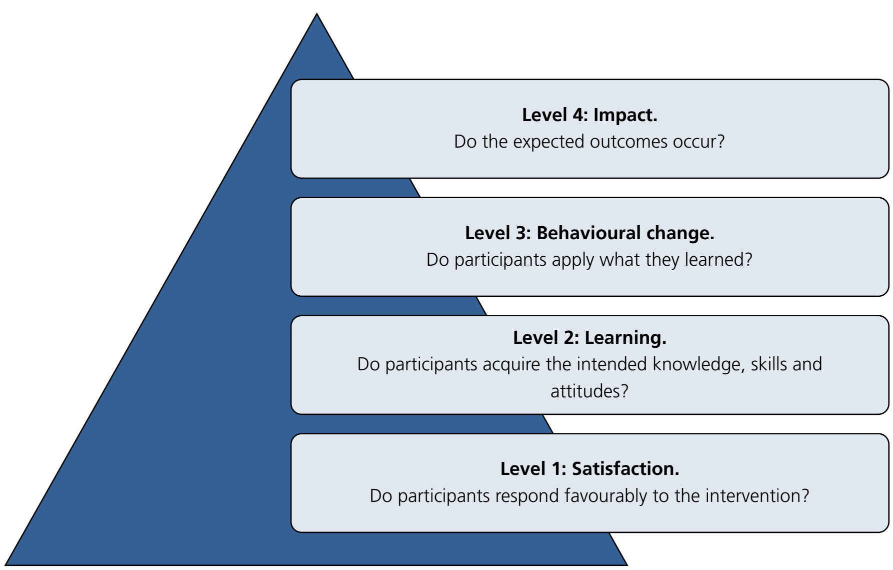


Figure 30.1 Kirkpatrick's hierarchy of evaluation.

focuses upon. Context evaluations address questions related to needs, problems, strengths, and opportunities as the basis for setting programme goals and objectives. Input evaluations examine different approaches to addressing needs and specify a plan for the resources that will be required to achieve goals and objectives. Process evaluation focuses on asking questions about implementation of the intervention and product evaluation assesses outcomes – both intended and unintended – that result from the intervention.

Participatory, Collaborative, and Empowerment Evaluation

Participatory, collaborative, and empowerment all refer to the involvement of those who have a stake in the programme, including funders, policy makers, students, faculty, staff, and members of the community. Participatory evaluation methods are based on the foundations of community-based participatory research and participatory action research. While the level of participation can vary across studies using this approach, the focus is upon valuing and using the knowledge and expertise of those involved with or benefiting from the programme. Participation can include involving stakeholders in identifying evaluation questions, developing indicators and measures, collecting data, analysing and interpreting data, and disseminating results of the evaluation. In general, there has been an increased use of collaborative methods in evaluating education and social programmes, with the movement toward wider accountability to government, citizens, and students themselves.

There are many advantages to using collaborative approaches, including the empowerment of those involved, building evaluation capacity, and reinforcing organisational learning [20]. However, participatory evaluation is not always appropriate for every evaluation. There are a number of constraints to be considered, such as the cost and time of involving individuals with a broad range of experience and understanding of evaluation. In addition, the process can sometimes be unpredictable, requiring an experienced evaluator who can negotiate the process. An important advantage of involving stakeholders in the evaluation is the significant increase in the chances that the findings will be used.

Fetterman et al. [21] describes the use of empowerment approaches when evaluating and transforming a medical school curriculum. Five tools are outlined as being central to implementing the approach: (i) developing a culture of evidence; (ii) using a critical friend; (iii) encouraging a cycle of reflection and action; (iv) cultivating a community of learners; and (v) developing reflective practitioners. Fetterman et al. [21] report that the application of these methods during a curriculum reform process fostered greater institutional self-reflection, an evidence-based model of decision-making, and expanded opportunities for collaboration among faculty, students, and staff.

Logic Models and Theories of Change

An important tool for conducting theory-driven evaluation is logic modelling [22]. A logic model provides a concise graphic representation that communicates the purpose of a programme, its components, and the sequence of activities and outcomes anticipated. In effect it is a causal model, or theory of change, that links inputs, activities, outputs, and outcomes. The basic components of a logic model are shown in Figure 30.2. Inputs are defined as those resources dedicated to or consumed by the programme (e.g. money, staff, faculty, facilities, equipment), activities are the tactical actions (e.g. curriculum, support services) that occur to achieve the objectives of the programme, and outputs are what the programme does with the inputs to fulfil its mission (e.g. admit students to a programme, deliver courses, provide clinical training, provide student support services). Outcomes refer to the benefits for students during and after their training (e.g. knowledge, skills, licensing, and practice) and they are often specified as short-term, intermediate, and long-term [23]. Impacts are the even longer system changes that are anticipated to result, such as wider societal benefit.

When you begin the process of developing a logic model (see Box 30.4), assume it will require multiple versions

BOX 30.4 HOW TO: Develop a logic model

1. Specify the need, or problem, the programme is addressing to help define the outcomes
2. Identify major programme resources/inputs necessary to run the programme
3. Define the key activities that comprise the programme
4. Determine the programme outputs or actions that will be implemented
5. Specify the programme short-term, intermediate, and long-term outcomes expected
6. Identify the longer term fundamental system change or impact anticipated
7. Consider the external factors that will influence outputs and outcomes
8. Construct a draft logic model and refine with collaborators

Figure 30.2 The basic components of a logic model.

before finalised. Logic modelling is a tool. The process facilitates building consensus about a programme, identifying of limits and logical gaps, and whether it is realistic to expect the effects anticipated with the programme activities that will be implemented. It also facilitates the identification of outcomes and indicators and helps determine the boundaries of evaluation questions. For example, some programmes may examine evaluation questions that focus on outputs and short-term outcomes during a pilot phase, with an eye toward measuring longer term outcomes once the programme is stabilised.

Ethics in Programme Evaluation

Evaluators have the same obligations as researchers in considering the ethical issues involved in implementing studies. However, the overlap between evaluation and research often causes confusion as to whether formal approval by an ethics review board or committee is required. Bedward et al. [24], in their editorial in *Medical Teacher* in 2005, discussed the lack of clarity on what requires ethical approval, the reliance on one procedure for all types of research applications, confusion over the scope of responsibilities, and the scale of the work involved.

When participants give their informed consent to participate in a programme evaluation, this means that they know the risks and benefits of participating in the study and what it will involve. They also agree to the terms of participation and know their rights as research participants. While programme evaluation is sometimes exempt from formal ethical review, participants still need to provide informed consent. In some circumstances it may not be necessary to distribute consent forms to participants. For example, it may be sufficient to include a statement on a survey that alerts students that the information will be used for the purposes of programme evaluation. The use of administrative data (e.g. learning analytics, assessment data, clerkship, or residency placement information), leads to ethical concerns that need to be considered, ranging from the confidentiality of students whose information is being used to the nature of the study and its potential impact on participants. In any case, it is important to check with your institution's policies regarding collection and use of data for purposes of programme evaluation.

Institutions don't always have the same requirements on when an evaluation project is or is not research. However, the primary determining factor is primarily related to who will have access to the results and in what form. In general, if there is intent to publish or otherwise disseminate findings as a contribution to knowledge, then ethical review is likely to be required. This is an important consideration for many evaluators working in medical education since the redefinition of scholarship by Boyer [25] emphasises the educational environment as a setting for academic inquiry. On the other hand, if the evaluation will be communicated to an internal committee and will not be used outside your institution, it may not be considered research because you are not seeking to generalise the knowledge. Regardless, it is highly recommended that medical education evaluators seek advice from their own institution, as there is evidence that considerable variability exists in requirements for medical education research studies [26].

Among the tips Egan-Lee et al. [27] provide for obtaining ethical approval for research in health professions education are examining your intent, planning for early communication with your local ethics review system (e.g. Institutional Review Board, or Research Ethics Committee), and determining if your study is exempt from review. As they note, many institutions exempt programme evaluation studies from requiring approval but requirements for approval will vary depending on your national context.

Morrison [28] describes ethical issues in evaluation practice as rising at any stage in the evaluation process, but most commonly during the entry and design phase, communication of results, and utilisation of findings. Examples of issues include:

- stakeholders who have already decided what the findings 'should be'
- different expectations or purposes for the evaluation as viewed by diverse stakeholders
- leaving out certain stakeholder voices from the evaluation
- pressure by stakeholders to alter the presentation of findings
- stakeholders who suppress or ignore findings
- stakeholders who misinterpret the findings.

Goldie [3] cites seven ethical standards for evaluators, drawn from a number of national bodies by Worthen et al. [29] (see Box 30.5). Ultimately though, it is the individual evaluator's responsibility to work ethically to bring to their work a principles-based or virtue-based approach rather than to merely follow external policies and procedures.

As in other professions, evaluators have developed standards and guidelines for practice that are informative in both anticipating and dealing with ethical issues that may arise in evaluation practice. Box 30.6 provides a synopsis and internet links for the UK Evaluation Society and the American Evaluation Association guidelines for practice.

Evaluation Methods

The evaluation design that you use will depend on the nature of question(s) that you are asking. Are the questions focused on the programme process, implementation,

BOX 30.5 Ethical standards in evaluation [3, 29]

- Service orientation
- Formal agreements
- Rights of human subjects
- Complete and fair assessment
- Disclosure of findings
- Conflicts of interest
- Fiscal responsibility

BOX 30.6 FOCUS ON: Good practice in evaluation

A number of organisations have defined standards for good evaluation practice. The European Evaluation Society lists the standards laid out by a number of European countries and can be found at http://www.europeanevaluation.org/resources/evaluation-standards.

The UK version spells out what constitutes good practice for:

- evaluators
- participants
- commissioners
- institutions conducting self-evaluation.

The UK Evaluation Society Guidelines for Good Practice in Evaluation are available at www.evaluation.org.uk.

The American Evaluation Association has also produced guiding principles. These are built around five areas:

1. *Systematic inquiry:* Evaluators conduct systematic, data-based inquiries.
2. *Competence:* Evaluators provide competent performance to stakeholders.
3. *Integrity/honesty:* Evaluators display honesty and integrity in their own behaviour, and attempt to ensure the honesty and integrity of the entire evaluation process.
4. *Respect for people:* Evaluators respect the security, dignity, and self-worth of the respondents, programme participants, clients, and other evaluation stakeholders.
5. *Responsibilities for general and public welfare:* Evaluators articulate and take into account the diversity of general and public interests and values that may be related to the evaluation.

The full guide can be downloaded from http://www.eval.org/Publications/GuidingPrinciples.asp.

outputs, outcomes, or impact? Is the purpose of your evaluation to monitor progress or demonstrate attribution? Are quantitative or qualitative methods better suited to addressing the questions of interest? A comprehensive evaluation that examines the range of evaluation questions related to process, implementation, and outcome will commonly require both numbers and narratives. Quantitative methods are best suited for addressing questions related to causality or attribution and the magnitude of change, while qualitative methods rest in people's experience. Mixing these two research approaches gives the evaluator a broader understanding of the programme and offsets the weaknesses associated with each approach. An important advantage of using both approaches is the possibility of triangulation, that is, using several means to study the same phenomenon.

Chapter 29 of this book provides a comprehensive overview of qualitative methods, including the different research paradigms and methods used in addressing qualitative research questions which will be useful to readers. Typically these questions focus on examining 'how' or 'what'. For example, 'how' do students experience a programme or policy or 'what' does being a peer mentor mean to students.

Evaluations focusing on outcomes and using quantitative methods commonly use experimental, quasi-experimental, and observational designs. An experimental design uses random assignment to compare outcomes in 'intervention' and 'non-intervention' group(s), while a quasi-experimental design compares 'intervention' and non-equivalent 'comparison' group(s) that are not randomly assigned. While the most powerful design for establishing causality is the randomised control trial, it remains less common in medical education for practical reasons. Chapter 28 of this book provides an extensive description about the range of quantitative research designs available to medical educational researchers, including experimental and quasi-experimental designs. As the authors note, these designs are more typically used in clinical research, but are not always relevant to the kinds of questions researchers in medical education want to address.

In many cases it will not be feasible to implement any of the designs described in Chapter 28. Contribution analysis, developed by Mayne [30, 31], is an alternative to establishing cause and effect. This approach explores attribution by assessing the contribution a programme is making to the outcomes that are observed. The steps for conducting contribution analysis are listed in Box 30.7. Causality is inferred by taking into consideration both the theory of change underlying the programme and other factors that might influence outcomes.

BOX 30.7 HOW TO: Conduct a contribution analysis [32]

Step 1: Set out the problem to be addressed
Step 2: Develop a theory of change and risks to it
Step 3: Gather the existing evidence on the theory of change
Step 4: Assemble and assess the contribution story and challenges to it
Step 5: Seek out additional evidence
Step 6: Revise and strengthen the contribution story

Causality is inferred from the following evidence:

1 The programme is based on a reasoned theory of change: the assumptions behind why the programme is expected to work are sound, are plausible, and are agreed upon by at least some of the key players.
2 The activities of the programme were implemented.
3 The theory of change is verified by evidence: the chain of expected results occurred.
4 Other factors influencing the programme were assessed and were either shown not to have made a significant contribution or, if they did, the relative contribution was recognised [32].

It should be noted that in cases where the purpose of the evaluation is accountability alone, it may be acceptable to use a goal-based evaluation model, by using predetermined goals and objectives as the standards of comparison for the evaluation. In these cases, demonstrating attribution is not the primary purpose. For example, you may need to show that a particular programme was delivered or that certain objectives have been achieved.

Data Collection

This section includes brief descriptions of the common methods used in programme evaluation, including questionnaires, interviews, focus groups, site visits, administrative records, and group methods.

Several methods may be used serially, e.g. when a focus group is used to generate items for a questionnaire. Alternatively, methods may be used in parallel where multiple methods are adopted to tap into different data sources in order to build the richest possible picture of the educational initiative under study.

Questionnaires

There are several advantages of using a questionnaire for evaluation purposes. The basic concepts underlying the psychometrics of measurement instruments such as questionnaires are discussed in Chapter 28 of this book and will not be repeated here. Questionnaires are feasible and economical in terms of the time and effort to collect information for an evaluation and are commonly used for this purpose. Using a survey that has been validated and published is the ideal choice, not only because it will save you time and money, but also because it will allow you to compare your results with those from other evaluations. Whenever possible, rely on previously validated questionnaires. If you use a standardised survey that has been administered in a context that is different from your own, you will need to consider adapting and pre-testing it to ensure it is relevant to your setting. If you develop your own survey we recommend that you consult the resources referenced at the end of this chapter.

Be aware that there are some disadvantages to the questionnaire method. For example, there is a well-recognised problem with pre-coded responses, which may not be sufficiently comprehensive to accommodate all answers, forcing participants to choose a view that does not represent their true perceptions [33]. We make assumptions that all respondents will understand the questions in the same way, and there is no way of clarifying the question as in a one-to-one interview. Non-response affects the quality of the data and thus the generalisability of the results. To help overcome these challenges, an evaluation questionnaire can be designed to include open-ended questions, closed questions (e.g. yes/no), tick box questions with specified categories, scaled items (e.g. Likert rating scale with strongly agree to strongly disagree), and the opportunity for participant comments to catch any other relevant information that may not otherwise be collected.

Individual Interviews

Cohen et al. [34] defined a research interview as a 'two-person conversation initiated by the interviewer, for the purpose of gathering research relevant information'. The interview has multiple uses within educational evaluation and research. It can be used:

- to gather information about the evaluation questions to ask
- to develop ideas for new hypotheses or research questions
- as a primary source of data or in conjunction with other evaluation methods
- to validate results from a study
- to go deeper and explore new themes generated from other evaluation methods
- to test hypotheses that have already been generated [35].

Also, the interview method is a powerful way to gain internal validity in case study work, to go deeper and explore new themes generated from other educational evaluation methods in this work.

Focus Groups

The focus group is a form of group interview in which discussion and interaction within the group is part of the methodology [36]. People are encouraged to talk, exchange ideas, tell stories, comment on each other's ideas, and ask one another questions. The method is useful in evaluation in exploring learners' knowledge and experiences, and also in determining what they think and why. The idea of a focus group is that it may help to clarify ideas and views that might be less accessible in a one-to-one interview.

Krueger and Casey [36] provide advice on group composition, running the group (four to eight people is an ideal number), analysis, and writing up results. It is highly recommended that focus groups be recorded and transcribed for detailed analysis. Recording the group using a digital recorder and boundary microphone (for 360° capture of what people say) will give good sound quality. In addition, audio files may be stored on a computer and sent to participants and colleagues for further comments.

Group Consensus Techniques

A number of consensus techniques have been developed for the evaluation of educational events involving medium-to-large groups. Two commonly used approaches are described here.

Snowball Review

This is a group-based evaluation [6] that uses a series of steps where comment and opinion are suggested, discussed, shared, and agreed, before going on to the next step, until a final list of good and not so good points about, for example, a course has been agreed upon. The steps are as follows:

- Each individual alone lists, say, three good and three not so good points about the course being evaluated.
- Participants form pairs and discuss their suggestions, and then come to an agreement as a pair.
- Two such pairs then form a group of four, and this again debates the views and comes to an agreement.
- Two groups of four join up to form a group of eight. Again, they debate and agree upon their conclusions.
- A reporter presents the group's views to all participants. This is a good method in that it involves everyone and ideally reaches a consensus and a conclusion, but it does take time.

Nominal Group Technique

This is another group-based consensus method. It differs from the snowball review (above) in that each person gives their views and then all the views are collected and voted on. The steps are as follows:

- Each individual is asked in turn for feedback on the best and least valuable aspects of the course.
- Comments are collected and listed (once) on a flip chart, that is, if two members of the group thought that the catering was not very good, this is only listed once.
- The facilitator continues to go around the group until all (unique) comments have been exhausted.

Promoting the Use of Evaluation Findings

An essential question in judging the quality of an evaluation is 'are the results used?'. No matter how beautifully designed the evaluation strategy, the key question for evaluators is: Can the results be fed back into the system and acted upon? 'Use' can mean different things to different people. There are five different types of 'use' described in the evaluation literature [37]:

1. *Instrumental use:* when an evaluation directly affects decision-making and influences change. Evidence for this type of use involves decisions and actions that come from the evaluation, including the implementation of recommendations.
2. *Conceptual use:* when the evaluation findings help individuals understand the programme in a new way or influence thinking in a general way without any immediate new decisions being made about the programme.
3. *Enlightenment:* related to conceptual use, but more focused on whether the evaluation findings add knowledge to the field and thus may be used by anyone, not just those involved with the programme or evaluation of the programme.
4. *Process use:* how individuals and organisations are impacted as a result of participating in an evaluation. It acknowledges that being involved in an evaluation can lead to changes in how people think, what they do, and how they make decisions, which then result in cultural and organisational change.
5. *Symbolic use:* when evaluations are done as a requirement or political move, rather than serving an identifiable need. This is not an ideal or recommended use of evaluation.

There is a substantial amount of theoretical and empirical research on the use of results from evaluation studies. In a systematic review of the literature in this area, Johnson et al. [37] identified three broad categories of factors that increase the use of evaluation findings: (i) stakeholder involvement, (ii) characteristics related to implementation, and (iii) characteristics of the decision or policy setting; the context in which the evaluation was implemented. Their findings are elaborated and discussed below.

The strongest evidence regarding use of evaluation falls under stakeholder involvement, including early involvement, inclusion of different stakeholder groups, and an evaluator who communicated well and sustained engagement throughout the evaluation.

Characteristics related to implementation included the quality and credibility of the evaluation approach, evaluator competence, and timely/relevant reporting. However, the most critical factor in this category relates to evaluator/user communication. Frequency and quality of communication are important factors, as well as dissemination. The other interesting element in this category is the type of recommendations evaluators make; more specifically, if the recommendations are detailed and actionable items, they are more likely to result in changes, as compared to broad platitudes that have no practical significance.

The most important characteristics of context for increasing use are the personal characteristics of the individual(s) who will use the evaluation. To get the most value out of resources spent in conducting evaluations, the right decision makers must engage with the results. Important characteristics include the organisational role of the decision maker, the kind of authority they have, where they are in the organisational structure, and their level of experience and these are critical to increasing the probability that results will be used. An important consideration for the evaluator is the 'information processing style' of the user – dissemination of some 'reports' verbally, in briefing style, or by simply providing key findings and recommendations may be better than written reports.

Challenges in Educational Evaluation

There are several common pitfalls to avoid in any evaluation, some of which were mentioned earlier. The following are some common problems we have encountered in our careers as evaluators in medical education.

Timeliness

The timing of results is an important consideration. Too often planning and decision-making occur before data have been analysed or are even available. Providing a polished final report is not as important as providing results when planners and decision makers need them. It is often necessary to provide rapid turnaround. Using alternative vehicles for reporting, such as presentations, memos, or open discussion, are all good (and welcomed) substitutes to the long formal report.

Generalisability

Programme evaluation is often very local and the results may not be generalisable to other contexts. While the results may be useful at the local level, they may be challenging to publish. Because most evaluators working in medical education evaluation are academics, this represents an important challenge. It is impingent upon the evaluator to make a case for how the study contributes to the literature and to provide a thorough description of the context in which the work was conducted so that readers can determine its applicability to their context.

Standard of Acceptability

E-*valu*-ation is about 'valuing' or 'judging' the success of a programme. Therefore, some standard of comparison is necessary. It can often be challenging to get stakeholders to define what will be considered 'successful'. Is it 100% of trainees achieving licensure? If 80% of trainees are satisfied, will that be considered successful?

Measuring only what is Easy to Measure

Focusing on measuring what's easy to measure is a common pitfall. For instance, it might be quite easy to send out a questionnaire, but more difficult to use an alternate approach such as a case study, where problems such as bullying or negative feedback are the focus. Similarly, it may be attractive to gather only one source of evidence, rather than trying to triangulate evidence from three different areas.

Validity and Reliability of Surveys

An important activity involved in collecting evaluation information includes demonstrating the reliability (consistency) and validity (accuracy) of the measures. While perfect reliability and validity are impossible to achieve, it is important to know any instrument's limitations and how confident you can be in the information you have collected. Whether you use an instrument published in the literature, adapt an instrument, or develop your own, it is important to address the validity and reliability of the measures used.

Survey Fatigue and Low Response Rates

Trainees and students in medical education receive many surveys. Low response rates can pose a serious validity problem. While a response rate of 70% was previously considered a general standard for survey research, it is widely known that response rates have gone down over the years and are now often much lower. You will want to do as much as possible to keep the number and length of surveys reasonable. Providing incentives for special one-off surveys are worth considering to increase response rate, particularly among students. Keep in mind that it is not the response rate per se that is most important, but rather whether you are getting an accurate picture (or representative sample) since the purpose is to know about the general population who are recipients of the programme or policy.

Balancing the Positive and Negative

It is easy to fall into the trap of ignoring the positive accomplishments of a programme and to emphasise the negative. We have seen examples of this in evaluation studies and even on national quality assurance visits. Everyone is happy with the teaching except for one or two disaffected and sometimes dysfunctional individuals. The visiting team takes into consideration all of the concerns of those who are disaffected, to the exclusion of all other evidence. All of this appears in the report, but the emphasis on the negative, small though it may be, distorts the conclusions. Even if further investigations show much of this to be factually incorrect, the damage is done.

Conclusions

Evaluation is an essential part of medical education and needs to be carried out rigorously and systematically, all the while focusing on its utility. A good knowledge of many different areas is needed if we are to be able to carry out meaningful evaluations that can be fed back to improve medical education programmes. In the end, the ultimate goal is to design, implement, and use evaluation to deliver training that will ultimately produce quality patient care and a healthy population.

Examples of Evaluations in Medical Education

Example 1: Promoting Continuous Improvement in the Curriculum

Context

An internal evaluation unit of a medical undergraduate programme wanted to reinforce the use of evaluation findings and embed a new routine for addressing recommendations for curriculum improvement. To accomplish this an enhanced process was developed that included active collaboration in reviewing evaluation results and developing recommendations, assignment of a faculty lead for each recommendation, monitoring of actions taken, and regular reports on recommendations and their status to the relevant committees. A logic model was also developed that illustrated inputs, activities, outputs, and outcomes.

Programme Goals

Contribute to building a learning organisation by promoting continuous improvement of the curriculum.

Programme Objectives

1. Bring forward recommendations for programme improvements on a continuous basis.
2. Ensure recommended programme improvements are addressed.
3. Increase the use of evaluation results in planning and decision-making.
4. Demonstrate compliance with national accreditation standards regarding programme improvement.

Design

Mixed methods, including interviews to address process, and a pre–post comparison of key quantitative outcomes (Table 30.1).

Table 30.1 Evaluation outline design.

Evaluation question	Indicators	Standards of acceptability	Data collection method	Sources of data
Has the new process been implemented as planned?	Extent to which activities have been implemented	All activities are being fulfilled	Survey	Internal evaluators, relevant committee members
What have been the facilitators and barriers to implementation?	Challenges and facilitating factors	Facilitating factors will out-number challenges	Interviews	Internal evaluators, relevant committee members
Is there a greater uptake of recommended programme improvements?	Number of quality report recommendations addressed or being addressed	Increased uptake of recommendations, with 100% uptake by third year	Admin. records	
Are accreditation requirements achieved?	Accreditation report before and after implementation	Strong evaluation report from accrediting body at next visit	Review report	Accreditation report

Communication Plan

A communication plan was outlined to ensure that internal and external stakeholders were engaged throughout the evaluation. Products include a written evaluation report, a slide presentation, a two-page summary, and a brief and informal video that can be accessed through the Internet.

Example 2: Evaluating e-clips: an Evaluation Capacity Building Programme

The Communities of Learning, Inquiry, and Practice (CLIPs) model was developed within the higher education context to support faculty and staff. The model is designed to help professionals use evaluation to improve practice by building relationships and connecting with the larger purpose and goals of an institution [38]. CLIPs are a type of community of practice [39, 40] or groups that form to accomplish a purpose valued by all participants. These informal, dynamic groups support professionals in learning together within a supportive structure. The programme was implemented by an internal evaluation service supporting a medical educational programme. The CLIPs groups consisted of three to seven people who chose a shared evaluation question and project. One of the group members served as a facilitator. A 'guide' from the evaluation unit with an evaluation and facilitation expertise supported the work of multiple groups. The programme was piloted with 18 projects.

An evaluation conducted by an external evaluator addressed the following questions: (i) were the expectations and intended outcomes achieved, (ii) did the programme build evaluation capacity, and (iii) what factors influenced implementation, delivery, and utilisation? The external evaluator used a qualitative approach and conducted individual interviews with CLIPs' group leads and ran a focus group with those from the evaluation unit who were involved in the programme.

Interviews with project team leaders indicated that the CLIPs programme successfully delivered on a number of anticipated evaluation capacity-building outcomes occurring at the participant, programme, evaluation unit, and Faculty of Medicine levels. At the participant level, there was an increased appreciation and knowledge of evaluation as well as more independent evaluation activity by participating faculty. At the programme level, CLIPs have driven better quality evaluations and facilitated the increased use of evidence for programme improvement and decision-making. At the evaluation unit level, CLIPs led to the development of relationships with new stakeholders and created more knowledgeable and active evaluation stakeholders around faculty committee tables. At the faculty level, it created a group of CLIPs alumni who began functioning as internal evaluation champions. This 'learn by doing' design of the programme appears to be a very cost-effective mechanism for building evaluation capacity within the context where it was implemented.

The evaluation highlights the achievement of a number of important capacity-building outcomes, even among those with limited involvement, from what was a relatively small, simple, and low-cost intervention. Some changes were made to the way the programme was delivered, based on feedback, but the general approach was considered a success and was continued.

Example 3: Evaluating Readiness for Clerkship and Residency

There are times when evaluation questions relate to the effectiveness of a larger programme, rather than to individual components of a curriculum or course. Students' readiness to begin full-time patient-based learning, commonly called 'clerkship', is an important outcome of the first two years of medical training. Likewise, new medical graduates' readiness to begin the first year of residency training is an important outcome of the last two years of medical training. 'Readiness' is a key indicator as to whether a curriculum is functioning to enable students to achieve competencies expected mid-way through and at the end of the undergraduate training. Although the most efficient way to determine readiness is to ask students how much they have learned, there is a body of evidence showing that self-assessment data are not accurate measures of individual performance. However, recent findings using the *Readiness for Clerkship Survey* and the *Readiness for Residency Survey* [41–43] indicate that aggregated self-assessments from medical students and new residents were reliable for the purpose of evaluating overall programme effectiveness.

At the University of British Columbia Undergraduate Medical Education Program the *Readiness for Clerkship Survey* and the *Readiness for Residency Survey* were used to determine if undergraduate medical students and new residents were prepared for the next level of training. These surveys are unique in four ways: first, the items are competency-based key physician tasks that are blueprinted using the educational programme objectives; second, they are framed as self-assessments; third, the rating scale is based on the construct of independence; and fourth, they are administered four and eight months after the start of clerkship and residency, respectively.

Relative strengths and weaknesses of cohorts of trainees were identified when scores assigned to the items are aggregated and items are ranked from highest to lowest. By comparing medical student and resident ratings on the items common to both surveys, the function of the curriculum to enable learners to achieve greater competence where expected or its failure to do so, were identified. These data aided faculty in setting priorities for programme improvement and identifying and learning from effective educational practices.

In the process of using results it was found that summaries of key findings, use of visuals, tracking change over time, provision of student and resident comments, and working closely with faculty to develop recommendations were powerful in engaging faculty and facilitating use of the data. Sharing the results with students was also very beneficial to learning; junior students wanted to know the particular areas where new graduates felt they could have benefited from more assistance and guidance. Knowing what a near-peer sees as a challenge can be very powerful in guiding learning earlier in the programme. Students also appreciated knowing that the results are being used by the faculty to improve the programme and to support them in their learning.

Acknowledgement

We wish to acknowledge Professor David Wall whose authorship of a previous edition of this chapter has provided the foundation for this one.

References

1 Fink, A. (2005). *Evaluation Fundamentals: Insights into the Outcomes, Effectiveness, and Quality of Health Programs*, 2e. Thousand Oaks, CA: Sage.

2 Blanchard, R.D., Torbeck, L., and Blondeau, W. (2013). A snapshot of three common program evaluation approaches for medical education. *Academic Medicine* 88: 146.

3 Goldie, J. (2006). AMEE educational guide no. 29: evaluating educational programmes. *Medical Teacher* 28 (3): 210–224.

4 Norcini, J.J., Blank, L.L., Duffy, F.D., and Fortna, G.S. (2003). The miniCEX: a method of assessing clinical skills. *Anuals of Internal Medicine* 138: 476–483.

5 Mohanna, K., Wall, D., and Chambers, R. (2004). *Teaching Made Easy – A Manual for Health Professionals*, 2e. Oxford: Radcliffe Medical Press.

6 SCOPME (1996). *Appraising Doctors and Dentists in Training*. London: Standing Committee on Postgraduate Medical and Dental Education.

7 General Medical Council (2013). The Good Medical Practice Framework for Appraisal and Revalidation. https://www.gmc-uk.org/-/media/documents/The_Good_medical_practice_framework_for_appraisal_and_revalidation___DC5707.pdf_56235089.pdf (accessed 26 June 2018).

8 Preskill, H. and Russ-Eft, D. (2005). *Building Evaluation Capacity*. Thousand Oaks, CA: Sage.

9 Patton, M.Q. (2008). *Utilization-Focused Evaluation*. Thousand Oaks, CA: Sage.

10 Donaldson, S.I., Patton, M.Q., Fetterman, D.M., and Scriven, M. (2010). The 2009 Claremont debates: the promise and pitfalls of utilization-focused and empowerment evaluation. *Journal of Multidisciplinary Evaluation* 6: 15–57.

11 Vassar, M., Wheeler, D.I., Davison, M., and Franklin, J. (2010). Program evaluation in medical education: an overview of the utilization-focused approach. *Journal of Educational Evaluation for Health Professions* 7: 1.

12 Kirkpatrick, D.I. (1967). Evaluation of training. In: *Training and Development Handbook* (ed. R. Craig and I. Mittel), 87–112. New York: McGraw Hill.

13 Belfield, C.R., Thomas, H.R., Bullock, A.D. et al. (2001). Measuring effectiveness for best evidence medical education: a discussion. *Medical Teacher* 23: 164–170.

14 Parker, K. (2013). A better hammer in a better toolbox: considerations for the future of programme evaluation. *Medical Education* 47: 434–442.

15 Kirkpatrick, J.D. and Kirkpatrick, W.K. (2016). *Four Levels of Training Evaluation*. Alexandria, VA: ATD Press.

16 Moreau, K.A. (2017). Has the new Kirkpatrick generation built a better hammer for our evaluation toolbox? *Medical Teacher* 39: 999–1001.

17 Stufflebeam, D.L. (1983). The CIPP model for program evaluation. In: *Evaluation Models. Evaluation in Education and Human Services*, vol. 6 (ed. G.F. Madaus, M.S. Scriven and D.L. Stufflebeam), 117–141. Dordrecht: Springer.

18 Stufflebeam, D.L. (2003). The CIPP model for evaluation. In: *International Handbook of Educational Evaluation*, vol. 9 (ed. T. Kellaghan and D.L. Stufflebeam), 31–62. Kluwer International Handbooks of Education. Dordrecht: Springer.

19 Stufflebeam, D.L. and Shinkfield, A.J. (2007). *Evaluation Theory, Models, & Applications*. San Francisco, CA: Jossey-Bass.

20 Fetterman, D.M., Wandersman, A., and Millett, R.A. (2005). *Empowerment Evaluation Principles in Practice*. Thousand Oaks, CA: Sage.

21 Fetterman, D.M., Deitz, J., and Gesundheit, N. (2010). Empowerment evaluation: a collaborative approach to evaluating and transforming a medical school curriculum. *Academic Medicine* 85: 813–820.

22 Kellogg Foundation (2004). Logic Model Development Guide. https://ag.purdue.edu/extension/pdehs/Documents/Pub3669.pdf (accessed 25 June 2017).

23 Knowlton, L.W. and Phillips, C.C. (2009). *The Logic Model Guidebook: Better Strategies for Great Results*. Thousand Oaks, CA: Sage.

24 Bedward, J., Davison, I., Field, S., and Thomas, H. (2005). Audit, educational development and research: what counts for ethics and research governance? *Medical Teacher* 27: 99–101.

25 Boyer, E.L. (1990). *Scholarship Reconsidered: Priorities of the Professoriate*. Menlo Park, CA: Carnegie Foundation for the Advancement of Teaching.

26 Dyrbye, L.N., Thomas, M.R., Mechaber, A.J. et al. (2007). Medical education research and IRB review: an analysis and comparison of the IRB review process at six institutions. *Academic Medicine* 82: 654–660.

27 Egan-Lee, E., Freitag, S., Leblanc, V. et al. (2011). Twelve tips for ethical approval for research in health professions education. *Medical Teacher* 33: 268–272.
28 Morrison, M. (2005). Ethics. In: *Encyclopedia of Evaluation* (ed. S. Mathison). Thousand Oaks, CA: Sage.
29 Worthen, B.L., Sanders, J.R., and Fitzpatrick, J.L. (1997). *Program Evaluation: Alternative Approaches and Practical Guidelines*, 2e. New York: Longman.
30 Mayne, J. (2001). Addressing attribution through contribution analysis using performance measures sensibly. *Canadian Journal of Program Evaluation* 16: 1–24.
31 Mayne, J. (2012). Contribution analysis: coming of age? *Evaluation* 18 (3): 270–280.
32 Mayne, J. (2008). Contribution Analysis: An Approach to Exploring Cause and Effect. *ILIC Brief* (16): Rome: The Institutional Learning and Change Initiative. https://cgspace.cgiar.org/handle/10568/70124 (accessed 10 July 2017).
33 Bowling, A. (1997). *Research Methods in Health*. Buckingham: Open University Press.
34 Cohen, L., Manion, L., and Morrison, K. (2000). *Research Methods in Education*, 5e. London: Routledge Falmer.
35 Henry, R.C. and Wright, D.E. (2001). When do medical students become human subjects of research? The case of program evaluation. *Academic Medicine* 76: 871–875.
36 Krueger, R.A. and Casey, M.A. (2009). *Focus Groups: A Practical Guide for Applied Research*. Thousand Oaks, CA: Sage.
37 Johnson, K., Greenseid, L.O., Toal, S.A. et al. (2009). Research on evaluation use: a review of the empirical literature from 1986 to 2005. *American Journal of Evaluation* 30: 377–410.
38 Parsons, B. (2010). Communities of Learning, Inquiry, and Practice (CLIPS). AEA365: A Tip-a-Day by and for Evaluators. http://aea365.org/blog/beverly-parsons-on-communities-of-learning-inquiry-and-practice-clips (accessed 10 July 2017).
39 Wenger, E., McDermott, R., and Snyder, W. (2002). *Cultivating Communities of Practice*. Boston: Harvard Business School Press.
40 Parsons, B.L., Lovato, C.Y., Hutchinson, K., and Wilson, D. (2016). Building evaluation capacity through CLIPs: communities of learning, inquiry and practice. *New Directions for Program Evaluation* 151: 21–36. Jossey-Bass and the American Evaluation Association.
41 Peterson, L.N., Rusticus, S.A., Eva, K.W., and Lovato, C.Y. (2012). A readiness for clerkship survey; can self-assessment data be used to evaluate program effectiveness? *Academic Medicine* 87 (10): 1355–1360.
42 Peterson, L.N., Rusticus, S.A., Wilson, D.A. et al. (2015). Readiness for residency: survey to evaluate undergraduate medical education programs. *Academic Medicine* 90: S36–S42.
43 Peterson, L.N., Rusticus, S.A., Eva, K.W. et al. (2017). Using the readiness for clerkship and residency surveys to evaluate the effectiveness of four MD Programs: a cross-institutional generalizability study. *Academic Medicine* 92 (11): S100–S109.

Further Reading and Resources

Andres, L. (2012). *Designing and Doing Survey Research*. London: Sage.

Better Evaluation. An international collaboration to improve evaluation practice and theory by sharing and generating information about options (methods or processes) and approaches. https://www.betterevaluation.org/ (accessed 26 June 2018).

British Educational Research Association (2011). Ethical guidelines for educational research. British Educational Research Association, Southwell. www.bera.ac.uk/researchers-resources/publications/ethical-guidelines-for-educational-research-2011 (accessed 10 July 2017).

Cohen, L., Manion, L., and Morrison, K. (2000). *Research Methods in Education*, 5e. London: Routledge Falmer.

Cook, D. (2010). Twelve tips for evaluating educational programs. *Medical Teacher* 32: 296–301.

Donaldson, S.I., Christie, C.A., and Mark, M.M. (2009). *What Counts as Credible Evidence in Applied Research and Evaluation Practice?* Thousand Oaks, CA: Sage.

Fry, D.A. and Hemmer, P.A. (2012). Program evaluation models and related theories: AMEE guide 67. *Medical Teacher* 34: e288–e299.

Owen, J.M. (2007). *Program Evaluation: Forms and Approaches*, 3e. New York: Guildford Press.

Torre, D.M. and Common, D.B. (2013). Evaluation designs in medical education I. *Academic Medicine* 88 (11): 1784.

Torre, D.M., Ferris, A., Daley, B., and Durning, S.J. (2016). Common evaluation designs in medical education II. *Academic Medicine* 91 (11): 1584.

Torre, D.M., Ferris, A., Daley, B., and Durning, S. (2016). Common evaluation designs in medical education III. *Academic Medicine* 91 (12): 1707.

Western Michigan University. Evaluation Checklists. https://wmich.edu/evaluation/checklists (accessed 10 July 2017).

Yarbrough, D.B., Shulha, L.M., Hopson, R.K., and Caruthers, F.A. (2011). *The Program Evaluation Standards: A Guide for Evaluators and Evaluation Users*, 3e. Thousand Oaks, CA: Sage.

31 Knowledge Synthesis

Lauren A. Maggio[1], *Aliki Thomas*[2], *and Steven J. Durning*[1]

[1]Division of Health Professions Education, Department of Medicine, Uniformed Services University of the Health Sciences, Bethesda, MD, USA

[2]Centre for Medical Education, School of Physical and Occupational Therapy, Faculty of Medicine, McGill University, Montreal, Quebec, Canada

KEY MESSAGES

- Knowledge syntheses summarise a body of work thereby providing the state of what is known about a topic and highlighting areas for future research.
- Literature reviews and knowledge syntheses are distinct, but related, review processes and although the terms are often used interchangeably, they are not synonymous.
- There are multiple types of knowledge syntheses (e.g. narrative reviews, systematic reviews, umbrella reviews, scoping reviews, realist reviews, etc.).
- Conducting a knowledge synthesis is an intensive multi-step process.
- Knowledge syntheses can inform educational practice, policy, and research.

Introduction

Over the last 20 years, the volume of publications on topics in health professions education (HPE) has increased dramatically [1]. In 2016 alone there were over 25 000 articles about HPE topics added to PubMed. For most in the HPE community, it is impossible to keep up with this influx of articles. To help integrate some of these multiple studies, researchers are increasingly creating, consuming, and citing knowledge syntheses.

In this chapter, we define and describe the characteristics of a knowledge synthesis. We also situate knowledge syntheses in the context of HPE, including a discussion of how they may be used and the five types of knowledge syntheses that are prevalent in, or appropriate for use by, those in HPE. Next, we outline a seven-step process for those seeking to undertake knowledge syntheses. Lastly, we explore available training for knowledge syntheses in HPE.

What are Knowledge Syntheses?

Knowledge syntheses, sometimes also referred to as integrative scholarship [2], are comprehensive reviews of an extant body of literature. The 'idea' of knowledge synthesis is not a new one. Knowledge syntheses in philosophy actually date back to the twelfth century. During the seventeenth century, new statistical techniques and procedures were developed to synthesise the literature and were particularly common practice in astronomy. In the early twentieth century, basic meta-analytical (a type of knowledge synthesis) procedures were developed to address a growing body of health-related research. It was not until the 1970s that researchers and statisticians in education, psychology, and the social sciences began to advance more modern techniques for meta-analysis. The decades that followed saw an emergence of methods for conducting knowledge syntheses, including systematic reviews and meta-analyses by international organisations such as the Cochrane collaboration. Though these primarily quantitative syntheses methods used methodologically rigorous and sophisticated techniques aimed at determining the effectiveness of specific interventions, health care systems worldwide faced new, multifaceted, and increasingly complex issues that could not be examined using traditional quantitative methods [3]. Multifaceted biopsychosocial issues faced by individual patients and patient populations necessitated an examination of more 'complex' research evidence using newer methods or including a combination of qualitative and quantitative methods.

> Systematic reviews based on these methods have changed the landscape of health research, but the methods themselves have been criticized as being unable to address questions of great complexity, such as exploring patients' perceptions of disease, identifying underlying theories to explain the effectiveness of an intervention, or understanding the facilitators of and barriers to the uptake of an intervention. Because traditional systematic review methods may be inadequate to address these questions, other types of search approaches (e.g., snowballing of papers, focusing on identification of

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

key theories) and analysis techniques (e.g., thematic analysis, grounded theory) may be required [4, p. 20].

This evolution in the nature and breadth of research evidence in health care led to the emergence of new knowledge synthesis methods (e.g. critical interpretive synthesis, integrative review, meta-synthesis, meta-ethnography, narrative synthesis, realist review, scoping review, mapping review, rapid review) some of which will be described further in this chapter [4]. A detailed description of each of these methods is outside the scope of this chapter, but we provide a reference list and recommended reading section for those interested in greater depth.

Literature reviews and knowledge syntheses are distinct but related review processes. Although the terms are often used interchangeably, they are not synonymous. Whether undertaking a literature review or knowledge synthesis, the author must be aware of the distinctions and make an informed decision about the type of review most appropriate for the question being posed or the task at hand (e.g. a preliminary review of the literature prior to any empirical study).

Literature reviews, unlike knowledge syntheses, are often considered more cursory methods for obtaining a general overview of the state of the knowledge in a given area or field of study. As such, they are ideal for identifying gaps in the literature and for informing the next steps in a programme of research. There is no expectation that a formal critical appraisal be conducted and the output tends to be a narrative account of the state of knowledge in the chosen domain. Literature reviews are therefore an important starting point for individuals conducting empirical work, for situating a study's research question(s), methodology, and potential contributions to the field, as well as for avoiding unnecessary duplication of research [5].

Why Use Knowledge Syntheses in Health Professions Education?

HPE is a complex field. Educational researchers often ask complex questions grounded in different epistemological perspectives and for which a multitude of research methods exist. Given the numerous synthesis methods available, it can be challenging to select the method for the right question. Indeed, the notion of alignment – the right synthesis method for the right question – is as central to a knowledge synthesis as it is to an empirical study.

It is important to note that well-conducted knowledge syntheses require many judgements and decisions both before embarking on the review, as well as during the process. Decisions prior to embarking on the synthesis are key because studies to be included are identified after they have been completed and reported. As such, the authors may already know the results of many of the included studies. Importantly, the authors' knowledge of previous studies may facilitate the formulation of the review question, the criteria for study selection, the comparisons for analyses, and the outcomes to be reported in the synthesis by helping to focus authors on key areas of need.

There are four main reasons for conducting a knowledge synthesis: (i) to summarise a body of work, thereby providing the state of knowledge in an area; (ii) to identify gaps in the literature and thus in the knowledge base; (iii) to provide avenues for future research; and (iv) to create a knowledge 'tool' that can be used to inform educational practice and/or policy [5]. Knowledge syntheses can be used to create knowledge tools such as a set of guidelines and/or recommendations or a specialised publication type like '12 Tips', which is featured in *Medical Teacher*. These resources can be useful for HPE researchers and practitioners to consult and utilise to make important teaching, assessment, or educational policy decisions. As such, the knowledge synthesis serves a practical purpose for various HPE stakeholder groups.

Types of Knowledge Syntheses

Authors approach synthesising the literature for a variety of reasons and from a multitude of perspectives. Thus, there is no single approach or 'gold standard' for undertaking a knowledge synthesis [6]. More than 25 different types of knowledge syntheses have been identified in the literature [4]. In this chapter, we discuss five types most commonly used in HPE: narrative reviews, systematic reviews, umbrella reviews (aka meta-syntheses), scoping reviews, and realist reviews. We chose these based on their prevalence in HPE and because they could potentially help HPE researchers answer the complex, inter-disciplinary questions that arise in HPE. Our selection also represents examples of knowledge syntheses with a range of theoretical and epistemological underpinnings – from systematic reviews that have roots in positivism (e.g. knowledge can be verified through observation) to scoping reviews, which are influenced by constructivist theories in that they are rooted in the belief that knowledge is socially constructed. See Chapter 27 in this book for more on research paradigms generally. More specifically, Gordon provides an overview on selecting paradigms for HPE knowledge syntheses [7].

When appropriate in the following sections, for each knowledge synthesis type, we discuss typical research questions, relevant methodological frameworks, use of a critical appraisal process, and practical tips. Readers interested in knowledge syntheses methods beyond those featured here should consult resources such as Tricco's typology of knowledge syntheses [4] and Kastner's guidance on how to select an appropriate synthesis type [8]. For each knowledge synthesis type, we describe key characteristics and provide examples of its use in HPE practice and research. When appropriate, we also include potential challenges to undertaking the review in an HPE context. As we only provide an overview of each type, we have included supplementary references for those interested in the details of executing each knowledge synthesis in the recommended resources section at the end of the chapter.

The approach and execution of knowledge syntheses can vary. For example, depending on knowledge synthesis type (e.g. scoping review versus systematic review), a synthesis might incorporate various sources of information, from empirical research to editorials. Also, depending on the methodology and review question, knowledge syntheses may or may not use formal critical appraisal processes to assess the quality of the included research [8]. In this chapter, we describe similarities and differences among the five selected knowledge synthesis types and highlight them in Box 31.1 [9–14].

Narrative Reviews

Narrative reviews, which have been characterised as the 'simplest form' of knowledge synthesis, generally attempt to describe what is known on a topic or subject based on the existing published literature [15]. Narrative reviews rarely contain descriptions of a formal search or how articles were selected for inclusion. Instead, authors take a constructivist approach using their discretion to select evidence from a variety of materials that they see fit for inclusion and provide commentary on their contents. For example, in the narrative review, 'Good teaching is good teaching: A narrative review for effective medical educators' [16], the author notes that his descriptions of the characteristics of good teachers are based on the literature, but also informed by his interactions with teachers across his career as a professional educator.

Narrative reviews have been criticised for being difficult or impossible to replicate and prone to bias [17]. Nevertheless, narrative reviews can be helpful to HPE researchers as a starting point for understanding a topic, learning the vocabulary of a topic, and identifying related trends, controversies, and needs for further research [15]. Furthermore, narrative reviews tend to be well cited in HPE [18]. For example, the narrative review article 'Problem-based learning: A review of literature on its outcomes and implementation issues' [19] is currently the top cited article of all time in HPE.

Systematic Reviews

A systematic review is defined as: 'A review of a clearly formulated question that uses systematic and explicit methods to identify, select, and critically appraise relevant research, and to collect and analyze data from the studies that are included in the review' [20, p. 1.1.2]. Popularised by the evidence-based medicine movement, systematic reviews take root in the positivist tradition and seek to provide an observable or measurable truth that can be applied in practice [6]. In HPE, systematic reviews have become increasingly prevalent, perhaps in part due to the efforts of the Best Evidence in Medical Education (BEME) Collaboration, which supports their creation and dissemination and is increasingly emphasising the importance of uptake of BEME reviews in educational practice and policy [21]. (See Box 31.2 [22–25].)

Systematic reviews include a well-defined question, clearly documented methods, including a comprehensive literature search, explicit criteria for the inclusion and exclusion of studies, and a critical appraisal of those included. Systematic reviews are resource intensive with a documented average of five review team members and an average completion time of 67.3 weeks [26]. Increasingly

BOX 31.1 FOCUS ON: Five types of knowledge synthesis

We provide a general comparison of components of knowledge synthesis types. However, there are no hard and fast rules and knowledge syntheses may vary from these generalisations.

	Narrative review[a]	Systematic review	Umbrella review/ meta-synthesis	Scoping review	Realist review
Methodological framework	To our knowledge unavailable	BEME [9]	Aromataris [10]	Arksey [11]	Pawson [12]
Epistemology	Varies	Positivist	Positivist	Constructivist	Realist
Search	Generally not expected to be comprehensive	Comprehensive	Comprehensive	Comprehensive	Comprehensive
Types of materials included	Generally journal articles	Journal articles	Systematic reviews	All formats of materials are eligible	All formats of materials are eligible with an emphasis on theory
Critical appraisal	Not required	Required	Required	Not required	Not required
Team members	May be a single author, but generally multiple authors	Generally multiple authors	Generally multiple authors	Generally multiple authors, including community stakeholders	Generally multiple authors, including community stakeholders
Reporting guidelines	Unavailable	PRISMA [13]	Unavailable	Unavailable	RAMESES [14]

[a] Narrative reviews and literature reviews are very similar.

BOX 31.2 FOCUS ON: The Best Evidence in Medical Education (BEME) Collaboration

The BEME Collaboration was established in 1999 as a rejection of basing educational approaches on anecdotal evidence [22]. To this end, this international group of health professions educators is committed to fostering the dissemination of available best evidence, the production and peer-review of high-quality reviews geared towards HPE practitioners, and the creation of an evidence-informed culture in HPE (BEME website). The BEME Collaboration has supported more than 40 BEME Guides that cover a variety of HPE topics ranging from the effects of interprofessional interventions [23] to an exploration of career choice [24]. While the majority of BEME Guides are systematic reviews there are also other types of knowledge syntheses, such as realist reviews [25]. For additional information about the BEME Collaboration and instructions on how to submit a BEME Guide topic consult: http://www.bemecollaboration.org

journals require that researchers undertaking systematic reviews use the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [13] to guide the reporting of the review. A PRISMA flowchart includes a 27-point checklist of the necessary details to enable future researchers to replicate the review and to increase transparency.

Authors of systematic reviews are primarily interested in answering a specific research question about what works best (e.g. what is most effective) – in relation to a practical problem [15]. For example, an HPE researcher and/or practitioner interested in understanding the impact of online learning compared to traditional face-to-face teaching approaches would be well served by conducting or reading a systematic review. A systematic review may be less appropriate for the researcher or practitioner attempting to examine a broad question or a new topic for which the literature base is not yet developed.

In systematic reviews, analysis of included studies includes quantitative, qualitative, or a combination of both approaches. If taking a quantitative approach, some systematic reviews include a meta-analysis of data from individual studies in the review. A meta-analysis seeks to use data extraction techniques and statistical methods to aggregate quantitative data from each included study in order to evaluate the consistency of the results across studies, investigate potential reasons for variation in study findings, and calculate summary effect sizes with a confidence interval to determine the effectiveness of an intervention [15]. For example, Brydges et al. published a systematic review and meta-analysis focused on simulation-based training that analysed and pooled the effects of the interventions designed to support learners in self-regulated learning activities [27]. For further information on systematic reviews, including meta-analyses, please see Chapter 28 in this book.

HPE systematic reviews have been criticised for providing limited details of the included studies and for being abstract [28]. This lack of detail renders the systematic review of limited value for the educator attempting to utilise the evidence in their educational practice. Additionally, systematic reviews have been criticised for not yielding results that are applicable or relevant. To help combat this issue, Gordon has proposed the STructured apprOach to the Reporting in health care education of Evidence Synthesis (STORIES) statement, which builds on resources, such as the PRISMA statement, by incorporating key educational elements to be reported [29]. (See Box 31.3.)

Umbrella Reviews

An umbrella review or meta-synthesis integrates and synthesises evidence from multiple systematic reviews and/or meta-analyses that address the same topic [15, 30]. Umbrella reviews can include qualitative and/or quantitative studies. With the increase in number of published systematic reviews, umbrella reviews allow researchers to summarise the overall body of knowledge available on a topic in one resource with the goal of answering a specific question. Additionally, umbrella reviews provide the opportunity to compare and contrast the findings from systematic reviews to shed light on the consistency or discordance of available evidence on a topic [10]. Based on systematic reviews and with a similar purpose, umbrella reviews are also rooted in a positivist paradigm. However, as researchers are increasingly undertaking qualitative meta-syntheses they are also approaching umbrella reviews based on alternate paradigms as to best fit their review.

In HPE, umbrella reviews can be helpful for researchers and practitioners seeking information on topics that have been well covered by previous syntheses. For example, many systematic reviews examine the efficacy of simulation education. To pool these reviews, a recent umbrella review, 'Use of simulation-based learning in

BOX 31.3 FOCUS ON: The STORIES statement

To guide the reporting of knowledge syntheses in HPE, in 2014 Gordon and Gibbs published the STructured apprOach to the Reporting in health care education of Evidence Syntheses (STORIES) statement [29]. The STORIES statement contains 25 items that guide users through title construction to conclusion of a knowledge synthesis. A component of the items are derived from earlier knowledge syntheses guidelines [13, 14]; however, there are several items focused on education. An example is item 17 'Describe methods of quality assessment of education reported' [29]. Additionally, users are asked to consider the implications of a synthesis for educators. Although not a prescriptive checklist, the STORIES statement can be helpful to those authoring and peer reviewing HPE knowledge syntheses.

undergraduate nurse education: An umbrella review' was published and drew conclusions based on 25 systematic reviews [31].

A criticism of umbrella reviews is that they are limited by the quality and comprehensiveness of the original systematic reviews available on a topic [30]. Additionally, as a distillation of many systematic reviews, umbrella reviews are two steps removed from the original studies that are included. Therefore, for some HPE practitioners, this type of knowledge synthesis may not yield the level of detail necessary for implementing findings into practice without needing to revisit the original studies or systematic reviews.

Scoping Reviews

Scoping reviews attempt to answer broad, exploratory questions in order to map key concepts, characterise types of available evidence, and identify gaps in research on a particular topic [32]. Scoping reviews are also completed to determine whether or not researchers should undertake a full systematic review on a topic [33]. For example, the authors of 'Feedback for learners in medical education: What is known? A scoping review' [34] present a detailed road map of the feedback literature and conclude by providing readers with suggested research questions for future systematic reviews. In HPE, practitioners may find scoping reviews valuable in orienting them to the entirety of a topic, especially if it is a new and uncharted subject area in HPE.

When conducting a scoping review, authors are expected to undertake a systematic and comprehensive search to include a variety of material types, such as primary research articles, opinion pieces, grey literature, information from websites, etc. Casting a wide net, scoping reviews also include materials that focus broadly to include a range of study designs, diverse learner groups, variety of educational settings, and range of skills and competencies. In this way, scoping review authors operate from a constructivist approach. Once search results are obtained, authors synthesise and characterise the findings to generate a map – often in the form of tables, figures, or graphical representations [33]. Increasingly, researchers are encouraged to engage stakeholders in scoping reviews and to actively distribute results in formats accessible to them. Unlike a systematic review, the synthesis of the findings is iterative and flexible. There is growing debate as to whether or not it is necessary to have a protocol prior to beginning the review as is the case with systematic reviews and whether to critically appraise the included materials for their methodological qualities in a scoping review [15]. Lastly, despite being described as aiming to 'rapidly map the key concepts underpinning a research area' [11] scoping reviews are time consuming and require significant resources.

Realist Reviews

Authors of realist reviews focus on attempting to understand the mechanisms by which an intervention works or does not work, for whom, in what circumstances, in what respects, and why? [12, 35]. Realist syntheses have been suggested for HPE practitioners attempting to implement or research educational approaches, which often occur in unique and complex learning contexts [35]. Realist reviews are rooted in realism, which posits that 'causal explanations are achievable; social reality is mainly an interpretative reality of social actors; and social actors evaluate their social reality' [36].

The execution of a realist review is iterative and focused on testing theory to understand and explain intervention mechanisms. For example, Kehoe published the realist review 'Supporting international medical graduates' transition to their host-country: A realist synthesis' [37]. This realist review begins by identifying and testing several programme models and theories against examples of international medical graduate programmes in the literature. Through this process, in which the examined programmes act as case studies, the authors tested the fitness of existing theory and generated their own refined programme theory which they felt explained why programmes work or fail and in which contexts. Similar to scoping reviews, guidelines to inform this process encourage authors to incorporate stakeholders into various components of the research, including in the formulation of the question and for providing feedback on findings. For example, Kehoe included international medical graduates, programme directors, policy makers, and practising physicians to discuss relevant theories and to judge the applicability of the review's results [37].

For HPE, realist reviews are well suited for examining evidence about implementation and evaluation of complex interventions [36]. Wong suggests the usefulness of this approach in HPE in several instances, including when trials of interventions have produced inconsistent efficacy results and consensus is limited on when, who, and with whom to use the intervention; when newly developed interventions are being tested to determine how they are effective and for whom; and when the materials available on an intervention are predominantly qualitative or drawn from the grey literature [35]. However, due to the iterative and flexible nature of realist reviews, experts in the field do not recommend this approach for novices [12]. Lastly, realist reviews tend to be quite time consuming and therefore resource intensive due to their complexity [36].

Conducting a Knowledge Synthesis

In the above section, we focused on five specific types of knowledge syntheses. In these descriptions, we identified some of the ways in which they differ, such as in the questions they pose or the search strategies they employ. Despite these potential differences, knowledge syntheses generally share several key steps, which are critical to the success of their conduct [7]. Based on the stepwise approaches presented by Tricco [38] for knowledge syntheses in general and by Cook [39] to execute a systematic review, we next describe seven shared steps for conducting knowledge syntheses (Figure 31.1). We acknowledge that these steps are not exhaustive and vary depending on the chosen knowledge synthesis type. We encourage readers wanting more in-depth discussions of a

Figure 31.1 How to conduct a knowledge synthesis.

specific review type to consult the suggested resources at the end of this chapter. We have also provided examples of the steps as executed in each of the five knowledge syntheses types in Box 31.5 [40–46], located near the end of this chapter.

Define a Focused Research Question

First and foremost, authors should clearly define their research question(s). The research question should be important to the authors undertaking the review and be relevant to the broader educational community. The research question will shape the conduct of the entire knowledge synthesis and serve as a touchstone for the authors as they move through the steps of the process.

To formulate questions, some researchers have suggested the use of the mnemonic PICO (Population, Intervention, Comparison or Context, Outcome of interest) [39]. Although valuable, particularly for systematic reviews, the PICO format is not appropriate for all knowledge syntheses types. For example, PICO does not work well for scoping reviews where questions are often formulated as, or begin with, 'What is known on …'. We caution authors against trying to 'fit' a research question into a particular format. Rather, we suggest that the authors consider the constructs of interest, the possible relationships among them, and the potential contribution to knowledge when formulating a synthesis question.

Determine Knowledge Synthesis Type

The determination of which knowledge synthesis type to select can be a somewhat iterative process. Typically, authors determine the type of knowledge synthesis based on the type appropriate for answering their research question. In some cases, authors may select a particular knowledge synthesis type that then guides the formulation of the question. Additionally, selection of knowledge synthesis type is informed by the purpose of the knowledge synthesis and the needs of its anticipated end users [8]. For example, in selecting a synthesis type, an author might consider whether they are hoping to describe the effectiveness of an intervention (e.g. a systematic review), identify gaps in the existing literature (e.g. a scoping review), or generate theory (e.g. a realist review). Each of these desired outcomes would impact the type of synthesis selected. Other factors, such as the current state of the literature on a topic, will also impact the selection of knowledge synthesis type. A potential pitfall at this step is that authors may be unaware of the variety of knowledge syntheses types available and therefore select a type that is suboptimal. Thus, authors should explore various knowledge syntheses types, including those perhaps not well established in HPE, to determine the best type for their research question. Authors should consider consulting Tricco's list of 25 knowledge syntheses types [4]. We also urge authors to articulate a clear and answerable question, and, as is the case with most scientific endeavours, to then find the most appropriate synthesis method to answer the question.

Recruit the Research Team

Undertaking a knowledge synthesis is rarely a solo endeavour and the recruitment of a diverse team can have a positive impact on the process and the outcome [39]. Currently, there is no gold standard for the makeup of a knowledge synthesis team. Generally, the team will be determined depending on the research question and knowledge synthesis type. Based on these factors, authors might recruit team members with expertise in the following:

- the subject matter
- methodological approaches (e.g. statistician, qualitative methodologist, and synthesis methodologist as well)
- literature searching (e.g. medical librarian) (see Box 31.4 [47–53])
- data management.

To ensure that a knowledge synthesis meets end users' needs, when that is one of its objectives, authors should consider the inclusion or consultation of an appropriate stakeholder. For example, authors synthesising the literature on the use of e-learning for nurse educators may benefit from consulting or including a front-line nurse educator to determine the relevance of their findings to their teaching practice.

When recruiting the research team, authors may face timing challenges such that members with specific expertise may not be incorporated from the start of a project. For

BOX 31.4 WHERE'S THE EVIDENCE: For including a librarian in a knowledge synthesis

Increasingly, medical librarians are being added to knowledge synthesis teams [47]. Additionally, journals, such as *JAMA* [48], and organisations, including the Institute of Medicine [49] and the Cochrane Collaboration [20], recommend their inclusion. As collaborators, medical librarians provide expertise in searching the biomedical literature, documenting the review process, and facilitating the management of found information [50]. Researchers have demonstrated that the inclusion of medical librarians on systematic review teams correlates with significantly higher quality reported search strategies [51] and their role has been highlighted for facilitating scoping reviews [52]. Additionally, searches conducted and reported by medical librarians tend to be better described and therefore more reproducible than those undertaken by non-librarians [53].

example, a statistician may not be added to a team until the authors are preparing to analyse data. Ideally, the statistician would have been included from the conceptualisation of the project to inform its design and ensure appropriate data collection and presentation. Therefore, authors should carefully consider team member recruitment early in the project.

Identify Materials for Inclusion

The selection of materials (e.g. websites, journal articles, policy papers) for inclusion will depend on the type of knowledge synthesis undertaken. For example, an umbrella review will include only systematic reviews, whereas a scoping review may include a variety of sources – including journal articles, theses, and meeting abstracts. Regardless of the knowledge synthesis type selected, except for narrative reviews, it is important that authors report the selection of materials in a clear and auditable fashion.

For the identification and retrieval of materials, many helpful resources have been published to facilitate literature searching [54–56]. These resources address how to formulate search queries, select databases, perform hand searches, and locate the grey literature. For most review types, authors utilise electronic databases to locate materials. To improve database search strategies and ensure the effective retrieval of materials, some authors elect to have medical librarians peer review their search strategies. To structure this peer review, librarians are encouraged to use the Peer Review of Electronic Search Strategies (PRESS) guideline [57].

When identifying materials for inclusion in a knowledge synthesis, an author may consider a variety of formats for incorporation. Including materials other than journal articles enables authors to provide more complete and, in some cases, timely coverage of a topic. In the case of scoping reviews, these other sources provide a more comprehensive and nuanced perspective on the knowledge base. For a realist review, the inclusion of theory and educational models play a major role. In the case of systematic reviews, authors of a systematic review of educational interventions for teaching nutrition included several educational interventions from MedEdPortal that were not reported in journal articles and therefore would have been missed [58]. In some cases, reviews also include conference proceedings and meeting abstracts. These materials can provide timely glimpses into cutting edge research that may not, due to publishing timelines, appear as journal articles for quite some time.

Despite benefits, the inclusion of non-journal articles presents challenges. Authors may find it difficult to pool information in a coherent fashion or to compare and contrast information that is presented in very different formats. Additionally, authors may be challenged to retrieve the full-text of these materials, as academic libraries do not necessarily collect them. Therefore, authors should account for additional time in their process to secure these materials through Internet searches and author outreach.

As authors identify materials for inclusion, they should also identify a strategy for managing them. Without a strategy, authors could quickly find themselves buried by the high volume of materials. Authors may find the use of a bibliographic reference manager (e.g. Zotero, EndNote, etc.) helpful for retrieving and organising references from multiple information resources. Additionally, many reference managers also support the location and storage of the full text of materials and synchronise with author's library subscription resources.

Once identified, authors must determine which materials to include or exclude. Inclusion and exclusion criteria depend on the research question and type of knowledge synthesis. For example, a knowledge synthesis focused on recent technology may exclude materials published before the technology was introduced. A knowledge synthesis addressing an assessment instrument for practising clinicians may exclude health professions students. Authors may predefine these criteria, but in some review types (e.g. scoping) criteria may also arise as the authors become familiar with the literature available on a topic. The selected criteria for inclusion notwithstanding, it is imperative that they are clearly defined as this impacts the generalizability of the knowledge synthesis [9]. To enhance clarity, some knowledge synthesis types, such as systematic and scoping reviews, provide a flow chart to graphically represent the inclusion and exclusion of studies (Figure 31.2).

Extract Key Data

Once materials for inclusion have been determined, authors can begin extracting key information. Data extraction should be driven by the research question and authors may find that PICO is helpful to inform the process [39]. For example, using the PICO format, researchers would extract key details about Participants (learner level, professional discipline, geographic region, etc.); the Intervention (characteristics of the design, implementation, timing, etc.); the Comparison (alternative intervention and/or characteristics of the design, other implementation strategy, etc.); and hoped for Outcomes (increased learning, superior performance on a task, decreased burnout, etc.).

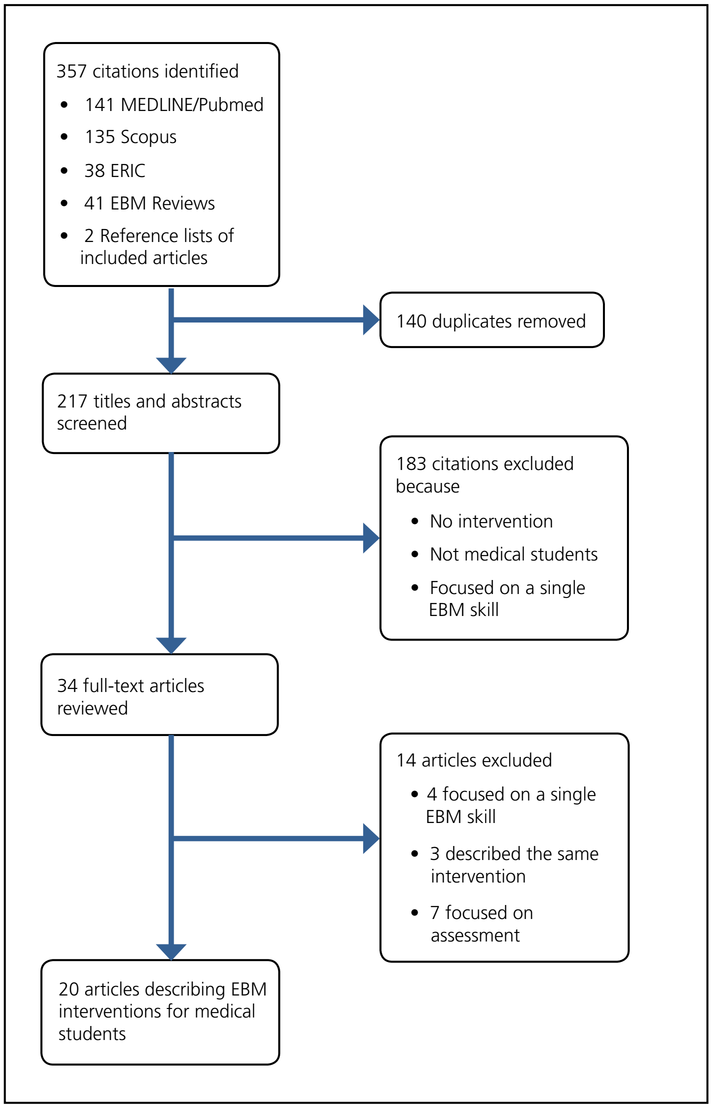


Figure 31.2 A flow diagram depicting the search and inclusion process for a systematic review on training medical students how to locate medical information [59].

As discussed above, the PICO approach to data extraction works well for systematic reviews and umbrella reviews but may not be ideal for all knowledge synthesis types. Therefore, Sharma has also suggested that those undertaking other types of reviews, such as a realist review, consider focusing on the following elements to extract:

- utilised conceptual frameworks or theories
- specified learning outcomes
- pedagogical approaches
- resources and equipment required [9].

For data extraction, researchers might consider using a data abstraction tool or sheet, which they develop, based on the review questions and units of analysis. For example, the BEME Collaboration has created a data abstraction tool for HPE systematic reviews, which can be modified to meet researcher needs (this can be found on the BEME website). It is suggested that authors create their data abstraction tool *a priori* so that it can be piloted prior to use for screening all materials. This can help researchers and their team avoid multiple data extraction cycles [38], which is a common pitfall at the data extraction stage. To facilitate the data extraction process, researchers might consider a literature review tool, such as Distiller or Covidence. These subscription resources enable the easy download of materials, provide a platform for extracting data, generate customised reports, and allow for real-time collaboration among team members.

Analyse and Synthesise Results

Writers of knowledge syntheses have been encouraged to ensure that they are synthesising the literature, not simply cataloguing it, to ensure the creation of new knowledge [39]. In knowledge syntheses, analysis might utilise quantitative and/or qualitative methods.

Quantitative analyses aim to map and aggregate information such as years of publication, countries of

BOX 31.5 HOW TO: Examples of the seven steps of knowledge synthesis

The examples below show how authors of knowledge syntheses executed the seven steps in published articles. These examples demonstrate that not all knowledge syntheses clearly document all seven steps. Instead, the authors used their judgement as to which details to report and what details might be less essential for their particular article and the journal to which it was submitted.

	Narrative review	Systematic review	Umbrella review	Scoping review	Realist review
Published knowledge synthesis title	Good teaching is good teaching: A narrative review for effective medical teachers [16]	Translating knowledge in rehabilitation: A systematic review [40]	Evidence for curricular and instructional design approaches in undergraduate medical education: An umbrella review [41]	Feedback for learners in medical education: What is known? A scoping review [34]	Key characteristics of successful quality improvement (QI) curricula in physician education: a realist review [42]
Step 1: Define a focused research question	The process is not defined, which is not uncommon or considered negatively for this synthesis type Process not described Question: What are characteristics of good teachers?	Process: identified based on previous knowledge translation (KT) reviews in allied health professions Question: What are the KT strategies that influence rehabilitation disciplines?	The process not described, which is uncommon for this type of synthesis type. Further details would be preferred. Question: What are the learning outcomes of curricular design and instructional design approaches employed in undergraduate medical education?	Process: Collaboratively defined by the research team Question: What has been broadly published in the literature about feedback to help learners in medical education?	Process: Defined collaboratively by research team Question: How does teaching (QI) in clinical settings enhance patient care and system performance?
Step 2: Determine knowledge synthesis type	The process is not defined, which is not uncommon or considered negatively for this synthesis type	Described a gap in the KT systematic review literature on rehabilitation	Described a need for an overarching scan of existing systematic reviews	Described the existing literature as 'far-ranging' A need to map key concepts and highlight existing gaps	Discussed that QI research has not assessed contextual or mechanistic factors that predict training success
Step 3: Recruit the research team	A single author	Five authors with backgrounds in nursing, education, pharmacy, and public health Acknowledged a librarian	Five authors with backgrounds in HPE and librarianship	Six authors with expertise in medicine, librarianship, and systematic reviews	Three authors with backgrounds in medicine, health policy, and HPE Acknowledged a librarian
Step 4: Identify materials for inclusion	The process is not defined, which is not uncommon or considered negatively for this synthesis type	A librarian searched multiple databases using free-text and controlled terms (MeSH) Citations managed in Excel PICO guided the inclusion criteria and incorporated studies using qualitative and quantitative methods Two reviewers assessed eligibility of full-text citations with discrepancies resolved by a third author	A librarian searched multiple databases and key HPE journals with free-text and MeSH Included qualitative and quantitative systematic reviews focused on undergraduate medical students	Librarian co-authors searched multiple databases and HPE journals using key words Citations managed in EndNote A two-stage screening process involving all authors informed inclusion of materials Included qualitative and quantitative articles relevant to the research question	The authors spoke with QI experts to identify theories and/or models A librarian searched multiple databases and key HPE journals with free-text and MeSH Two reviewers independently assessed the eligibility of articles with discrepancies resolved by consensus

(*Continued*)

BOX 31.5 (*Continued*)

	Narrative review	Systematic review	Umbrella review	Scoping review	Realist review
Step 5: Extract key data	The process is not defined, which is not uncommon or considered negatively for this synthesis type	A single author extracted data using a Cochrane checklist A second author verified the data extracted	Two authors extracted data from all full-text reviews using a data extraction sheet developed through iterative testing and revision A third author adjudicated disagreements	Developed a data charting form Two authors independently extracted data using the charting form	Created a standardised data extraction tool used by a single author
Step 6: Analyse and synthesise results	The process is not defined, which is not uncommon or considered negatively for this synthesis type	A PRISMA flow chart depicted search results and reasons for exclusion Quantitative studies assessed using Quality Assessment Tool for Quantitative Studies [43] Qualitative studies assessed using the Quality Assessment Tool for Qualitative Studies [44]	Quality assessed using AMSTAR [45] Learner outcomes assessed using a modified version of Kirkpatrick's model of educational outcomes A PRISMA flow chart depicted search results and reasons for exclusion	A PRISMA flow chart depicted search results and reasons for exclusion Methodological quality of results was not assessed as this was not a goal of this scoping review	Developed a quality assessment tool derived from the Standards for Quality Improvement Reporting Excellence Publication guidelines [46] Tested candidate conceptual frameworks and theories for fit with the literature by iteratively reviewing the articles and themes that had emerged
Step 7: Report	Author presented a summary of major findings in narrative form of what he considered essential characteristics of teachers	Results reported in narrative and tabular form Details of interventions described using the Workgroup for Intervention Development and Evaluation Research (WIDER) guidelines [60]	Results reported in narrative and tabular form with interventions organised and described as either curricular design or instructional design approaches	Results reported in narrative and tabular form with components addressing characteristics of articles, the process of giving feedback and the impact of feedback for learners	Reported themes organised as 'what works', 'for whom', 'under what circumstances' and to 'achieve what outcomes' Results reported in narrative and tabular form Presented a conceptual framework as a figure

origin, types of studies, research design, intervention, and population. The purpose of a quantitative analysis is to paint a portrait of the research activity and breadth of literature on the topic. If one is interested in the growth of research in a particular area, then tracking the year of publication of the papers could be quite useful. When the nature of the research is of greatest interest, review teams summarise the type of research and the different methods used across time.

Qualitative analyses, such as those used in scoping and realist reviews, identify the major themes emerging from the knowledge synthesis. If, for example, authors are interested in the conceptualisations of professional identity in HPE learners, a thematic analysis can be used to describe and compare ways in which professional identity is discussed and conceptualised across various health professions learners.

The synthesis and reporting of results should take into consideration the needs of end users so that the results are relevant to educators, researchers, and educational policy makers. With the increased attention to evidence-informed HPE, it is not surprising that although knowledge syntheses can be standalone pieces of scholarship, they have the potential to be vital for HPE stakeholders involved in educational decision-making. Moreover, as we witness a rise in empirical research in HPE, the community may be urged to address the applications and implications of synthesis scholarship as an accountability mechanism for those who are considered the potential users of the knowledge. It may behove the review teams and the HPE communities at large to consider how we can best use knowledge syntheses to advance HPE and ensure that we provide learners with state of the art education in order to ultimately improve patient care.

Report

HPE knowledge syntheses are generally published as journal articles and most journals accept these. However, it is important that authors check each publication's author instructions to determine the types of knowledge syntheses accepted and to verify the acceptable word limit – as knowledge syntheses often translate to lengthy manuscripts. Additionally, as knowledge syntheses should be carried out in a transparent manner to enable replication, journals are increasingly requiring that authors adhere to reporting guidelines, such as PRISMA [13] for systematic reviews or RAMESES [14] for realist reviews.

In addition to reporting knowledge syntheses through traditional outlets, such as journal articles, authors are also being encouraged to reach out to broader audiences [61]. For example, when an article is published in *Academic Medicine*, the journal editors encourage its authors to share news of the article with their institution's public affairs office and through their social media channels, such as their Twitter and Facebook accounts. By using these alternate dissemination channels, authors potentially reach academic readers in HPE, but also a much broader audience, including the public. A potential pitfall of this approach is that in some cases knowledge syntheses articles are available only to those with subscription access to the journal in which it is published or by paying a fee.

How to be a Critical Consumer of Knowledge Syntheses?

Becoming a consumer of knowledge syntheses is an emerging skill and we are unaware of any resources available to help the reader at present. With the burgeoning of HPE research, the opportunity to leverage knowledge synthesis types to enhance the continuum of theory to practice in HPE is greater than ever before. Therefore, we see creating resources and training opportunities for how to read and critically appraise knowledge syntheses as a priority area for the HPE community.

As is the case in clinical practice, using the results of a synthesis is no simple feat. It requires careful reading and appraisal of the review for its validity and usefulness. It also requires the reader and/or user to consider the evidence in light of their own context, their specific learners, the resources available (e.g. when space or new technology is needed), and the support from the leadership of the organisation, particularly if the evidence suggests modifying existing practices [62].

Consumers of knowledge syntheses may wish to ask the following questions when reading a review:

- What have I learned?
- Are the findings relevant and applicable to my context?
- How could I use the findings to inform or change my practice?
- What would I have to do to use the findings in my practice?
- With whom can I discuss the findings?

Conducting knowledge syntheses can be quite challenging for novice researchers [6]. Current training on knowledge synthesis is almost non-existent in terms of the creation of knowledge syntheses or in its critical and informed consumption or use. Relevant coursework (e.g. through formal degree programmes), workshops (e.g. through certificate programmes like the Medical Education Research Certificate [MERC]), and group discussions can help improve both the creation and use of knowledge syntheses. At a minimum, we recommend that readers review several of the articles cited in this chapter. We have found these resources helpful for both producing and consuming knowledge syntheses.

Conclusions

In this chapter, we present the characteristics of knowledge syntheses in HPE and provide details of five knowledge syntheses types that are prevalent in HPE and/or useful to HPE researchers and practitioners. We also describe seven key steps for undertaking a knowledge synthesis. As the creation and use of knowledge syntheses continues to grow in HPE, awareness of the types of knowledge syntheses and related methods are increasingly important.

Acknowledgements

We would like to acknowledge Doug Dworkin for his design and execution of Figure 31.1: *How to conduct a knowledge synthesis*.

References

1 Doja, A., Horsley, T., and Sampson, M. (2014). Productivity in medical education research: an examination of countries of origin. *BMC Medical Education* 14: 243.
2 McGaghie, W.C. (2015). Varieties of integrative scholarship: why rules of evidence, criteria, and standards matter. *Academic Medicine* 90: 294–302.
3 Grant, M.J. and Booth, A. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Information and Libraries Journal* 26: 91–108.
4 Tricco, A.C., Soobiah, C., Antony, J. et al. (2016). A scoping review identifies multiple emerging knowledge synthesis methods, but few studies operationalize the method. *Journal of Clinical Epidemiology* 73: 19–28.
5 Steinert, Y. and Thomas, A. (2016). When I say … literature reviews. *Medical Education* 50: 398–399.
6 Boell, S.K. and Cecez-Kecmanovic, D. (2015). On being 'systematic' in literature reviews in IS. *Journal of Information Technology* 30: 161–173.
7 Gordon, M. (2016). Are we talking the same paradigm? Considering methodological choices in health education systematic review. *Medical Teacher* 38: 746–750.
8 Kastner, M., Antony, J., Soobiah, C. et al. (2016). Conceptual recommendations for selecting the most appropriate knowledge synthesis method to answer research questions related to complex evidence. *Journal of Clinical Epidemiology* 73: 43–49.
9 Sharma, R., Gordon, M., Dharamsi, S., and Gibbs, T. (2015). Systematic reviews in medical education: a practical approach: AMEE guide 94. *Medical Teacher* 37: 108–124.
10 Aromataris, E., Fernandez, R., Godfrey, C.M. et al. (2015). Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. *International Journal of Evidence-Based Healthcare* 13: 132–140.
11 Arksey, H. and O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology* 8: 19–32.
12 Pawson, R., Greenhalgh, T., Harvey, G., and Walshe, K. (2005). Realist review – a new method of systematic review designed for complex policy interventions. *Journal of Health Services Research & Policy* 10: 21–34.
13 Moher, D., Liberati, A., Tetzlaff, J., and Altman, D.G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *British Medical Journal* 339: b2535.
14 Wong, G., Greenhalgh, T., Westhorp, G. et al. (2009). RAMESES publication standards: realist syntheses. *BMC Medicine* 11: 21.
15 Paré, G., Trudel, M.-C., Jaana, M., and Kitsiou, S. (2015). Synthesizing information systems knowledge: a typology of literature reviews. *Information & Management* 52: 183–199.
16 Berman, A.C. (2015). Good teaching is good teaching: a narrative review for effective medical educators. *Anatomical Sciences Education* 8: 386–394.
17 Boell, S.K. and Cecez-Kecmanovic, D. (2010). Literature reviews and the hermeneutic circle. *Australian Academic & Research Libraries* 41: 129–144.
18 Azer, S.A. (2015). The top-cited articles in medical education: a bibliometric analysis. *Academic Medicine* 90: 1147–1161.
19 Albanese, M.A. and Mitchell, S. (1993). Problem-based learning: a review of literature on its outcomes and implementation issues. *Academic Medicine* 68: 52–81.
20 Cochrane Collaboration. Cochrane Handbook. http://handbook.cochrane.org (accessed 24 March 2017).
21 About BEME website. https://www.bemecollaboration.org/About+BEME (accesed 24 March 2017).
22 Harden, R.M., Grant, J., Buckley, G., and Hart, I.R. (1999). BEME Guide No. 1: best evidence medical education. *Medical Teacher* 21: 553–562.
23 Reeves, S., Fletcher, S., Barr, H. et al. (2016). A BEME systematic review of the effects of interprofessional education: BEME Guide No. 39. *Medical Teacher* 38: 656–668.
24 Querido, S.J., Vergouw, D., Wigersma, L. et al. (2016). Dynamics of career choice among students in undergraduate medical courses. A BEME systematic review: BEME Guide No. 33. *Medical Teacher* 38: 18–29.
25 Ellaway, R.H., O'Gorman, L., Strasser, R. et al. (2016). A critical hybrid realist-outcomes systematic review of relationships between medical education programmes and communities: BEME Guide No. 35. *Medical Teacher* 38: 229–245.
26 Borah, R., Brown, A.W., Capers, P.L., and Kaiser, K.A. (2017). Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. *BMJ Open* 7.
27 Brydges, R., Manzone, J., Shanks, D. et al. (2015). Self-regulated learning in simulation-based training: a systematic review and meta-analysis. *Medical Education* 49: 368–378.
28 Green, C., Taylor, C., Buckley, S., and Hean, S. (2016). Beyond synthesis: augmenting systematic review procedures with practical principles to optimise impact and uptake in educational policy and practice. *International Journal of Research & Method in Education* 39: 329–344.
29 Gordon, M. and Gibbs, T. (2014). STORIES statement: publication standards for healthcare education evidence synthesis. *BMC Medicine* 12: 143.
30 Ioannidis, J.P. (2009). Integration of evidence from multiple meta-analyses: a primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. *Canadian Medical Association Journal* 181: 488–493.
31 Cant, R.P. and Cooper, S.J. (2017). Use of simulation-based learning in undergraduate nurse education: an umbrella systematic review. *Nurse Education Today* 49: 63–71.
32. Colquhoun, H.L., Levac, D., O'Brien, K.K. et al. (2014). Scoping reviews: time for clarity in definition, methods, and reporting. *Journal of Clinical Epidemiology* 67: 1291–1294.
33 Thomas, A., Lubarsky, S., Durning, S.J., and Young, M.E. (2017). Knowledge syntheses in medical education: demystifying scoping reviews. *Academic Medicine* 92: 160–166.
34. Bing-You, R., Hayes, V., Varaklis, K. et al. (2017). Feedback for learners in medical education: what is known? A scoping review. *Academic Medicine* 92 (9): 1346–1354.
35 Wong, G., Greenhalgh, T., Westhorp, G., and Pawson, R. (2012). Realist methods in medical education research: what are they and what can they contribute? *Medical Education* 46: 89–96.
36 Rycroft-Malone, J., McCormack, B., Hutchinson, A. et al. (2012). Realist synthesis: illustrating the method for implementation research. *Implementation Science* 7: 33.
37 Kehoe, A., McLachlan, J., Metcalf, J. et al. (2016). Supporting international medical graduates' transition to their host-country: realist synthesis. *Medical Education* 50: 1015–1032.
38 Tricco, A.C., Tetzlaff, J., and Moher, D. (2011). The art and science of knowledge synthesis. *Journal of Clinical Epidemiology* 64: 11–20.
39 Cook, D.A. and West, C.P. (2012). Conducting systematic reviews in medical education: a stepwise approach. *Medical Education* 46: 943–952.
40 Jones, C.A., Roop, S.C., Pohar, S.L. et al. (2015). Translating knowledge in rehabilitation: systematic review. *Physical Therapy* 95: 663–677.

41 Onyura, B., Baker, L., Cameron, B. et al. (2016). Evidence for curricular and instructional design approaches in undergraduate medical education: an umbrella review. *Medical Teacher* 38: 150–161.

42 Jones, A.C., Shipman, S.A., and Ogrinc, G. (2015). Key characteristics of successful quality improvement curricula in physician education: a realist review. *BMJ Quality & Safety* 24: 77–88.

43 National Collaborating Centre for Methods and Tool website. http://www.nccmt.ca/resources/search/15 (accessed 24 March 2017).

44 Hutchinson, A.M., Mallidou, A., and Toth, F. (2010). *Review and Synthesis of Literature Examining Characteristics or Organizational Context that Influence Knowledge Translation in Healthcare*. Alberta, Canada: Technical Report: University of Alberta.

45 Shea, B.J., Hamel, C., Wells, G.A. et al. (2009). AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *Journal of Clinical Epidemiology* 62: 1013–1020.

46 Ogrinc, G., Davies, L., Goodman, D. et al. (2016). SQUIRE 2.0 (Standards for QUality Improvement Reporting Excellence): revised publication guidelines from a detailed consensus process. *BMJ Quality & Safety* 25: 986–992.

47 Dudden, R.F. and Protzko, S.L. (2011). The systematic review team: contributions of the health sciences librarian. *Medical Reference Services Quarterly* 30: 301–315.

48 JAMA. Author Instructions website. http://jamanetwork.com/journals/jama/pages/instructions-for-authors (accessed 24 March 2017).

49 Institute of Medicine Committee on Standards for Systematic Reviews of Comparative Effectiveness Research (2011). Standards for reporting. In: *Finding What Works in Health Care: Standards for Systematic Reviews* (ed. J. Eden), 8. Washington, DC: National Academies Press.

50 Rethlefsen, M.L., Murad, M.H., and Livingston, E.H. (2014). Engaging medical librarians to improve the quality of review articles. *JAMA* 312: 999–1000.

51 Rethlefsen, M.L., Farrell, A.M., Osterhaus Trzasko, L.C., and Brigham, T.J. (2015). Librarian co-authors correlated with higher quality reported search strategies in general internal medicine systematic reviews. *Journal of Clinical Epidemiology* 68: 617–626.

52 Morris, M., Boruff, J.T., and Gore, G.C. (2016). Scoping reviews: establishing the role of the librarian. *Journal of the Medical Library Association* 104: 346–354.

53 Golder, S., Loke, Y.K., and Zorzela, L. (2013). Some improvements are apparent in identifying adverse effects in systematic reviews from 1994 to 2011. *Journal of Clinical Epidemiology* 66: 253–260.

54 Haig, A. and Dozier, M. (2013). BEME guide no. 3: systematic searching for evidence in medical education – part 2: constructing searches. *Medical Teacher* 25: 463–484.

55 Haig, A. and Dozier, M. (2013). BEME guide no 3: systematic searching for evidence in medical education – part 1: sources of information. *Medical Teacher* 25: 352–363.

56 Maggio, L.A., Tannery, N.H., and Kanter, S.L. (2011). AM last page: how to perform an effective database search. *Academic Medicine* 86: 1057.

57 McGowan, J., Sampson, M., Salzwedel, D.M. et al. (2016). PRESS peer review of electronic search strategies: 2015 guideline statement. *Journal of Clinical Epidemiology* 75: 40–46.

58 Dang, T.M. and Maggio, L.A. (2017). Supporting the call to action: a review of nutrition educational interventions in the health professions literature and MedEdPORTAL. *Academic Medicine* 92: 403–416.

59 Maggio, L.A. and Kung, J.Y. (2014). How are medical students trained to locate biomedical information to practice evidence-based medicine? A review of the 2007–2012 literature. *Journal of the Medical Library Association* 102: 184–191.

60 Meisel, Z.F., Gollust, S.E., and Grande, D. (2016). Translating research for health policy decisions: is it time for researchers to join social media? *Academic Medicine* 91: 1341–1343.

61 Albrecht, L., Archibald, M., Arseneau, D., and Scott, S.D. (2013). Development of a checklist to assess the quality of reporting of knowledge translation interventions using the Workgroup for Intervention Development and Evaluation Research (WIDER) recommendations. *Implementation Science.* 8 (1): 52. doi: 10.1186/1748-5908-8-52.

62 Gordon, M., Carneiro, A.V., and Patricio, M.F. (2015). Enhancing the impact of BEME systematic reviews on educational practice. *Medical Teacher* 37: 789–792.

Further Reading

Booth, A., Sutton, A., and Papaioannou, D. (2016). *Systematic Approaches to a Successful Literature Review*. Los Angeles, CA: Sage Publications.

Cooper, H. (2016). *Research Synthesis and Meta-Analysis: A Step-by-Step Approach*. Los Angeles, CA: Sage Publications.

Pope, C., Mays, N., and Popay, J. (2007). *Synthesising Qualitative and Quantitative Health Research*. London: Open University Press.

Part 5

Faculty and Learners

32 Career Progression and Support

Caroline Elton[1] *and Nicole J. Borges*[2]
[1]Career Planning for Doctors and Dentists, London, UK
[2]Department of Neurobiology and Anatomical Sciences, University of Mississippi Medical Center, Jackson, MS, USA

KEY MESSAGES

- The effectiveness of careers support is enhanced if both the provider and the recipient work from a shared framework.
- A four-stage model of self-assessment, career exploration, decision-making, and implementation planning provides an effective framework for the provision of careers support.
- Contemporary practice highlights the potential limitations of offering directive careers advice. Instead, a more facilitative stance is favoured.
- The literature on personality and medical specialty choice suggests that there is more variation in personality within than between different specialties.
- Any psychometric measurement should be complemented by qualitative self-assessment activities.
- Career planning skills are not only useful to guide specialty choice, but are essential throughout a medical career, up to and including retirement.

Introduction

Drawing on medicine-specific and more general careers research, this chapter provides an evidence-based approach to the provision of careers support to medical students and doctors in postgraduate training. Some of the different career decisions that one might encounter over the length of a medical career are described and a practical model for the provision of careers support is outlined. The literature on the role of psychometric testing is examined in detail in order to make recommendations on how best to incorporate psychometric test results into a career discussion. Some of the potential limitations of adopting an overly directive approach to the provision of careers support are described. The chapter works from the premise that for the majority of students and trainees, the provider of careers support will be a senior clinician, but situations in which a referral to a specialist careers service is indicated are also discussed. Policies and practices to which the chapter refers are drawn predominantly from the UK and USA context. Throughout, the argument is made that medical careers research and practice needs to draw more heavily on contemporary developments in vocational psychology.

Need for Support in Medical Careers

Until the introduction of a raft of major reforms in postgraduate medical education in 2005, designated careers support services for doctors in training in the UK were not widely available. Even since the introduction of *Modernising Medical Careers* (MMC) trainees have had to make significant career decisions about specialty choice less than 18 months into their postgraduate training [1]. Yet it is known that prior to MMC, many trainees had not yet chosen their final specialty so soon after leaving medical school [2–4], with women being less likely to be confident about early specialty choices than men [5]. In response to this new, condensed career trajectory, the need for appropriate careers support services was recognised, and subsequently embedded into the reforms. Today, all regions in the UK have established careers support services for trainees, and all of the UK medical schools now have a named careers adviser, often linked to a university careers service [6].

In the USA, where the majority of medical schools have postgraduate entry, medical students choose their specialty on exit from medical school. Since 1999 the Association of American Medical Colleges (AAMC) Careers in Medicine web-based resource offers online help for students in career planning, as well as for training faculty in how best to use the resource with their students [7]. However, earlier studies suggested minimal careers advice provision in American medical schools [8].

In part, the reason for the earlier lack of careers support offered during medical training may be an unstated assumption that a medical degree is a vocational training so career choices have already been made, rendering further careers support superfluous. Yet in all medical training systems it is clear that this assumption is unjustified as career-related decisions are made from the point of

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

application to medical school. For example, medical schools in the UK differ in the ways in which they integrate the clinical and non-clinical training, in the extent of exposure to primary care, in opportunities for studying for an intercalated degree, and in the range of student-selected components. Studies have indicated that factors such as opting for an intercalated degree [9, 10], participating in research [11], or studying medicine as a postgraduate [12] can have an impact on final career decisions. In the USA, medical students choose their specialty during the final year prior to graduating. As a result, the final year is largely comprised of electives and is regarded as an important part of testing one's suitability for an intended specialty and also as a way of improving one's chance of a given residency application being successful. So, it is clear that medical students are making career-related decisions during medical school.

Whether decisions about specialty choice are made during medical school (as in the USA and Canada) or after completing a generic trainee programme (as in the UK and Australia), there are a large number of different specialties from which to choose. For example, in the UK, on completion of the Foundation Programme, trainees have a choice of 20 different training programmes, as well as deciding if they want to apply for an academic training pathway which would allow them to combine clinical training with opportunities to complete a research degree. In subsequent years further specialty options branch out from these 20 different post-foundation options and at the time of writing there are 66 specialties from which to choose (with opportunities for further sub-specialisation in 32 sub-specialties) [13]. In the US there are 37 specialties which then branch out into 87 sub-specialties [14], whilst in Australia there are 85 specialties but no sub-specialties [15].

It is also apparent that career decision-making does not end once the specialty/sub-specialty choice phase has been successfully navigated. For example, decisions have to be made about continuing with research, or fellowship training, about whether to take a leading role in clinical management or clinical education, or whether to switch to part-time work. Seen in this way, medical career decision-making is a process that begins before entry to medical school and continues up to the point of retirement.

Provision of Effective Careers Support

Concerns about the quality of careers support that doctors receive are not new [16–19]. More recently, but prior to the implementation of the UK MMC reforms, a national survey on the availability and quality of careers support within medical training recommended that there was a need for high-quality careers information, for the development of self-assessment and career planning tools, for trained advisers who could provide expert careers advice, and for national coordination of the careers support available across the UK [20]. Six years later a further survey reported that at least in terms of medical school provision, there was little evidence that these recommendations had been implemented [6].

MMC recognised that the previous inattention given to the provision of careers support for doctors needed to be remedied, and for the first time in the UK a comprehensive careers support strategy for medical students and trainees was produced [21]. This report also recognised that certain groups of doctors, such as those with childcare responsibilities, international medical graduates, and doctors with significant health issues, may need specialist careers support.

However, it could also be argued that this initiative was something of a missed opportunity. Given the absence in the policy of any linkage to an underlying theoretical framework of career support, there was no understanding embedded in the recommendations that it can be helpful if both the providers and recipients of careers support share a common framework for the overall career planning process. Yet a large-scale study has clearly demonstrated that sharing an underlying framework enhances the effectiveness of career discussions [22]. This lack of an underpinning theoretical framework also meant that insufficient attention was given to the fact that when providing support for career planning, some tasks come before others and it is important to map these different tasks across the length of the medical training continuum. In effect MMC was perhaps characteristic of the tendency highlighted by Petrides and McManus for developments in medical careers support to become divorced from broader developments in occupational psychology [23].

Who Should Provide Careers Support?

In the UK, national guidance for postgraduate training recommends that the responsibility for the provision of careers support should fall within the remit of the senior clinician supervising the trainee. It is also recognised that some students/trainees with more complex career needs (e.g. those with significant health concerns) may need specialist input from trained careers advisers or occupational psychologists.

This notion that, in the main, careers support will be provided by senior clinicians tallies with the trainees' own expectations. For example, in a survey of specialty trainees, Lloyd and Becker reported that study participants looked for careers advice and support from their educational supervisors, rather than from careers professionals [24].

Improving the Quality of Careers Support

In the non-medical context, an extensive survey of the components of effective career discussions in the workplace identified the importance of both parties having a simple shared framework to structure the discussion [22]. This finding has been incorporated into the approach to careers support described below and the specific framework is the four-stage model of careers guidance that is used throughout the higher education sector, namely:

1 self-assessment
2 career exploration
3 decision-making
4 plan implementation.

Evidence for this approach comes from an international review of best practice in career development [25]. Health Workforce New Zealand [26] and the 'Careers in Medicine' programme devised by the Association of American Medical Colleges, and used throughout medical schools in the USA, also use the four-stage approach.

When using this approach in the medical education context, it is also possible to draw a parallel between career decision-making and clinical decision-making (see Box 32.1).

A number of important points can be drawn from this parallel. First, it would make no clinical sense to start with the treatment plan, then formulate a diagnosis and continue working backwards to the history; it is just so with career decision-making where it is not helpful to concentrate on the details of implementing one's career plan if one has not done adequate preparatory work in terms of self-assessment or exploring different options. There is of course one exception and that is in the context of a clinical emergency where one may have to implement a treatment plan before formulating the diagnosis, carrying out investigations, or taking a history. But here too there is a career correlate in that sometimes the overriding need is to get any job as a holding position, before working out one's career in the longer term, which requires self-assessment, career exploration, etc.

A further point highlighted by the parallel is that in both clinical and career decision-making, the first two stages are linked. The particular details of the patient history will inform the examination/investigation stage. Similarly, the results of the self-assessment stage inform not only the specific career options that are explored in greater depth, but also the specific questions that should be researched. Furthermore, with both career decision-making and clinical decision-making there may be to-ing and fro-ing between the first two stages: clinically something found on examination may necessitate asking a more detailed history on certain points; in career decision-making, something that the person discovers when they are exploring a particular option might lead them to go back and rethink the self-assessment stage.

However, whilst the parallel is useful in highlighting a systematic approach to career planning, it also has limitations. In clinical decision-making the aim (although it is not always achieved) is to make a definitive diagnosis. In career decision-making, the notion of a definitive diagnosis can be unhelpful if it is taken to suggest that each doctor could be happy in only one specialty. Instead, the literature suggests that each doctor is likely to be suited to more than one particular specialty [27].

BOX 32.1 The parallel between clinical decision-making and career decision-making

Clinical decision-making	Career decision-making
Taking a history	Self-assessment
Examining the patient/ investigations	Career exploration
Formulating a diagnosis	Decision-making
Implementing the treatment plan	Implementing the career plan

Specialty Choice

A vast literature spanning decades of research exists on factors influencing medical specialty choice. Hutt identified six areas on which the literature on career choice in medicine has focused: background, personality and attitude, educational system, career, working conditions, and intrinsic differences within specialties [28]. Other factors associated with specialty choice cited in the literature range from internal, personal, or individual characteristics, including personality, values, and interests [29, 30], to external factors related to lifestyle [31, 32], such as work hours and income, clinical experiences during medical school [33, 34], and exposure to positive and negative role models [35–37].

The concept of a 'controllable lifestyle' has been referenced for several decades but can be traced back to the work of Schwartz et al. in 1989 [32]. Recent trends in this literature [38] illuminate lifestyle factors as major influences in the current generation's choice of medical specialty. In the USA, controllable lifestyle specialties, such as anaesthesiology, dermatology, emergency medicine, pathology, psychiatry, and radiology, are on the rise as popular medical specialties and as a result it is increasingly competitive for students to gain residency in these specialities. Specialties referred to as 'uncontrollable' in lifestyle include surgery as well as primary care specialties of family medicine, internal medicine, and paediatrics. Currently these specialties may be less competitive and all residency match slots may not fill.

A recent study carried out by the UK Royal College of Physicians [39] concluded that women were drawn to choosing specialties that offered a more controllable lifestyle. These authors also point out that given that the number of female medical students exceeds that of male students, the issue of shying away from specialties that have considerable out-of-hours commitments is likely to have a significant impact on future medical workforce planning.

In considering this area, two further points must be made. First, due to the ways in which a given specialty is practised in different health care systems, a career choice such as family medicine may be perceived as offering poor control of lifestyle in the USA, but good control in the UK. Second, research has emphasised that the issue of controllable lifestyle is not only of importance to female doctors [40, 41].

More recently researchers in the USA have investigated the relation between medical school debt to choice of specialty, yielding mixed results. Phillips et al. found no overall relationship between anticipated debt upon graduating from medical school, but did find that medical students from middle-income households who anticipated more debt were less likely to consider a career in a primary care specialty [42]. Rosenblatt and Andrilla found demographic factors mediated their finding that medical students with more debt were less likely to enter primary care specialties [43]. In their longitudinal study (data from 1997 to 2006), medical students with higher debt were less likely to choose internal medicine and paediatrics [44]. The issue of the impact of student debt on specialty choice may become increasingly important in the UK context given the rise in university fees that took place in 2011.

Much of the literature on specialty choice has focused on students choosing primary care versus non-primary care careers, and researchers such as Fincher et al. [45], Bland et al. [46], and Senf et al. [47] have made important contributions to help medical educators understand the factors related to primary care specialty choice. In particular, the theoretical model developed by Bland et al., although over 15 years old, contains many of the factors related to specialty choice that researchers continue to investigate today [46]. Caution is needed, however, when applying research findings from one national context to another because of differences in the way specialties are considered, e.g. paediatrics and internal medicine are regarded as primary care specialties in the USA but secondary care ones in the UK.

Although much of the literature focuses on factors related to specialty choice and how to predict medical specialty choice, a more novel approach comes from the work of Reed et al., who posit using decision theory to understand medical specialty choice and focus on the *process* of medical specialty choice decision-making rather than the actual content of the decision [48]. While theoretical, these models have applicability for careers professionals as they help highlight the complexity of medical specialty decision-making and may thereby normalise the experience of a student/trainee who is experiencing difficulty with making a decision. In addition, these models highlight the fact that career decision-making is a developmental task that unfolds over time, and there is therefore a need to prepare students and trainees over a period of time, rather than just before the point when they have to make a specialty choice decision.

One might predict that after decades and decades of research the factors associated with choosing a specialty would be well articulated and no longer require further investigative inquiry. To borrow the words of Barzanky, 'What much research has revealed, however, is that the simplicity of the question is deceptive' [49]. Researchers would agree that understanding medical specialty choice is a complex process affected by a host of factors, and that the factors vary from individual to individual. This contributes to the difficulty in synthesising and summarising the literature.

While individual factors and specialty choice decisions have been closely linked, Leduc and colleagues [50] suggest contextual influences in the health care system that can shape specialty choice, including population needs and health care policy changes. In the US, the Affordable Care Act with projections indicating a need for a larger primary care workforce and the graduate medical education crisis caused by an increasing number of medical school graduates but a limited number of medical residency slots across specialties are examples [51].

Psychometric Testing

In the USA where specialty choices have to be made by the end of the fourth year of medical school, a number of medical schools offer psychometric testing to assist medical students with their specialty choice decisions. A review of the relevant literature carried out by Borges and Savickas [27] suggested that the Myers–Briggs Type Indicator (MBTI) [52] was the most frequently used psychometric instrument in helping medical students/trainees choose their specialty. However, other researchers have constructed medicine-specific indicators, such as the Medical Specialty Preference Inventory [53, 54], developed in the USA, and Sci45/Sci59 [55], which is a specialty choice inventory developed in the UK.

The Myers–Briggs Type Indicator

The Myers–Briggs Type Indicator (see Box 32.2) is a self-report measure of normal personality that assesses differences in how people perceive information and the differences in how they use that information [52]. The inventory assesses the individual's preferences on four dichotomous scales: Extraversion/Intraversion (E/I), Sensing/Intuition (S/N), Thinking/Feeling (T/F), and Judging/Perceiving (J/P). In this way each individual's personality can be categorised into one of 16 different four-letter 'types' (e.g. ESTJ; ISTJ, etc.).

BOX 32.2 FOCUS ON: The Myers–Briggs Type Indicator [52]

Extraversion/Intraversion (E/I)

People who prefer *Extraversion* tend to focus on the outer world of people and external events.

People who prefer *Introversion* tend to focus on their own inner world of ideas and experiences.

Sensing/Intuition (S/N)

People who prefer *Sensing* like to take in information through their senses to find out what is actually happening in the present.

People who prefer *Intuition* like to take in information by seeing the big picture and grasping patterns.

Thinking/Feeling (T/F)

People who prefer to make judgements using *Thinking* tend to look at the logical consequences of a choice.

People who prefer to make judgements using *Feeling* tend to consider what is important to them and other people.

Judging/Perceiving (J/P)

People who prefer to orientate themselves to the outer world using *Judging* live in a planned orderly way.

People who prefer to orientate themselves to the outer world using *Perceiving* live in a flexible, spontaneous way.

Since the 1950s, Myers and others have researched whether there are consistent relationships between MBTI personality type and the choice of particular medical specialties [56–59]. Based on these studies, some American career resources, such as *The Ultimate Guide to Choosing a Specialty* [60] and the Careers in Medicine website [7], give information about MBTI codes that are frequently associated with particular specialties. So, for example, Freeman [60], using data from McCaulley [56], lists specialties frequently associated with 'ESTJ' as obstetrics/gynaecology, general practice, general surgery, orthopaedic surgery, and paediatrics.

An initial problem associated with the use of personality type to choose a medical specialty thus immediately becomes apparent: this list of specialties commonly associated with this one particular MBTI type covers a broad range of different specialties, spanning both primary and secondary care and surgical/non-surgical specialties. Borges and Savickas [27], in a comprehensive review of the relevant literature, highlight two further points. First, links between personality and specialty type do not necessarily remain constant over time. For example, there are some American data to suggest that different personality types were attracted to family medicine in the 1970s compared to those in the 1980s. This finding strongly suggests that one needs to be cautious about over-reliance on suggested links between MBTI preferences and specific specialties based on data that may have initially been collected over 30 years ago.

The second point to emerge from the Borges and Savickas review is that there is actually more variation in personality type (using both the MBTI and other personality measures) *within* each specialty than *between* specialties [27]. They therefore conclude that all personality types appear in all specialties (although some types are more common than others) and that more than one particular specialty will fit the personality of any particular medical student or doctor. Similarly, a German study published in 2010 found significant differences in personality between board-certified internists and surgeons [61].

However, on the basis of their review, Borges and Savickas do not conclude that personality assessment has no role in supporting trainees through the process of specialty choice. Instead, they argue that personality assessment should be viewed as one of a number of different factors that trainees should consider when choosing a specialty. Furthermore, they recommend that if a personality test such as MBTI is carried out, it should be used to increase self-knowledge (see Box 32.2) rather than used as a simple diagnostic process that makes a link between a particular personality type and a particular specialty.

Medical Specialty Preference Inventory

The original version of the Medical Specialty Preference Inventory (MSPI) [53] was a 199-item questionnaire that assessed medical students' interest in 40 areas of medical practice and preferences for 6 specialties: family practice, internal medicine, obstetrics and gynaecology, paediatrics, psychiatry, and surgery. Whilst Zimny [54] described predictive validity of the instrument to be in the region of 50%, Savickas et al. [62] found predictive validity indices in the range of 59% and Glavin et al. found that the MSPI predicts medical specialty choice 58% of the time [63]. In a more recent study, however, Borges et al. reported that the questionnaire correctly predicted the preferred specialty in only 33% of their sample [64]. In addition, these latter authors also reported that nearly half of their sample (47%) chose specialties not listed in the six specialty fields included in the MSPI. So, clearly both the accuracy of prediction and the breadth of specialties included in the inventory are problematic.

The MSPI has undergone revisions since its original version. In 2009 Sodano and Richard reduced the MSPI down from 38 to 18 factors [65]. Additional research in 2010 supported an expanded number of specialty scales [66]. The Revised MSPI [67] consists of a 150-item scale predicting specialty choice in 16 specialties, and reports two additional studies in which predictive validity were 52% and 46%. The 18 Medical Interest Scales relate interest in different activities to medicine. However, caution is needed when using the instrument with students/trainees outside of the USA because aspects of a particular specialty (e.g. typical patterns of working hours or financial rewards) may differ significantly in different countries.

Sci59 Specialty Choice Inventory

The Sci59 Specialty Choice Inventory is a 130-item questionnaire designed in the UK to help medical students and trainees with the task of choosing an appropriate specialty. The first version of the questionnaire included 45 specialties [55], but the current version now contains additional options, hence the change of name from Sci45 to Sci59. Having completed the questionnaire, respondents are provided with a computer-generated printout that lists the 10 specialties to which there is greatest fit and the 10 to which there is the least fit. In addition, the printout also contains a graph showing how the respondent has scored on 12 different underlying subscales, such as 'action orientation', 'coping with uncertainty', 'need for assertiveness', etc.

Sci59 is currently widely available and, in addition to access through the British Medical Association (BMA) website for BMA members, a number of medical schools also offer the instrument. It is therefore of more than academic interest to provide a critical appraisal of the use of Sci59 for helping medical students or trainees with the task of identifying an appropriate specialty.

In terms of the instrument design, although the original list of items considered for inclusion in the questionnaire was drawn from interviews with consultants and principals in general practice (GP), when drawing up the final list of items, the questionnaire was actually calibrated using responses from trainees who had not yet embarked on higher specialty training. The authors point out that the validity of the scale therefore rests on the assumption that the trainees 'had insight into the properties of work in their specialty and of consultant or principal posts and into the skills needed to occupy them' [55]. Yet these authors provide no evidence of the validity of this fundamental assumption (i.e. that a junior's understanding of the nature of work as a consultant or GP principal accords closely with that of the consultant or GP principal who actually occupies these posts).

Another concern is that of the predictive validity of the questionnaire. The original paper in which the design of the questionnaire is described [55] does not include data on predictive validity, and there is only one report in a peer-reviewed journal that examines this question [68]. This was a study of Foundation Year 2 (F2) doctors who filled out the Sci45 questionnaire within the first two months of their F2 year, and then again in the last two months of the programme, and it reported that 30% of the doctors were successfully appointed to a specialty predicted by their initial Sci45 scores. However, given the design of the study, it is impossible to tease out the different possible explanations for this low predictive validity, such as the questionnaire failing to identify the specialties that most interested the trainee; the interests of the trainee changing over the F2 year; and the trainees applying for, but not being accepted into, the specialties predicted by the Sci45 questionnaire.

Using Psychometric Test Results to Help with Specialty Choice

Given the complexity of factors that may impact on specialty choice, it is not surprising that studies using instruments such as Sci59 or the Revised MSPI often report relatively low predictive validity. In turn, given this low predictive validity, it is clear that caution is needed when using test results to help a medical student or trainee choose his or her future career. In other words, the fact that a particular career is listed as appropriate does not 'prove' that the respondent is well suited to that particular specialty. The appropriate use of any set of test results is to view it as a stimulus for further discussion. Examples of the types of issues that a supervisor could discuss with a medical student or trainee are given in Box 32.3. Used in this way, psychometric testing undoubtedly has a role to play in helping the trainee consider possible implications of aspects of their personality, offering reassurance that they might be suited to an emerging area of interest or broadening the range of specialties that they might research further.

BOX 32.3 HOW TO: Ask questions that explore the implications of psychometric test results

- What have you learnt about yourself from completing the particular questionnaire?
- To what extent do the results accord with the results of assessments that you have carried out during your training, or feedback you have received from your supervisors?
- Are there any surprises: How do you account for these?
- Which specialities might you be interested in researching further on the basis of the questionnaire results?

Deciding on a Career in Academic Medicine

Another major area of concern in the medical careers literature is the question of what factors predict whether individuals will choose an academic career pathway. In the USA, individuals typically first decide on the vocation of medicine, then make a specialty choice, followed by a decision about whether they wish to pursue an academic pathway. In the UK, it is possible to pursue the academic route alongside medical training (in the small number of MBBS/PhD programmes), as part of foundation training, during the early years of post-foundation training or at some later point.

In the USA, Strauss et al. sought to understand factors impacting academic medicine as a career choice [69]. Their systematic review revealed the following as influential in pursuing a career in academic medicine: (i) completion of a graduate degree or fellowship in addition to obtaining a MD degree; (ii) being involved in research and publishing while a student or resident; and (iii) an interest in teaching and/or the intellectual stimulation that the academic career path provides. An academic role model or mentor was also found to have a significant influence on decision-making. Additional factors cited in the literature include perceived status of a career in academic medicine [70], work–life balance [71], and autonomy [72].

Satisfaction with academic medicine is a growing area of concern in the literature. In their study of medical school faculty, Lowenstein et al. [73] found that 42% were considering leaving academic medicine. More recently, another study suggested that about one-quarter of current faculty from 26 US medical schools have considered leaving their careers in academic medicine [74]. They identify factors associated with dissatisfaction (i.e. incongruence in values, lack of institutional support), as well as factors that were unrelated (i.e. gender, faculty rank, and lack of mentoring). In the UK, lack of mentoring has been identified as one of a number of factors that contribute to the under-representation of women in senior academic positions, alongside the difficulty pursuing an academic career when working part-time or flexibly [75].

The literature on careers in academic medicine tends to be mostly focused in specialty areas rather than a collective assessment of this pathway in general. Two recent articles, however, sought to summarise the literature and address the question of choice of academic medicine as a career pathway [76, 77]. Findings by Borges et al. [77] suggest that, at least for women, a career in academic medicine tended to happen serendipitously and, as medical students, women knew little if anything about this career path. The authors suggest that medical schools and residency programmes provide specific programming and opportunities to expose their trainees at various stages of their education to academic medicine as a career path.

Contemporary Developments in Vocational Psychology

The vast literature on specialty choice demonstrates the breadth of factors that influence medical specialty decisions. The conceptual models currently available in the medical education literature are primarily focused on factors influencing specialty choice and largely were developed by medical educators. While they serve an important purpose, they may be limited in scope and fail to address the complexity of medical specialty decision-making. Future work in the field may benefit from incorporating developments in vocational psychology, in order to guide both research and careers interventions. As an example of the vocational psychology literature, Rogers et al. [78] have applied Social Cognitive Career Theory (SCCT) to the development of a measure of specialty choice. The SCCT model of career decision-making [79] proposes that personal, contextual, and experiential factors are responsible for shaping experiences that lead to self-efficacy beliefs, outcome expectations, and career goals. In turn, these beliefs and expectations underpin career interests and choices. Even though SCCT is not specifically focused on medical career and specialty choice, the complexity of the theory offers multiple pathways to explore significant factors in career planning with medical students.

Currently in the wider world of vocational psychology there is also considerable interest in bringing together qualitative and quantitative forms of career assessment. So, for example, in a key article in a leading vocational psychology journal, Walsh advocated expanding career assessment beyond traditional psychometric methods 'to consider idiographic, qualitative, and other creative approaches to assessing multiple aspects of both people and contexts' [80]. In the American context, researchers are already exploring the possible relevance of qualitative assessment methods in assisting medical students with career choice, to sit alongside more traditional quantitative psychometric methods [81].

Some UK medical schools have developed excellent career handbooks that encourage students to reflect on their potential suitability for particular specialties as they complete their undergraduate 'firms' in that speciality. Other methods of qualitative assessment in medical school include a compulsory reflective exercise on career planning linked to the undergraduate elective [82]. Undoubtedly these and other comparable initiatives are crucially important, and they are congruent with the recommendation for a greater emphasis on career planning in the undergraduate years [83].

At the postgraduate level in the UK, some foundation schools use resources such as 'Windmills' [84] or 'ROADS to Success' [85], whilst in the USA, the Careers in Medicine resource incorporates both objective psychometric testing and more subjective qualitative assessments [7]. Navarro, Taylor, and Pokorny have also described three innovative ways of incorporating Careers in Medicine resources into the undergraduate curriculum [86].

One further trend in vocational psychology that could usefully be incorporated into medical careers support is the switch from methods that concentrate on helping individuals make a one-off career decision, to an approach that emphasises the inevitability of change [87, 88]. Seen in this way, the task of careers support is to equip individuals with career planning skills so they can become more adept at navigating their way through the decisions that they will face throughout their career.

Helping Those Whose Career Plans Appear Unrealistic

This aspect of the supervisor's role could be paraphrased as 'how to help the student/trainee make best use of the feedback that they have received in order to develop realistic career plans'. In effect the supervisor's role in this situation is to facilitate the trainee's reflection on the *implications* of the feedback for their career plans. So, for example, if an individual wants to pursue a surgical career yet feedback from relevant work-based assessments does not indicate particular surgical aptitude, what careers advice should the trainee be given?

Whilst there is a paucity of literature on this question in the medical education field, it is an issue that is described at length in the generic careers counselling literature [89, 90]. This broader literature would suggest that the supporting clinician should focus on posing the sorts of questions outlined in Box 32.4 and avoid directive advice. We caution against giving such advice for a number of reasons. First, telling somebody what career to follow removes responsibility for the decision from the person. This is contradictory to the educational ethos of most medical training curricula which emphasise the importance of developing the trainee's responsibility for their own learning and professional development. Second, directive advice – 'if I were you, I would ditch surgery' – is not the most effective way of getting somebody to take on board the key points. There are parallels here with the health behaviour literature, which consistently reports a high degree of non-compliance with directive advice. Third, in giving directive advice it is very

BOX 32.4 HOW TO: Ask questions of those whose career plans may be unrealistic

- Based on the feedback you have received in this post (and previously), what do you see as your key strengths?
- Based on the feedback you have received in this post (and previously), what do you see as the areas that you find more difficult?
- What are the key competences that will be assessed for recruitment into your specialty of interest?
- Who have you talked to/what have you read in order to assess the competitiveness of the particular specialty you are interested in?
- If the specialty is highly competitive and you have been getting average or below average assessments in relevant areas of your work, what gives you confidence that your career plans are realistic?
- Going back to review the areas of performance that you seem to perform better in, are there other specialties that might make a better match? Have you also looked at the specialties that in previous years seemed to be shortage specialties? Might any of these specialties suit your particular strengths?

easy for personal preferences to seep in inadvertently. For example, it is easy to omit options in which one is not particularly interested, or of which one has little knowledge. Finally, the person giving the directive advice may be in error. For example, the trainee may know that something in their private life is having a seriously detrimental effect on the quality of their work, but they do not want to discuss this with their supervisor. The trainee, however, knows that previously they have received excellent feedback on areas of their work that are currently causing concern.

It is for these sorts of reasons that professional training courses in careers support suggest that careers practitioners should be wary of giving directive advice about which career paths somebody should follow. A wariness about advice is also congruent with the conclusion reached by Woolf and McManus advocating that the senior clinician should concentrate on 'listening to the needs of students and trainees, understanding their points of view, and encouraging them to make their own decisions' [91].

Other Sources of Careers Support

A senior clinician who is trying to provide careers support to a student or doctor in postgraduate training whose career plans they believe to be unrealistic might feel that the suggestions made above are insufficient. Where clear constructive feedback has been given but the trainee insists on sticking to their plans, the educational supervisor can suggest that the trainee discuss their career plans with a more senior member of the educational faculty who has overall responsibility for the training programme. In this situation, the educational supervisor can write a brief summary outlining their concerns about the robustness of the trainee's proposed plans and give a copy to the trainee and to the person providing the additional careers support.

The suggestion can also be made that the trainee could consult with somebody from the local medical careers advisory service. In the UK people occupying these roles are qualified careers professionals, and therefore have had relevant training in how to approach delicate careers discussions. In addition, some trainees may find it easier to be more open with careers advisers (as opposed to the clinicians who supervise them) about their career concerns, as careers advisers are not involved in carrying out assessments or writing references, or sitting on future interview panels. But here too, the trainee may persist in holding on to their intended plan, which might in the long run turn out well – or it might not. Just as all patients cannot always be helped from making poor health decisions (continuing to smoke, drink too much, etc.), providers of careers support need to be aware that it is not always possible to stop someone from making poor career decisions [92].

Leaving Medicine

Senior clinicians may also be uncertain how to respond when a student or trainee for whom they are responsible tells them that they are considering leaving the profession. Concern about doctors leaving the profession is not new; in 1997 an editorial in the *Journal of the Royal Society of Medicine* was entitled 'Why do young doctors leave the profession?' [93]. However, what is striking is the fact that despite reports of disillusionment amongst trainees, the actual proportion of doctors leaving the profession is remarkably low. This was a conclusion reached in the 1997 editorial mentioned above and a review of more recent studies would concur with this position. For example, studies undertaken by the UK Medical Careers Research Group in Oxford suggest that, although quite high levels of dissatisfaction can be expressed, the actual numbers of UK medical graduates intending to leave the profession are low; less than 3% in one study [94] and less than 1% in another [3].

One possible explanation for the anxiety about loss to the medical workforce despite the figures suggesting that actual rates of leaving the profession are remarkably low could be the difference between expressed rates of dissatisfaction and actual rates of leaving medicine. Rittenhouse et al. reported that although physician dissatisfaction had a strong association with expressed intention to leave clinical practice, it was not associated with actual departure from practice [95]. These authors therefore concluded that self-reported intention to leave practice may be more of a proxy for dissatisfaction than a reliable predictor of actual behaviour.

In terms of underlying reasons for leaving the profession, a qualitative follow-up study by the BMA of 14 doctors who had left the profession reported that the main reasons were that they felt that they were not valued or supported and that medicine involved an unacceptable work–life balance [96]. It would seem that there is a need for further studies to explore these issues in more depth, but these preliminary findings suggest that these three issues could usefully be explored when faced with a trainee who reports that they are considering switching career. This BMA study also reported high levels of distress amongst the sample, together with regrets that they had been unable to access adequate careers support. This might suggest that students or doctors who are considering leaving the profession should be made aware of professional careers support services that could assist them with the major life decision to leave the profession or to remain working as a doctor.

Conclusions

Students and trainees tend to look to the senior clinicians who supervise them for careers support. The approach outlined in this chapter suggests that caution is needed when giving directive careers advice. Instead, the recommended model is based on a structured approach to careers support that emphasises the importance of thorough self-assessment (using both quantitative and qualitative methods) and careful exploration of options. Furthermore, career decision-making is not over when a specialty/sub-specialty choice has been made, but instead will re-occur up to the point of retirement, and the careers support model can be used at any point in one's working life. Some students or trainees whose career decisions are complicated by health or performance issues, or who are considering leaving medicine entirely, may benefit from a referral to a specialist careers support service. Historically, research, and interventions in the medical careers field have tended to be divorced from wider developments in vocational psychology and closer collaboration between the two disciplines is recommended.

References

1 Department of Health (2004). *Modernising Medical Careers: The Next Steps*. London: Department of Health.
2 Edwards, C., Lambert, T.W., Goldacre, M.J. et al. (1997). Early medical career choices and eventual careers. *Medical Education* 31: 237–242.
3 Lambert, T.W., Goldacre, M.J., and Turner, G. (2006). Career choices of the United Kingdom medical graduates of 2002: questionnaire survey. *Medical Education* 40: 514–521.
4 Goldacre, M.J., Laxton, L., and Lambert, T.W. (2010). Medical graduates' early career choices of specialty and their eventual specialty destinations: UK prospective cohort studies. *British Medical Journal* 341: c3199.
5 McManus, I.C. and Goldacre, M.J. (2008). Predicting career destinations. In: *Education and Training: From Theory to Delivery* (ed. Y. Carter and N. Jackson), 59–78. Oxford: Oxford University Press.
6 Tapper-Jones, L., Prout, H., Grant, A. et al. (2009). An evaluation of the sources and availability of careers advice in UK medical schools. *British Journal of Hospital Medicine* 70: 588–591.
7 Association of American Medical Colleges (2013). Careers in Medicine. http://www.aamc.org/students/cim (accessed 3 February 2013).
8 Zimny, G.H. and Senturia, A.G. (1973). Medical specialty counseling: a survey. *Journal of Medical Education* 48: 336–342.
9 Nguyen-Van-Tam, J.S., Logan, R.F.A., Logan, S.A.E. et al. (2001). What happens to medical students who complete an honours year in public health and epidemiology? *Medical Education* 35: 134–136.
10 McManus, I.C., Richards, P., and Winder, B.C. (1999). Intercalated degrees, learning styles and career preferences: prospective longitudinal study of UK medical students. *British Medical Journal* 319: 542–546.
11 Lloyd, T., Phillips, B., and Aber, R.C. (2004). Factors that influence doctors' participation in clinical research. *Medical Education* 38: 848–851.
12 Goldacre, M.J., Davidson, J.W., and Lambert, T.W. (2007). Career preferences of graduate and non-graduate entrants to medical schools in the United Kingdom. *Medical Education* 41: 349–361.
13 General Medical Council (2013). Approved Curricula and Assessment Systems. http://www.gmc-uk.org/education/A-Z_by_specialty.asp (accessed 3 February 2013).
14 American Board of Medical Specialties (2017). Specialty and Subspecialty Certificates. http://www.abms.org/member-boards/specialty-subspecialty-certificates (accessed 27 March 2017).
15 Medical Board of Australia (2013). List of specialties, fields of specialty practice and related specialist titles. www.medicalboard.gov.au/Registration/Types/Specialist-Registration/Medical-Specialties-and-Specialty-Fields.aspx (accessed 27 March 2017).
16 Allen, I. (1988). *Doctors and Their Careers*. London: Policy Studies Institute.
17 Dillner, L. (1993). Senior house officers: the lost tribes. *British Medical Journal* 307: 1549–1551.
18 Lambert, T.W., Goldacre, M.J., and Evans, J. (2000). Views of junior doctors about their work: survey of qualifiers of 1993 and 1996 from United Kingdom medical schools. *Medical Education* 34: 348–354.
19 Lambert, T.W. and Goldacre, M.J. (2007). Views of doctors in training on the importance and availability of career advice in UK medicine. *Medical Education* 41: 460–466.
20 Jackson, C., Ball, J., Hirsh, W. et al. (2003). *Informing Choices: The Need for Career Advice in Medical Training*. Cambridge, UK: National Institute for Careers Education and Counselling.
21 Department of Health (2005). *Career Management: An Approach for Medical Schools, Deaneries, Royal Colleges and Trusts*. London: MMC Working Group for Career Management.
22 Hirsh, W., Jackson, C., and Kidd, J. (2001). *Straight Talking: Effective Career Discussions at Work*. Cambridge: National Institute for Careers Education and Counselling.
23 Petrides, K.V. and McManus, I.C. (2004). Mapping medical careers: questionnaire assessment of career preferences in medical school applicants and final year students. *BMC Medical Education* 4: 18.
24 Lloyd, B.W. and Becker, D. (2009). Paediatric specialist registrars' views of educational supervision and how it can be improved: a questionnaire study. *Journal of the Royal Society of Medicine* 100: 375–378.
25 Richard, G.V. (2005). International best practices in career development: review of the literature. *International Journal for Educational and Vocational Guidance* 5: 189–201.
26 Ministry of Health New Zealand (2017). Career Planning. http://www.health.govt.nz/our-work/health-workforce/career-planning (accessed 27 March 2017).
27 Borges, N.J. and Savickas, M.L. (2002). Personality and medical specialty choice: a literature review and integration. *Journal of Career Assessment* 10 (3): 362–380.
28 Hutt, R. (1976). Doctors' career choice: previous research and its relevance for policy making. *Medical Education* 10: 463–473.
29 Smith, A.W., Glenn, R.C., Williams, V. et al. (2007). What do future (female) pediatricians value? *Journal of Pediatrics* 151: 443–444.
30 Taber, B.J., Hartung, P.J., and Borges, N.J. (2011). Personality and values as predictors of medical specialty choice. *Journal of Vocational Behavior* 78: 202–209.
31 Newton, D.A., Grayson, M.S., and Thompson, L.F. (2005). The variable influence of lifestyle and income on medical students' career specialty choices: data from two U.S. medical schools, 1998–2004. *Academic Medicine* 80: 809–814.
32 Schwartz, R.W., Jarecky, R.K., Strodel, W.E. et al. (1989). Controllable lifestyle: a new factor in career choice by medical students. *Academic Medicine* 64: 606–609.
33 Ellsbury, K.E., Carline, J.K., Irby, D.M. et al. (1998). Influence of third-year clerkships on medical student specialty preferences. *Advances in Health Sciences Education: Theory and Practice* 3: 177–186.
34 Paiva, R.E., Vu, N.V., and Verhulst, S.J. (1982). The effect of clinical experiences in medical school on specialty choice decisions. *Journal of Medical Education* 57: 666–674.
35 Campus-Outcalt, D., Senf, J., Watkins, A.J. et al. (1995). The effects of medical schools curricula, faculty role models and biomedical research support on choice of generalist physician careers: a review and quality assessment of the literature. *Academic Medicine* 70: 611–619.
36 Chang, J.C., Odrobina, M.R., and McIntyre-Seltman, K. (2010). Residents as role models: the effect of the obstetrics and gynecology clerkship on medical students' career interest. *Journal of Graduate Medical Education* 2 (3): 341–345.
37 Passi, V., Johnson, S., Peile, E. et al. (2013). Doctor role modelling in medical education: BEME Guide No. 27. *Medical Teacher* 35 (9): e1422–e1436.
38 Dorsey, E.R., Jarjoura, D., and Rutecki, G.W. (2003). Influence of controllable lifestyle on recent trends in specialty choice by US medical students. *Journal of the American Medical Association* 290: 1173–1178.
39 Royal College of Physicians (2009). *Women and Medicine: The Future*. London: Royal College of Physicians.

40 Gray, S.F. (2004). Women in medicine: doctors of both sexes are seeking balance between life and work. *British Medical Journal* 329: 742–743.

41 Lambert, E.M. and Holmboe, E.S. (2005). The relationship between specialty choice and gender of U.S. medical students, 1990–2003. *Academic Medicine* 80: 797–802.

42 Phillips, J.P., Weismantel, D.P., Gold, K.J. et al. (2010). Medical student debt and primary care specialty intentions. *Family Medicine* 42: 616–622.

43 Rosenblatt, R.A. and Andrilla, C.H. (2005). The impact of US medical students' debt on their choice of primary care careers: an analysis of data from the 2002 medical school graduation questionnaire. *Academic Medicine* 80: 815–819.

44 Jeffe, D.B., Whelan, A.J., and Andriole, D.A. (2010). Primary care specialty choices of United States medical graduates 1997–2006. *Academic Medicine* 85 (6): 947–958.

45 Fincher, R.M., Lewis, L.A., and Rogers, L.Q. (1992). Classification model that predicts medical students choices of primary care or non-primary care specialties. *Academic Medicine* 67: 324–327.

46 Bland, C.J., Meurer, L.N., and Maldonado, G. (1995). Determinants of primary care specialty choice: a non-statistical meta-analysis of the literature. *Academic Medicine* 70: 620–641.

47 Senf, J.H., Campos-Outcalt, D., Watkins, A.J. et al. (1997). A systematic analysis of how medical school characteristics relate to graduates' choices of primary care specialties. *Academic Medicine* 72: 524–533.

48 Reed, V.A., Jernstedt, G.C., and Reber, E.S. (2001). Understanding and improving medical student specialty choice: a synthesis of the literature using decision theory as a referent. *Teaching and Learning in Medicine* 13: 117–129.

49 Barzansky, B. (2000). Commentary: research on specialty choice: the challenge is in the details. *Education for Health* 13: 197–2000.

50 Leduc, N., Vanasse, A., Scott, I. et al. (2011). The Career Decision-Making Process of Medical Students and Residents and the Choice of Specialty and Practice Location: How Does Postgraduate Medical Education Fit In? Members of the Future of Medical Education in Canada Postgraduate Project consortium.

51 Association of American Medical Colleges (2016). The Complexities of Physician Supply and Demand: Projections from 2014 to 2025 Final Report (2016 Update). https://www.aamc.org/download/458082/data/2016_complexities_of_supply_and_demand_projections.pdf (accessed 3 May 2017).

52 Myers, I.B., McCaulley, M.H., Quenk, N.L. et al. (1998). *MBTI Manual: A Guide to the Development and Use of the Myers-Briggs Type Indicator*, 3e. Palo Alto, CA: Consulting Psychologists Press.

53 Zimny, G.H. (1977). *Manual for the Medical Specialty Preference Inventory*. St Louis, MO: St Louis University School of Medicine.

54 Zimny, G.H. (1980). Predictive validity of the Medical Specialty Preference Inventory. *Medical Education* 14: 414–418.

55 Gale, R. and Grant, J. (2002). Sci45: the development of a specialty choice inventory. *Medical Education* 36: 659–666.

56 McCaulley, M.H. (1977). *The Myers Longitudinal Medical Study (Monograph II)*. Gainesville, GA: Center for Applications of Psychological Type.

57 Taylor, A.D., Clark, C., and Sinclair, A.E. (1990). Personality types of family practice residents in the 1980s. *Academic Medicine* 65: 216–218.

58 Wallick, M., Cambre, K., and Randall, H. (1999). Personality type and medical specialty choice. *Journal of the Louisiana State Medical Society* 151: 463–469.

59 Stilwell, N.A., Wallick, M.M., Thal, S.E. et al. (2000). Myers-Briggs type and medical specialty choice: a new look at an old question. *Teaching and Learning in Medicine* 12: 14–20.

60 Freeman, B. (2004). *The Ultimate Guide to Choosing a Medical Specialty*. New York: Lange Medical Books/McGraw Hill.

61 Warschkow, R., Steffen, T., Spillmann, M. et al. (2010). A comparative cross-sectional study of personality traits in internists and surgeons. *Surgery* 148: 901–907.

62 Savickas, M.L., Brizzi, J.S., Brisbi, L.A. et al. (1988). Predictive validity of two medical specialty preference inventories. *Measurement and Evaluation in Counseling and Development* 21: 106–112.

63 Glavin, K.W., Richard, G.V., and Porfeli, E.J. (2009). Predictive validity of the Medical Specialty Choice Inventory. *Journal of Vocational Behavior* 74: 128–133.

64 Borges, M.J., Gibson, D.D., and Karnani, R.M. (2005). Job satisfaction of physicians with congruent versus incongruent specialty choice. *Evaluation and the Health Professions* 28: 400–413.

65 Sodano, S.M. and Richard, G.V. (2009). Construct validity of the Medical Specialty Preference Inventory: a critical analysis. *Journal of Vocational Behavior* 74: 30–37.

66 Porfeli, E.J., Richard, G.V., and Savickas, M.L. (2010). Development of specialization scales for the MSPI: a comparison of empirical and inductive strategies. *Journal of Vocational Behavior* 77: 227–237.

67 Richard, G.V. (2011). *Medical Specialty Preference Inventory, Revised Edition (MSPI-R), Technical Manual*. Washington, DC: Association of American Medical Colleges.

68 O'Donnell, M.E., Noad, R., Boohan, M. et al. (2010). The effect of Modernising Medical Careers on foundation doctor career orientation in the Northern Ireland Foundation School. *Ulster Medical Journal* 79: 62–69.

69 Straus, S.E., Straus, C., Tzanetos, K., and International Campaign to Revitalise Academic Medicine (2006). Career choice in academic medicine: systematic review. *Journal of General Internal Medicine* 21: 1222–1229.

70 Kelly, W.N. and Stross, J.K. (1992). Faculty tracks and academic success. *Annals of Internal Medicine* 116: 654–659.

71 Sanders, A.B., Fulginiti, J.V., Witzke, D.B. et al. (1994). Characteristics influencing career decisions of academic and non-academic emergency physicians. *Annals of Emergency Medicine* 23: 81–87.

72 Bilbey, J.H., Fache, J.S., and Burhenne, H.J. (1992). Are there predictors for future academic radiologists: a Canadian survey. *Canadian Association of Radiologists Journal* 43: 369–373.

73 Lowenstein, S.R., Fernandez, G., and Crane, L.A. (2007). Medical school faculty discontent: prevalence and predictors of intent to leave academic careers. *BMC Medical Education* 7: 37.

74 Pololi, L.H., Krupat, E., Civian, J.T. et al. (2012). Why are a quarter of faculty considering leaving academic medicine? A study of their perceptions of institutional culture and intentions to leave at 26 representative U.S. medical schools. *Academic Medicine* 87: 1–11.

75 Women in Academic Medicine (2007). *Developing Equality in Governance and Management for Career Progression*. London: Women in Academic Medicine Project.

76 Borges, N.J., Narvarro, A.M., Grover, A. et al. (2010). Academic medicine careers: a literature review. *Academic Medicine* 85: 680–686.

77 Borges, N.J., Navarro, A.M., and Grover, A. (2012). Women physicians: choosing a career in academic medicine. *Academic Medicine* 87: 105–114.

78 Rogers, M.E., Creed, P.A., and Searle, J. (2009). The development and initial validation of social cognitive career theory instruments to measure choice of medical specialty and practice location. *Journal of Career Assessment* 17: 324–337.

79 Lent, R.W., Brown, S.D., and Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice and performance. *Journal of Vocational Behavior* 45: 79–122.

80 Walsh, W.B. (2001). The changing nature of the science of vocational psychology. *Journal of Vocational Behavior* 59: 262–274.

81 Hartung, P.J. and Borges, N.J. (2005). Toward integrated career assessment: using story to appraise career dispositions and adaptability. *Journal of Career Assessment* 13: 439–451.

82 Elton, C. and Newport, M. (2008). The reflective elective: using the elective to develop medical students' skills in career planning. *British Journal of Hospital Medicine* 69: 409–411.

83 Tooke, J. (2008). *Aspiring to Excellence: Final Report of the Independent Inquiry into Modernising Medical Careers*. London: MMC Inquiry.

84 Hawkins, P. (2005). *The Art of Building Windmills*. Liverpool: GIEU, The University of Liverpool.
85 Elton, C. and Reid, J. (2010). *The ROADS to Success: A Practical Approach for Career Planning for Medical Students, Foundation Trainees and Their Supervisors*, 3e. London: Kent, Surrey and Sussex Deanery.
86 Navarro, A.M., Taylor, A.D., and Pokorny, A.P. (2011). Three innovative curricula for addressing medical students' career development. *Academic Medicine* 86: 1–5.
87 Mitchell, K.E., Levin, A.S., and Krumboltz, J.D. (1999). Planned happenstance: constructing unexpected career opportunities. *Journal of Counseling and Development* 77: 115–125.
88 Bright, J.E.H. and Pryor, R.G.L. (2008). Shiftwork: a chaos theory of careers agenda for change in career counselling. *Australian Journal of Career Development* 17: 63–72.
89 Nathan, R. and Hill, L. (2006). *Career Counselling*, 2e. London: Sage.
90 Ali, L. and Graham, B. (1996). *The Counselling Approach to Career Guidance*. London: Routledge.
91 Woolf, K. and McManus, I.C. (2010). Predicting and guiding career success in medicine. In: *Learning Medicine: A Text Linking Theory and Practice for Graduate Students and Educators Wanting to go Beyond the Basics* (ed. T. Dornan, K. Mann, A. Scherpbier and J. Spencer), 297–316. London: Elsevier.
92 Kidd, J.M. (2006). *Understanding Career Counselling: Theory Research and Practice*. London: Sage.
93 Paice, E. (1997). Why do young doctors leave the profession? *Journal of the Royal Society of Medicine* 90: 417–418.
94 Moss, P.J., Lambert, T.W., Goldacre, M.J. et al. (2004). Reasons for considering leaving UK medicine: questionnaire study of junior doctors' comments. *British Medical Journal* 329: 1263.
95 Rittenhouse, D.R., Mertz, E., Keane, D. et al. (2004). No exit: an evaluation of physician attrition. *Health Services Research* 39: 1571–1588.
96 British Medical Journal (2004). *Why Do Doctors Leave the Profession?* London: British Medical Association, Health Policy and Economic Research Unit.

Further Reading

Hirsch, G. (1999). *Strategic Career Management for the 21st Century Physician*. Chicago, IL: American Medical Association.
McMahon, M. ed. (2016). *Career Counselling: Constructivist Approaches*. London: Taylor & Francis.
Nathan, R. and Hall, L. (2006). *Career Counselling*, 2e. London: Sage.

33 Supporting Learner Well-being

Jo Bishop[1], Graeme Horton[2], Wendy Hu[3], and Claire Vogan[4]
[1]Faculty of Health Sciences and Medicine, Bond University, Gold Coast, Queensland, Australia
[2]School of Medicine and Public Health, University of Newcastle, Newcastle, New South Wales, Australia
[3]School of Medicine, Western Sydney University, Penrith, New South Wales, Australia
[4]Swansea University Medical School, Swansea, Wales, UK

KEY MESSAGES

- To care for patients, health professionals must care for themselves, and they need to learn about and develop this capability. An individual's well-being impacts on their capacity to learn and perform competently. Learner well-being and supporting their acquisition of skills in self-care for lifelong practice is therefore core business for educators.
- Designing and delivering a robust learner support system requires consideration at the level of individuals (learners, teachers, and staff), organisation (procedures, programmes, curricula), and system (regulations and policies). Programmes and interventions can be categorised as general support to all learners, preventive support in anticipation of challenges, and additional support for learners in need.
- A starting point for considering the extent to which educators and schools are responsible for learner health, beyond where it impacts on learning, is the principle of 'first, do no harm'. Educators should consider their roles as individual teachers, programme designers, and in creating supportive and safe learning environments that will help learners negotiate future challenges in the clinical workplace.
- Educators should consider professional boundaries in their actions to promote well-being and welfare. While many educators may draw upon their experiences as clinicians, learners are not their patients.

> 'The wellbeing of medical students and doctors alike is becoming an increasingly recognised issue. Wellbeing is about managing your mental, physical, emotional and financial health to ensure you have a balanced lifestyle, enjoy what you do, and set yourself up for a long and rewarding career'
> Australian Medical Students' Association (AMSA) [1].

Background: Why the Focus on Learner Well-being?

Well-being is a complex construct, integrating many aspects of life that impact on a person's perception of how their lives are progressing [2]. Globally, the importance of learner well-being, and its impact on learning, performance [3], and, ultimately, health care is now being recognised. Promoting well-being is now acknowledged as an issue that requires dedicated resources and effective strategy.

To care for others, a health practitioner must care for themselves. Their ability to do this has been included in medical graduate competencies and professional body initiatives worldwide. Examples are the following: Australian Medical Council (AMC) [4], General Medical Council (GMC) [5], Royal College of Physicians and Surgeons of Canada: CanMEDS framework [6], and the Accreditation Council for Graduate Medical Education [7]. It is therefore expected that medical schools and colleges prepare and train learners for this lifelong skill and professional capability.

Attention to learner well-being is brought into sharp focus at times of crisis or tragedy, many of which have been described recently [8]. Such events have triggered enquiries revealing a far greater extent of the problem than was previously documented [9]. There is a high prevalence revealed from the enquiries of stress and burnout amongst medical students and junior doctors, beyond that of age-matched population controls [10]. Burnout in the context of health care professionals [11] is a multifaceted construct which includes emotional exhaustion, depersonalisation, behaving insensitively towards patients, is associated with a low sense of personal accomplishment [12], and, in some cases, results in those affected leaving the profession.

This chapter explains why the well-being of learners is a priority for all education providers; many examples refer to medical schools and their students but the information is relevant to the well-being of all allied health learners. The chapter outlines a range of approaches and strategies to be considered when developing a robust system of learner support that will promote a healthy and sustainable workforce to care for future patients.

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

Challenges to Learner Well-being

Descriptive studies have revealed a range of challenges to learner well-being which may be considered in the following domains.

Personal Factors

Learners begin their educational journey following selection (and/or interview) with the academic abilities and recognised attributes to likely succeed in their chosen programme, but learners may develop illness or experience difficult personal circumstances during their studies which impact on their ability to complete course requirements. Personality traits which may predispose to depression and burnout, such as neuroticism and conscientiousness, are prevalent in medical students and doctors [13]. Do these personality traits lead to increased incidence of 'unwell-being'? The two studies below confirm that higher rates of depression have been reported within the medical student than that found in the general population. A systematic review and meta-analysis of the prevalence of depression amongst medical students in 43 countries was reported by Rotenstein et al. who found that the overall pooled crude prevalence of depression or depressive symptoms was 27% [14]. The prevalence of suicidal ideation was obtained in this review from 24 cross-sectional studies in 15 countries and found to be 11.1%. A study by Mata and colleagues found that interns who had symptoms of depression were more likely to also report cynicism, exhaustion, and stress, whereas those without symptoms of depression more commonly reported positive life-changing experiences and positive interactions with patients and colleagues [15].

This has raised questions about whether these associated features are causal or an effect of depression in the intern and longitudinal studies are needed to determine this.

Environmental Factors

Dyrbye and Shanafelt reviewed studies from 1990 to 2015 of burnout in medical students and doctors and concluded that whilst personality factors play a part, the work and study environment was more significant [12]. Considerations around transitions, the length and challenges of programmes and postgraduate working hours, and workplace harassment, along with the stigma associated with seeking support, will be explored as factors that require attention.

Learners such as medical students must traverse key transitions, from classroom to the clinical environment and from final year to graduate practice, with multiple rotations in-between, each with new demands to adjust to in unfamiliar surroundings, often in different locations and communities. Any support structures which have been in place for the trainee may be disrupted or discontinued. Whilst transitions can be viewed as learning opportunities [16], stress levels have been believed to increase at these times [17]. Levels of burnout were high for all groups of doctors, but emotional exhaustion was reported highest in younger doctors and declined across age groups [10].

During medical degrees, which are often of a long duration, medical students may be required to forego relationships and personal needs to meet complex programme requirements [18]. Also, they can incur relatively higher levels of debt, which can cause considerable stress – particularly for those whose continuation in the programme is uncertain.

Hospitals and other health care settings are high demand workplaces which may be under-resourced and in which doctors and other health professionals may need to work long hours in high-pressure situations.

Harassment in medical schools and in hospitals is another issue which adversely affects the well-being of doctors. The issue was first documented by paediatrician Henry K. Silver in 1990 who reported in *JAMA* that 46.4% of students at one medical school had been 'abused' and that by the time they were senior medical students this proportion had risen to 80.6% [19]. Being exposed to such negative role modelling, as well as the stressors of witnessing pain, suffering, and death, are continual challenges to learner well-being (see Box 33.1).

The stigma of mental health is prevalent in health care students and practitioners. This may result in learners being reluctant to identify as suffering from a mental health condition and from seeking the required support and management [20]. Students and doctors may be wary of accessing support because of fear of activating fitness to practise and registration authority reporting procedures, which they are worried may jeopardise their career (see Box 33.2).

Promoting Well-being: The Role of Education Providers

Medical schools have well established responsibilities to not only deliver systematic interventions to support learners in need but to develop proactive approaches to promoting well-being of learners. These are driven by accreditation

BOX 33.1 How the culture of medicine hinders the well-being movement

Slavin has identified four aspects of the culture of medicine which have worked against more widespread acceptance of the need to address learner mental health concerns.

1. The belief that medical education needs to be rigorous so that only the best survive, which is thought to be reflective of the work environment. This is linked to the concept that greater challenges lead to better educational outcomes and doctors who will weather the stresses of medical practice.
2. There is a lack of accountability of medical schools for the well-being of students; universities may prioritise research and clinical advances.
3. Units in schools which support students may be siloed from curriculum development units, the latter of which are likely to be better resourced.
4. Mental health problems may be considered less significant than physical health problems [17].

BOX 33.2 Busting the myth of mandatory reporting

Tragic cases of self-harm and burnout are reminders of the importance of secondary prevention strategies; if a student or doctor is struggling, help should be sought early. However, fears of being reported to the authorities, or of compromising future careers, are likely key barriers to seeking psychological or psychiatric advice.

Some jurisdictions require health practitioners to report any impaired practitioners whom they may be treating, including students or other practitioners, whose practice may put the public at risk. This unfortunately is the case even if the student or practitioner is participating in an approved programme of treatment or is being managed effectively with no concerns around future risk. New Zealand has also required that treating health practitioners report 'impaired' peers since 2003 [21]. Several American states have mandatory reporting obligations that extend to treating practitioners [22]. While safeguards are necessary to protect the public, in the absence of experience about the threshold of reporting, the likelihood of being reported has been magnified in the minds of many students and trainees.

It is hoped that a consistent approach to mandatory reporting provisions will provide confidence to health practitioners that they can seek treatment for their own health conditions without putting their careers in jeopardy [23]. A review found that reports from the treating practitioner to the medical board were rare – the received reports were about substance misuse or mental illness and were made by a doctor who was not the patient's regular care provider [24].

While it is apparent that the real risk of being reported is very low, such is the fear of being reported, especially amongst students and junior doctors, that it is invoked as yet another reason not to seek care, or to act responsibly about one's own well-being. It is therefore vitally important that students, trainees, practitioners, and educators accurately understand the regulations and legal requirements that pertain to their jurisdiction. They should be discussed openly to ensure transparency around self-care, the care of others, and, ultimately, the safety of patients. The appropriate help-seeking pathways that are accessible should be known to all and should be *visible*.

requirements, competency frameworks, including those relating to professionalism, and by stakeholder groups.

Student associations around the world have developed awareness campaigns and programmes, often in collaboration with professional associations and sponsors (see also Box 33.5).

Consistent with the activities of medical students, are broader initiatives in higher education with learners regarded as consumers or co-producers of education, and measures of student engagement being performance indicators for universities around the world. Learner involvement in the design and review of their courses, representation on committees at school and university levels, retention, and student performance statistics and national student satisfaction surveys (used as the basis for ranking universities) are now commonplace and signal the attention paid to the learner experience.

Such measures are reflected in national quality assurance and accreditation processes which often take a student-centric approach to reviews and have student representation on their panels (e.g. in the UK the Quality Assurance Agency for Higher Education, QAA) [25]. The medical school accreditation bodies, for example the World Federation of Medical Education (WFME) [26], the AMC [4], and the GMC [27] require schools to have systems in place for the support of learners, to detect concerns at an early stage, and to provide accessible support services, separate to assessment. Learners should feel safe to approach such sources of support, without fear of their help seeking affecting their progression.

Self-care and its importance to life-long learning, and how development as a competent health professional requires one to recognise one's own limits are now essential components of medical curricula. Kreitzer and Klatt describe v curricula with self-care as a core competency [18], which were highlighted above. If self-care is truly to be a competency within a curriculum it would therefore have to be blueprinted – and many programmes do not yet do this.

Medical schools are also increasingly seen to be responsible for ensuring that graduates are prepared for practice, and for the demands of a career in medicine. In recent years, there has been a marked increase in the number of programmes for learners to help them with stress management techniques (e.g. Saunders et al. [28], Hassed et al. [29], Wasson et al. [30]), and to manage key stress points. Examples include early clinical exposure, simulation scenarios, professionalism training, reflections following placements, shadowing periods, and preceptorship. Interventions such as pre-internship (PRINT) programmes [31] describe how to ease transition to clinical practice. The focus, however, is on revising clinical skills, rather than developing ways of dealing with the psychological demands of taking on new roles [32].

There is evidence that curriculum design and the learning environment have significant effects on learner well-being. Curriculum developers and medical schools can thus institute practical and effective measures to reduce unnecessary stress; these are described in more detail below.

A Well-being Support Framework

Educators can be viewed as taking a three-pronged approach to the type of support they offer their learner (see Figure 33.1), namely 'general support', which helps learners get through the course, 'preventive support', which helps learners prepare for their future careers, and 'extra support', which helps learners deal with individual, and sometimes difficult, circumstances.

Within these three areas, numerous models are available to help those involved in the design and implementation of learner support processes which can be aimed at faculty

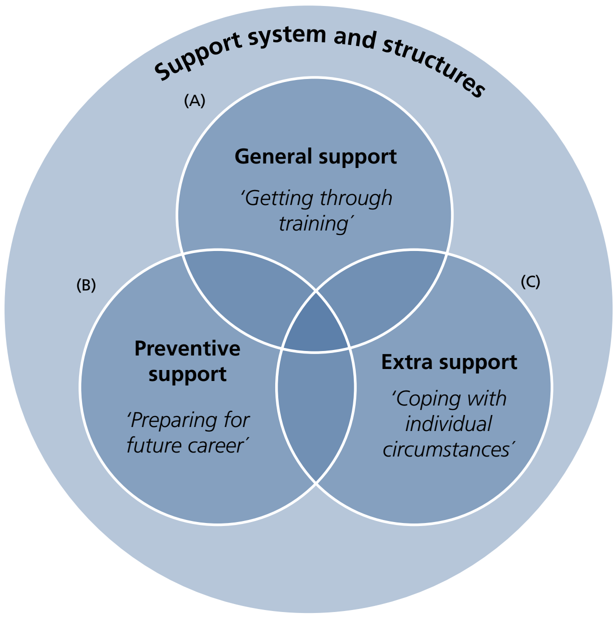


Figure 33.1 A diagram conceptualising the learner support system and approaches used by educators.

members and staff working at the 'coal-face' and/or at the organisational level. Underpinning the support, educators (e.g. medical schools) offer a generic support structure which usually involves multiple people and processes, linked into the medical school, wider university environment, and clinical arena. The most successful support systems tend to have multiple access points, target support delivered to times of vulnerability (e.g. exam periods, times of transition), and ensure transparency in their procedures and processes (see Box 33.3).

General Support ('Getting through training')

General or generic support (see Figure 33.1A) is that which is offered to all learners throughout their training. This can cover a multitude of areas such as handbooks, the pastoral care system [33, 36], formative exams [49], feedback [35], and briefing and debriefing on placements [50]. Educators need to ensure that the general support offered to learners is clearly sign-posted, easily accessible, transparent, and timely. They should also be mindful that learners can be reticent to access support [51] and therefore educators need to work hard at normalising the need to access support.

Preventive Support ('Preparing for future career')

Preventive support activities (see Figure 33.1B) are offered to help learners prepare for the rigours of their profession, e.g. medicine, and equip them with personal and professional skill sets for successful careers. This could include training in techniques to help with resilience [18], personal and professional development sessions [52], support and training strategies for learner support providers, and help with career planning. Whilst they generally tend to be available to all, except for activities directly linked to professionalism training within a curriculum, many educators seem to adopt an opt-in approach to these activities, thereby fostering engagement only amongst those that choose to attend.

Extra Support ('Coping with individual circumstances')

Some learners have circumstances which mean they have specific support needs or an increased likelihood of encountering difficulties at some point during their medical training (see Figure 33.1C). Whilst all courses should aim to ensure inclusivity in their curriculum design, delivery, and assessment processes, such learners can often benefit from needs-tailored (personalised) extra support programmes (e.g. peer social events for international students and/or specific, individualised support measures) being put in place. These needs may be identified early when educators have personal tutors [52, 53] or mentoring schemes.

Training adjustments are often required for learners who have disabilities [12, 44], temporary health conditions, dependents, and those dealing with significant personal issues (e.g. recovery following bereavements, serious financial difficulties) as well as academic issues that ultimately could result in them not progressing through the course. Depending on the circumstances, adjustments may be either temporary or permanent (i.e. extend into their working careers), or may take the form of additional tuition either with the aim of raising their general academic standard or learning coping strategies for safe clinical practice.

Whilst academic issues are frequently dealt with by education providers themselves, dealing with current health and well-being issues, and knowing the adjustments that may need to be put in place to enable a student to succeed on the course, frequently require onwards referral to and advice from specialist support units either internal or external to the university. To give the student the best chance of successful progression through their course and ease the 'mental burden' of a struggler, medical schools for example should aim to put systems and training in place to allow their faculty to identify those in difficulty early [40, 42, 43, 45–47].

Faculty and Peer Training in Student Support Roles

Academic and professional staff, as well as fellow learners, are often called upon to assist learners with concerns, and to respond to distressed learners [54]. Due to fear of stigma, medical and health professions students may be more likely to seek help from a professional staff member or a peer [55]. Staff and peers in the health professions may feel responsible for providing support, but are usually not trained to do so. Some may find the experience disturbing, with ongoing personal impact [56]. Over time, these experiences may lead to emotional exhaustion and burnout in support providers themselves [57]. Box 33.4 suggests practical strategies which health professions programmes can consider when supporting and training staff who may be called upon to respond to distressed learners (modified from Hochschild [56]). Boundaries need to be carefully considered for learners who take on peer support roles; in contrast with staff, there are no natural boundaries, with greater potential for negative impacts.

BOX 33.3 Examples of frameworks and approaches

These frameworks and approaches can be used to help educators in the development or implementation of learner support systems. They are linked to the three areas of general support (A), preventive support (B), and extra support (C) seen in Figure 33.1.

	Type of approach	What is it and how might it be used? (link to area or areas of support which the approach most relates to)
Cross and Lester [33]	Pastoral care: 10-step plan	A framework that can be followed to help review and improve a school's pastoral care programme (A/C).
Vogan et al. [34]	Twelve tips to supporting medical students	General guidance on setting up and maintaining a robust support system for medical students (A).
Wiggins [35]	Seven keys to effective feedback	A simple framework that can be followed to ensure any feedback being given is effective (A).
Ramani et al. [36]	Twelve tips for developing effective mentors	General guidance on establishing a good mentoring relationship with a trainee (A/B).
Sanders et al. [37]	Developmental student support	An approach that focuses on developing the medical student as a whole rather than solely providing academic and clinical support. It looks at the delivery of effective individualised support in a number of developmental areas including mentoring, resiliency, and careers as well as looking at how best to support those with disabilities (A/B/C).
Dunn et al. [38]	'Coping reservoir' model of well-being for medical students	A model that gives insight into the factors, both positive and negative, that influence medical student well-being and the effect these can have on promotion of resilience and prevention of burnout (B).
Gordon [39]	Personal and professional development framework	A model that gives insight into the factors that influence behaviour and how changes in behaviour may be learnt. It can be used to look at methods of fostering personal and professional development within a medical school curriculum (B).
Kreitzer and [18]	Educational innovations to foster resilience	A study that reviews the need to fostering resilience and ways of fostering resilience within the health care professions. It highlights a number of programmes that use mindfulness, mind–body skills, and resiliency training and, hence, would be of interest to those looking to incorporate more resilience-building techniques into their curriculum (B).
Bernstein [40]	Diagnosing the learner in difficulty	A clinical approach to making a 'differential diagnosis' of a struggling student and exploring the subsequent management options (C).
GMC [41]	Gateways to the professions	Guidance from the GMC in the UK on how best to support medical students with disabilities through their medical training and beyond (C).
Hays et al. [42]	Profiles of medical students needing support	A taxonomy that outlines seven distinct profiles of struggling medical students which can be used to help identify those in difficulty early and gives suggestions for remediation (C).
Hicks et al. [43]	Dealing with medical student difficulties in the clinical setting	A classification of types of medical student difficulties encountered in the clinical setting with suggestions of strategies to prevent, assess, and work effectively with these students (C).
Medical Schools' Council and General Medical Council [44]	Supporting medical students with mental health conditions	Guidance from the GMC in the UK on how best to support medical students with mental health conditions through their medical training and beyond (C).
Norrish et al. [45]	Interim identification of academically 'at risk' students	A study that identifies a number of criteria that can predict future academic performance on the basis of a student's previous and current performance. This information could be used for the early identification of students academically 'at risk' of not completing their studies, enabling extra support measures to be put in place (C).
Winston et al. [46]	Early identification of academic strugglers and subsequent successful remediation	A study that looks at ways of predicting medical students at risk of academic failure and gives insight into the factors that lead to successful remediation programmes (C).
Yates [47]	'Toolkit' to identify medical students at risk of failure to thrive	A study that identifies a number of criteria that could be monitored to help identify, at an early stage, those medical students at risk of struggling (C).
Plymouth University [48]	7 Steps to: Adopting Culturally Inclusive Teaching Practices	This succinct guide gives practical tips on how to improve inclusivity, engagement, and cultural sensitivity within everyday teaching practices (C).

BOX 33.4 Support and training strategies for those who provide learner support

Recruitment and selection
- Selection of staff or peers with a personal orientation towards emotional work, positive attitudes to students and their concerns, who are role models for self-care and well-being

Induction and role design
- Orientation to role and the fit with the medical school mission
- Role clarification of responsibilities, and personal and professional boundaries of what is acceptable and not acceptable
- Knowledge of and readily accessible current information about:
  - policies and procedures about when to refer, assessment, progression, privacy and confidentiality, reporting, and documentation
  - local support services and referral pathways
  - training in mental health first aid
  - recognition and initial response pathways for common student presentations
- Rotating duties with respite periods from being accessible to learners, defined times of day for receiving learner requests

Ongoing professional development
- Opportunities to exchange and reflect on experiences with peers to build a community of practice
- Skills training in student mental health and well-being, self-care, and boundary setting
- Formal recognition of support work in employment agreements and promotions

Emergency situations
- Readily accessible critical incident flowchart for use in crisis situations
- Opportunities for critical incident debriefings after distressing encounters
- Monitoring for early identification of ongoing trauma and need for assistance

A 'needy' learner may have the tendency to latch-on to their peer mentor; the peer mentor is also dealing with the stresses of being a learner and so this may result in them having difficulties too.

Examples of Support Programmes and Interventions

Here we highlight some more specific examples of ways in which universities and medical schools around the world have developed and implemented interventions to improve learner support. As with the frameworks section, we have grouped the examples into general, preventive, and extra support interventions. However, as we saw previously, it is important to realise these categories are not mutually exclusive. Learners have diverse backgrounds and cultures [48]. Along with general measures, which can be implemented by programmes to promote well-being for all learners, there are benefits of tailoring learner support for groups of learners. Medical schools can also work with student and doctors' organisations who may have taken the lead in such initiatives. We finish with a short section on programmes and interventions that have been tailored to address the diversity and cultural support needs for groups of learners. A participatory approach, engaging learners in the development and delivery of these services, helps to ensure that they are appropriate and welcomed.

General Support

Curricula which include more time on clinical placements have been found to be associated with fewer learners dropping out of study [15]. The St Louis University School of Medicine, as part of a programme of changes to improve student mental health, decreased curriculum time by 10% in some courses and negotiated with lecturers to decrease the amount of detail taught in certain courses. Longitudinal electives were instituted so that students could meet learning requirements over a longer period [4, 15].

Assessments can be stressful for many learners due to expectations of high standards of performance and potential barriers to progression if performance is deemed unsatisfactory. Formative assessments, in which the content tests similar knowledge and skills to what will be in summative examinations, have been shown to reduce negative perceptions of graded course exams [58, 59]. Evans has described how implementing a programme of a range of different formative assessments throughout a period of learning in anatomy can enhance learner engagement with course content [58]. This was in keeping with findings indicating that quizzes and fun ways of learning can reduce stress and anxiety levels about assessment [54]. Rolfe and McPherson describe the importance of having formative exams that mimic summative exams [49]. A study on programmatic assessment was found to indeed influence student learning, and this influence can either support or inhibit students' learning responses [60]. Chen et al. have found that the introduction of progress testing into their curricula reduced the level of stress [61].

Learners who are part of programmes which have pass/fail grading systems rather than several graded intervals (often five such as A, B, C, D, Fail) have been found to have reported significantly better well-being: specifically, less anxiety, depression, and stress; and better group cohesion scores [62]. The University of Virginia medical school was one such school which introduced pass/fail grading for the first two years of training for the class of 2007 and found no decrease in academic performance [62].

Peer-assisted learning programmes can help learners with adjustment to university life. For example, at the University of New South Wales, the student society allocates three mentors to a group of seven first-year medical students, who run weekly sessions about textbooks, referencing, study techniques, and exam preparation, as well as well-being, financial advice, and special interest groups [63]. The importance of near-peer support in running specific sessions to ease general anxieties (e.g. getting through exams) should not be under-estimated and can be easily facilitated by faculty at transition points [64].

Preventive Support

Discussions around preventative support have included options such as learning communities, mindfulness, and peer support.

A number of medical programmes have structured their learning environments into 'learning communities', which are groups of faculty and students working together. Activities include mentoring [65], mind–body training, and personal and professional development (including annual retreats). High levels of student satisfaction have been reported as well as decreased burnout rates [30].

Medical schools have delivered mindfulness, relaxation, and meditation classes for students, some of which include biofeedback and guided imagery. Findings from evaluations have included significant reduction in stress after the intervention, increased awareness of tension, increased ability to deal with stress, and less test anxiety [30]. The programme at Georgetown University School of Medicine has trained 3000 participants, including members of faculty from European schools, and involved students meeting weekly for 11 weeks in two-hour meetings in groups of 10. Each session included an opening meditation, followed by a sharing session, and then a new mind–body practice [25]. See below for more on meditation.

Peer support structures (or 'buddy systems') can complement a faculty-based network. Students may be more comfortable discussing issues with a peer. There needs to be careful consideration of training requirements including boundary issues and referral. AMSA has encouraged peer-led mentorship programmes along with other recommendations [66] (see Box 33.5). Spielman has described such a pastoral-based student peer-support programme in a UK Veterinary school which was positively evaluated by students [67].

Mental Health First Aid (MHFA) Australia is a national not-for-profit health promotion charity focused on mental health training and research into teaching people how to offer initial support to those who are experiencing a mental health problem [68]. The training can be online, face to face, or blended. Course participants learn about the signs and symptoms of common and disabling mental health problems, how to give initial support, how to get professional help, and how to provide first aid in a crisis. Mental health problems covered are: depression, anxiety problems, psychosis, substance use, and eating disorders. Mental health crisis situations are also explored. Bond et al. found that with 434 Australian nursing and medical students both online and face-to-face courses improved intentions towards other students experiencing depression, and improved confidence in providing help to them. Stigma and desire for social distance from those with depression were reduced [69]. Davies found similar improvements for students in the UK as well as reports from students that the training was interesting and informative for both their studies and their personal life [70].

Medical schools have promoted access to mental health services to learners and reduced barriers such as stigma by providing faculty education, student handbooks, and lectures about what is available [71]. The Centre for Spirituality and Healing at the University of Minnesota runs an eight-week online course that emphasises ways to manage stress, cope with challenging situations, and cultivate resiliency. The programme reported very positive feedback from participants of interprofessional iterations of the course, with residency programmes now building this course into their core training [18].

BOX 33.5 How do learners help each other?

Many accrediting bodies require that '*The medical education provider has formal processes and structures that facilitate and support student representation in the governance of their program*' [7].

It can be assumed that all education providers have a learner society that they can liaise with to ensure representation for academic and advocacy requirements.

How can learners thrive at medical school, not simply survive?

- Start a well-being initiative
- Share your stories
- Talk to others (faculty, peers)
- Run an event
- Get involved with local events
- Get informed
- Get social: Twitter, Facebook
- Use online programmes

In September 2013, the Australian Medical Student Association (AMSA) adopted a new health and well-being policy and in-line with this their student mental health and well-being committee reviewed the current informal initiatives that are run by the medical student societies of Australia and New Zealand. It was revealed that only 13 of the 20 medical societies included a position for a Health and Wellbeing officer. Common initiatives included mentoring programmes that included academic and pastoral issues, with some societies combining faculty and student initiatives. Well-being evenings and workshops were also common, but limitations around follow-up were implied. The review revealed that although events and classes around yoga and meditation were arranged the attendance was minimal. A best practice initiative which has now gained national discussion is the provision of a list of GPs and psychologists; the benefits are twofold – further promotion of health and well-being and increased access to professional advice [67].

Extra Support

The practice of faculty members volunteering to provide one-on-one advice to learners exists in many schools. Some programmes have been developed which involve faculty advising or mentoring groups of learners and these have been reported by Sastre as having significantly higher learner satisfaction levels about how wellness was promoted

[65]. Seritan describes the creation of a new Office of Student Wellness with evening hours and strict confidentiality [72]. Such interventions have been associated with lower rates of depression and suicidal thoughts in those who have been exposed to them. It has been reported that learners are more likely to self-refer to support services after having been made aware of them [30].

As discussed above, many learners will have disabilities or temporary health conditions and it is a responsibility of education providers to provide and facilitate reasonable adjustments so that these learners may proceed to fulfilling careers wherever possible. To determine what adjustments can be made in training and in future clinical practice, it is important to access available expert advice from equity or disability specialist units at Universities. Timely consultation with registration authorities for guidance may also be required.

International learners may face difficulties resulting from their move to a different country, including language barriers, financial burdens, heavy workloads, and cultural differences [73]. Measures which have been suggested may be utilised to support international learners include financial aid, promoting peer social support, nurturing cultural integration, and specific medical language courses [74, 75]. Some medical programmes provide the opportunity for international learners to seek support from a dedicated academic mentor. University international student offices often provide advice for international students, including referral to services for academic support, counselling, and accommodation.

Indigenous health support units, recruitment of indigenous staff, and cultural safety training are some of the ways in which support and academic mentoring can be provided to indigenous learners. The 'Best Practice in Student Support for Indigenous Medical Students' is an example of a resource which can be utilised [76].

The LIME Network (Leaders in Indigenous Medical Education) provides a forum for health education leaders to 'ensure the quality and effectiveness of teaching and learning of indigenous health, as well as best practice in the recruitment and graduation of indigenous medical students' [77]. The Committee of Deans of Australian Medical Schools (CDAMS) Indigenous Health Curriculum Framework and the Critical Reflection Tool are not designed to provide specific support and guidance for indigenous learners; however, they do help promote a culture and curriculum climate of inclusivity and awareness-raising of the primacy of indigenous leadership and knowledge.

The American Medical Student Association Committee on Gender and Sexuality has been proactive in providing advice to those students wishing to establish support groups for LGBTI learners. Recommendations from their handbook include finding an administrative and faculty advisor who is responsible for advocating for LGBTI patients as well as support for LGBTI students [78].

Resilience, Meditation, and Mindfulness: What's the Evidence?

Resilience has been defined broadly as the ability to withstand adversity and to recover from challenging circumstances [8, 79]. Until 2005, there were few reports of resilience in the medical education literature [80].

All health care professionals have the potential to face stressors that could lead to depression and burnout. Developing the skills to withstand these difficult times and facilitate personal recovery is the essence of programmes which aim to build resilience. According to Howe [80], resilience is a skill which can be learned and strengthened. Rutter [81] discussed the need for learners to have experiences or learning which in some instances may be out of their comfort zone, to develop resilience. Medical schools can be proactive in ensuring that when such learning occurs that it is not harmful to students, that there is close monitoring of learner responses to such situations, and that adequate opportunities for debriefing are provided. To introduce learners to such challenges, simulated encounters can be utilised, including graded exposure to complex scenarios that do not fit the 'rules' and case-based learning around real-life problems (e.g. pre-clinical problem-based learning).

Rogers reviewed several resilience programmes which used validated measures of resilience as outcome measures, but concluded that resilience training should be sustainable by being transferable to the clinical workplace; many programmes are classroom based with no follow up of the learners in their work environment [82]. The strongest evidence for improving resilience was using resilience workshops, cognitive behavioural training, or a combination of interventions. These interventions included problem solving in small groups, which showed some evidence of improving resilience, as did those interventions which included reflection, mindfulness, relaxation training, and mentoring. Limitations identified included the comparable timing of interventions and the tools used to measure resilience. Nevertheless, resilience scales [83] could be used by learners and educators as reflective tools to better understand their coping abilities and to develop greater resilience. Educational approaches to supporting the building of resilience have been reviewed [80, 84, 85]. We have provided examples of educational approaches which could be utilised to develop resilience according to respective theoretical constructs [80, 84, 85] (Box 33.6).

Meditation

Countless meditation practices have been developed in the Buddhist tradition over the millennia to improve health and well-being. Meditation practices known as focused attention meditation or open monitoring meditation incorporate either the concentration towards an object (external, corporal, or mental) whilst ignoring all other stimuli or the ability to enlarge the attentional focus to all incoming sensations, emotions, and thoughts from moment to moment without focusing specifically on any of them. Boccia et al. conducted a meta-analysis of neuroimaging data on the effects of meditation on brain structure and function that indicated that meditation leads to activation in brain areas involved in processing self-relevant information, self-regulation, focused problem solving, and adaptive behaviour [86]. Results also show that meditation practice induces functional and structural brain modifications in 'expert

BOX 33.6 Resilience construct theories and example educational approaches to build them

Theoretical construct	Examples of educational approaches
Development of self-efficacy 'confidence'	Learners require less didactic and more self-directed learning opportunities
Co-ordination 'planning'	Learners should have realistic goals and embrace feedback to build on their next performance
Control	Learner representation on committees is an opportunity for course change and to make a difference to the system that could include worker friendly policies
Composure 'low anxiety'	Assessment that supports progressive growth in learner confidence, e.g. programmatic assessment (described above) and continuous low stakes exposure to learning challenges (placements, communication skills with patients)
Commitment 'persistence'	Longer term goals are recognised; provide continuity through promoting reflection, mentoring, and peer networks systems, from the beginning of any training programmes and throughout challenging transitions

meditators'. Expert meditators also showed higher activations in the parahippocampal cortex, which is involved in memory formation and retrieval as well as in high-level perception, especially in perceiving complex and ambiguous visual stimuli [86].

Mindfulness can be defined as a state of awareness comprised of two interrelated components: the self-regulation of attention to your immediate experience; and the maintenance of a specific orientation towards that experience [87]. Mindfulness-based programmes are increasingly being promoted [88]. A self-administered intervention has been shown to be effective by significantly lowering stress and anxiety levels in senior medical students at 8 weeks which was maintained at 16 weeks [89]. See Box 33.7 for an example of a mindfulness programme in action.

Future Directions

This chapter has provided an overview of the importance of learner well-being in the education of future health professionals; the focus has been on the literature for medical students with the principles being applicable to all allied health learners from undergraduate to postgraduate. We have explored the responsibilities of medical schools (all education providers) to develop and promote robust support programmes at the level of individuals, organisations, and systems to address the challenges to well-being that impact unnecessarily on learning and performance. We have argued that securing the well-being of the next generation of health care providers requires a comprehensive, pre-emptive, and targeted approach so that learners are

BOX 33.7 A group-based workplace wellness initiative

The oneED wellness programme was initiated at the Gold Coast Health Emergency Department (ED) in 2016. It was recognised that health care workers had high risk of burnout, compassion fatigue, and depression [10]. It was also recognised that an effective programme would shift the focus away from *helping those who are struggling* and move towards *creating a collective culture of positive growth* whereby all employees could be nurtured to thrive.

The programme is based on mindfulness techniques and employs low-intensity practices that were designed to be embedded into work activity:

- a weekly 30 minute session including a 'sit', journaling, and sharing
- a weekly 4-minute pause at clinical handover consisting of introductory videos, a thought for the day, and/or a 'sit'
- flyers posted around the department.

Mindfulness was selected for its strong neuroscientific basis not only for *well-being* but also for *attention focus training*. The basis for mindfulness practice is curbing the stress response, especially when it becomes chronic. Studies have shown that, with training, there are anatomical and physiological changes to the brain [90], and that these changes then translate to improved attention control, emotional regulation, and facets of emotional intelligence – collaboration, communication, shared goals [91].

A key concept of mindfulness is embodied in the following quote by Holocaust survivor and psychiatrist, Victor Frankl: *Between stimulus and response, there lies a space. In that space is our power to choose our response* [92].

This is particularly apt for working in emergency medicine. Managing a constant stream of stimuli and interruptions by pausing before an otherwise reflexive response is a skill that allows ED health care workers not only to keep their emotions in check, but to respond in a manner that is considered and intentioned. This should then lead to greater self-awareness, improved team dynamics, and ultimately to enhanced patient care.

equipped to cope with and thrive in their future careers and to care for patients to the best of their ability.

Nevertheless, there are questions which require further investigation and exploration, including:

- What is the impact of current accreditation requirements and national regulatory systems on well-being, help seeking, remediation, and practitioner rehabilitation?
- What interventions will lead to sustained reduction in the stigma of mental illness amongst medical and health professions?
- What strategies will promote healthy workplace cultures and positively impact on learning and professional development?
- How transferable is the well-being research to date to non-Western cultures and medical education systems?

There is a need for research on learner well-being in medicine that is more theoretically informed and empirical. Much of the research to date has comprised of cross-sectional self-report prevalence studies and their associations with measures of poor mental health, single institution programmes with limited outcome measures, and only a few systematic reviews.

Longitudinal studies and research on programmatic approaches to promoting well-being are challenging to conduct, but as they will be better at identifying causative factors and the conditions under which interventions work, they may lead to more effective and long-lasting effects. The role of medical schools should extend beyond removing inappropriate curricular stressors and learner barriers, and should move towards providing effective interventions and promoting better learning environments in health care settings that will benefit all who work there.

Acknowledgements

We thank Shahina Braganza for her contributions to the OneED. Follow her work at: http://shahinabraganza.com

References

1 Medical Student Wellbeing (2011). AMSA Factsheet. http://media.amsa.org.au/publications/student_wellbeing_fact_sheet_2011_may.pdf (accessed 2 August 2017).

2 Well-being Concepts, Centers for Disease Control and Prevention. https://www.cdc.gov/hrqol/wellbeing.htm (accessed 7 August 2017).

3 Enhancing Student Wellbeing; Resources for University Educators. http://unistudentwellbeing.edu.au (accessed 7 August 2017).

4 Australian Medical Council 2013. Standards for Assessment and Accreditation of Primary Medical Programs. www.amc.org.au/files/d0ffcecda9608cf49c66c93a79a4ad549638bea0_original.pdf (accessed 7 August 2017).

5 General Medical Council 2013. Outcomes for Graduates (originally published in *Tomorrow's Doctors*) sets out the knowledge, skills and behaviours that new UK medical graduates must be able to show. http://www.gmc-uk.org/education/undergraduate/undergrad_outcomes.asp (accessed 27 August 2017).

6 CanMEDS, Royal College of Physicians and Surgeons of Canada. http://www.royalcollege.ca/rcsite/canmeds/canmeds-framework-e (accessed 7 August 2017).

7 Accreditation Council for Graduate Medical Education. http://www.acgme.org/What-We-Do/Initiatives/Physician-Well-Being (accessed 2 August 2017).

8 Munn, F. (2017). Medical students and suicide. *Student BMJ* doi: 10.1136/sbmj.j1460.

9 Slavin, S.J. and Chibnall, J.T. (2016). Finding the why, changing the how: improving the mental health of medical students, residents and physicians. *Academic Medicine* 91: 1194–1196.

10 Beyond Blue, National Mental Health Survey of Doctors and Medical Students. http://resources.beyondblue.org.au/prism/file?token=BL/1132 (accessed 7 August 2017).

11 Willcock, S.M., Daly, M.G., Tennant, C.C., and Allard, J. (2004). Burnout and psychiatric morbidity in new medical graduates. *Medical Journal of Australia* 181: 357–360.

12 Dyrbye, L. and Shanafelt, T. (2016). A narrative review on burnout experienced by medical students and residents. *Medical Education* 50 (1): 132–149.

13 Tyssen, R., Dolatowski, F.C., Rovik, J.O. et al. (2007). Personality traits and types predict medical school stress: a six-year longitudinal and nationwide study. *Medical Education* 41 (8): 781–787.

14 Rotenstein, L.S., Ramos, M.A., Torre, M. et al. (2016). Prevalence of depression, depressive symptoms, and suicidal ideation among medical students: a systematic review and meta-analysis. *Journal of the American Medical Association* 316 (21): 2214–2236.

15 Mata, D.A., Ramos, M.A., Kim, M.M. et al. (2016). In their own words: an analysis of the experiences of medical interns participating in a prospective cohort study of depression. *Academic Medicine* 91 (9): 1244–1250.

16 Teunissen, P.W. and Westerman, M. (2011). Opportunity or threat: the ambiguity of the consequences of transitions in medical education. *Medical Education* 45 (1): 51–59.

17 Slavin, S.J. (2016). Medical student mental health. Culture, environment and the need for change. *Journal of the American Medical Association* 316 (21): 2195–2196.

18 Kreitzer, M.J. and Klatt, M. (2017). Educational innovations to foster resilience in the health professions. *Medical Teacher* 39 (2): 153–159.

19 Silver, K., Duhl, A., and Glicken, M.S.W. (1990). Medical student abuse incidence, severity, and significance. *Journal of the American Medical Association* 263 (4): 527–532.

20 Stubbs, A. (2014). Reducing mental illness stigma in health care students and professionals: a review of the literature. *Australasian Psychiatry* 22 (6): 579–584.

21 Health Practitioners Competence Assurance Act 2003. Wellington: Government of New Zealand, 2003. http://www.legislation.govt.nz/act/public/2003/0048/latest/DLM203312.html (accessed 6 August 2017).

22 Brown, R.L. and Schneidman, B.S. (2004). Physicians' health programs: what's happening in the USA? *Medical Journal of Australia* 181: 390–391.

23 Mandatory Reporting; COAG Health Council 2017. www.coaghealthcouncil.gov.au/Portals/0/COAG%20Health%20Council%20Communique%20-%204%20August%202017.pdf (accessed 13 August 2017).

24 Bismark, M.M., Spittal, M.J., Plueckhahn, T.M., and Studdert, D.M. (2014). Mandatory reports of concerns about the health, performance and conduct of health practitioners. *Medical Journal of Australia* 201: 399–403.

25 Quality Assurance Agency for Higher Education (QAA). www.qaa.ac.uk/en (accessed 13 August 2017).

26 World Federation of Medical Education (WFME). http://wfme.org/accreditation (accessed 7 August 2017).

27 General Medical Council (2016). Promoting Excellence: Standards for Medical Education and Training

28 Saunders, P.A., Tractenberg, R.E., Chaterji, R. et al. (2007). Promoting self-awareness and reflection through an experiential mind-body skills course for first year medical students. *Medical Teacher* 29 (8): 778–784.

29 Hassed, C., de Lisle, S., Sullivan, G., and Pier, C. (2009). Enhancing the health of medical students: outcomes of an integrated mindfulness and lifestyle program. *Advances in Health Sciences Educations: Theory and Practice* 14 (3): 387–398.

30 Wasson, L.T., Cusmano, A., Meli, L. et al. (2016). Association between learning environment interventions and medical student well-being. *Journal of the American Medical Association* 316: 2237–2252.

31 Tweed, M.J., Bagg, W., Child, S. et al. (2010). How the trainee intern year can ease the transition from undergraduate education to postgraduate practice. *New Zealand Medical Journal* 123 (1318): 81–91.

32 O'Brien, B.C., Poncelet, A.N., Hansen, L. et al. (2012). Students' workplace learning in two clerkship models: a multi-site observational study. *Medical Education* 46 (6): 613–624.

33 Cross, D. and Lester, L. (2014). Pastoral care: A 10-step plan. www.teachermagazine.com.au/articles/pastoral-care-a-10-step-action-plan (accessed 28 June 2018).

34 Vogan, C.L., McKimm, J., Da Silva, A.L. and Grant, A. (2014). Twelve tips for providing effective student support in undergraduate medical education. *Medical Teacher* 36: 480–485.

35 Wiggins, G. (2012). Seven keys to effective feedback. *Education Leadership* 70: 10–16.

36 Ramani, S., Gruppen, L., and Kachur, E.K. (2006). Twelve tips for developing effective mentors. *Medical Teacher* 28: 404–408.

37 Sandars, J., Patel, R., Steele, H., and McAreavey, M. (2014). Association for Medical Education Europe. Developmental student support in undergraduate medical education: AMEE guide no. 92. *Medical Teacher* 36 (12): 1015–1026.

38 Dunn, L.B., Iglewicz, A., and Moutier, C. (2008). A conceptual model of medical student well-being: promoting resilience and preventing burnout. *Academic Psychiatry* 32: 44.

39 Gordon, J. (2003). Fostering students' personal and professional development in medicine: a new framework for PPD. *Medical Education* 37: 341–349.

40 Bernstein, S., Atkinson, A.R., and Martimianakis, M.A. (2013). Diagnosing the learner in difficulty. *Pediatrics* 132: 210–212.

41 General Medical Council (2015). Gateways to the Professions: Advising medical schools: encouraging disabled students. London: General Medical Council (published 2008; updated July 2015). http://www.gmc-uk.org/education/undergraduate/gateways_guidance.asp (accessed 11 July 2017).

42 Hays, R.B., Lawson, M., and Gray, C. (2011). Problems presented by medical students seeking support: a possible intervention framework. *Medical Teacher* 33: 161–164.

43 Hicks, P.J., Cox, S.M., Espey, E.L. et al. (2005). To the point: medical education reviews – dealing with student difficulties in the clinical setting. *American Journal of Obstetrics and Gynecology* 193: 1915–1922.

44 Medical Schools' Council and General Medical Council (2015). Supporting medical students with mental health conditions (published July 2013; updated July 2015). http://www.gmc-uk.org/education/undergraduate/23289.asp (accessed 11 July 2017).

45 Norrish, M.I., Kumar, P.A., and Heming, T.A. (2014). Interim identification of 'at risk' students: a predictive model. *Journal of Contemporary Medical Education* 2: 199–203.

46 Winston, K.A., van der Vleuten, C.P., and Scherpbier, A.J. (2014). Prediction and prevention of failure: an early intervention to assist at-risk medical students. *Medical Teacher* 36: 25–31.

47 Yates, J. (2011). Development of a 'toolkit' to identify medical students at risk of failure to thrive on the course: an exploratory retrospective case study. *BMC Medical Education* 11: 95.

48 Thomas, L. and May, H. (2010). Inclusive learning and teaching in Higher Education. www.heacademy.ac.uk/knowledge-hub/inclusive-learning-and-teaching-higher-education (accessed 28 June 2018).

49 Rolfe, I. and McPherson, J. (1995). Formative assessment: how am I doing? *Lancet* 345: 837–839.

50 Mackenzie, L. (2002). Briefing and debriefing of student fieldwork experiences: exploring concerns and reflecting on practice. *Australian Occupational Therapy Journal* 49: 82–92.

51 Shin, H.I. and Jeon, W.T. (2011). 'I'm not happy, but I don't care': help-seeking behavior, academic difficulties, and happiness. *Korean Journal of Medical Education* 23: 7–14.

52 Sandars, J., Patel, R., Steele, H., and McAreavey, M. (2014). Developmental student support in undergraduate medical education: AMEE guide no. 92. *Medical Teacher* 36: 1015–1026.

53 Flynn, E., Woodward-Kron, R., and Hu, W. (2015). 2015 training for staff who support students. *Clinical Teacher* 12: 1–6.

54 Amarasuriya, S.D., Reavley, N.J., Rossetto, A., and Jorm, A.F. (2017). Helping intentions of undergraduates towards their depressed peers: a cross-sectional study in Sri Lanka. *BMC Psychiatry* 17: 40.

55 Hu, W.C., Flynn, E., and Mann, R. (2017). Woodward-Kron from paperwork to parenting: experiences of professional staff in student support. *Medical Education* 51 (3): 290–301.

56 Hochschild, A. (1983). *The Managed Heart: Commercialization of Human Feeling*. Berkeley: University of California Press.

57 Lyndon, M.P., Strom, J.M., Alyami, H.M. et al. (2014). The relationship between academic assessment and psychological distress among medical students: a systematic review. *Perspectives in Medical Education* 3: 405–418.

58 Evans, D.J., Zeun, P., and Stanier, R.A. (2014). Motivating student learning using a formative assessment journey. *Journal of Anatomy* 224 (3): 296–303.

59 Cassady, J.C. and Gridley, B.E. (2005). The effects of online formative and summative assessment on test anxiety and performance. *Journal of Technology, Learning and Assessment* 4 (1).

60 Heeneman, S., Oudkerk, P.A., Schuwirth, L.W., Van der Vleuten, C.P. and Driessen, E.W. (2015). The impact of programmatic assessment on student learning: theory versus practice. *Medical Education* 49 (5): 487–98.

61 Chen, Y., Henning, M., Yielder, J. et al. (2015). Progress testing in the medical curriculum: students' approaches to learning and perceived stress. *BMC Medical Education* 15: 147.

62 Bloodgood, R.A., Short, J.G., Jackson, J.M., and Martindale, J.R. (2009). A change to pass/fail grading in the first two years at one medical school results in improved psychological well-being. *Academic Medicine* 84 (5): 655–662.

63 Medicine Society (MEDSOC). First Year Peer Mentoring Program, University of New South Wales, Sydney. https://student.unsw.edu.au/medsoc (accessed 13 August 2017).

64 Mistry, R.D. (2016). A near-peer talk to allay year 1 student anxieties and misconceptions over assessments. *Medical Education* 50: 1154–1155.

65 Sastre, E.A., Burke, E.E., Silerstein, E. et al. (2010). Improvements in medical school wellness and career counselling: a comparison of one-on-one advising to an advisory college program. *Medical Teacher* 32 (10): 429–435.

66 AMSA (2013). Australian Medical Students' Association Mental Health publication. Keeping your Grass Greener. http://mentalhealth.amsa.org.au/wp-content/uploads/2014/08/KYGGWebVersion.pdf (accessed 13 August 2017).

67 Spielman, S., Hughes, K., and Rhind, S. (2015). Development, evaluation, and evolution of a peer support program in veterinary medical education. *Journal of Veterinary Medical Education* 42 (3, Fall): 176–183.

68 Mental Health First Aid website. https://mhfa.com.au (accessed 21 July 2017).

69 Bond, K., Jorm, A.F., Kitchener, B.A., and Reavley, N.J. (2015). Mental Health First Aid training for Australian medical and nursing students: an evaluation study. *BMC Psychology* doi: 10.1186/s40359-015-0069-0.

70 Davies, B., Beever, E., and Glazebrook, C. (2016). The mental health first aid eLearning course for medical students: a pilot evaluation study. *European Health Psychologist* 18 (S): 861.

71 Thompson, D., Goebert, D., and Takeshita, J. (2010). A program for reducing depressive symptoms and suicidal ideation in medical students. *Academic Medicine* 85 (10): 1635–1639.

72 Seritan, A.L., Rai, G., Servis, M., and Pomeroy, C. (2015). The Office of Student Wellness: innovating to improve student mental health. *Academic Psychiatry* 39 (1): 80–84.

73 Malau-Aduli, B.S. (2011). Exploring the experiences and coping strategies of international medical students. *BMC Medical Education* 11: 40.

74 Yamada, Y., Klugar, M., Ivanova, K., and Oborna, I. (2014). Psychological distress and academic self-perception among international medical students: the role of peer social support. *BMC Medical Education* 14: 256.

75 Huhn, D., Huber, J., Ippen, F.M. et al. (2016). International medical students' expectations and worries at the beginning of their medical education: a qualitative focus group study. *BMC Medical Education* 16: 33.

76 Australian Indigenous Doctors Association (2016). Policy Statement: Best Practice in Student Support of Indigenous Medical Students. www.aida.org.au/wp-content/uploads/2015/03/20161201-Best-Practice-in-Student-Support-for-Indigenous-Medical-Students-AIDA-policy-statement.pdf (accessed 8 July 2017).

77 LIME Network (Leaders in Indigenous Medical Education). www.limenetwork.net.au (accessed 28 June 2018).

78 American Medical Students Association (2017). Gender and Sexuality Student Resources. http://www.amsa.org/advocacy/action-committees/gender-sexuality/gender-and-sexuality-student-resources (accessed 8 July 2017).

79 Walker, C., Gleaves, A., and Grey, J. (2006). Can students within higher education learn to be resilient and, educationally speaking, does it matter? *Educational Studies* 32: 251–264.

80 Howe, A., Smajdor, A., and Stöckl, A. (2012). Towards an understanding of resilience and its relevance to medical training. *Medical Education* 46 (4): 349–356.

81 Rutter, M. (2006). Implications of resilience concepts for scientific understanding. *Annals of the New York Academy of Sciences* 1094: 1–12.

82 Rogers, D. (2016). Which educational interventions improve healthcare professionals' resilience? *Medical Teacher* 38 (12): 1236–1241.

83 Pidgeon, A.M., Ford, L., and Klaassen, F. (2014). Evaluating the effectiveness of enhancing resilience in human service professionals using a retreat based mindfulness with Metta training program: a randomised control trial. *Psychology, Health & Medicine* 19: 355–364.

84 McAllister, M. and McKinnon, J. (2009). The importance of teaching and learning resilience in the health disciplines: a critical review of the literature. *Nurse Education Today* 29 (4): 371–379.

85 Martin, A.J. and Marsh, H.W. (2006). Academic resilience and its psychological and educational correlates: a construct validity approach. *Psychology in the Schools* 43: 267–281.

86 Boccia, M., Piccardi, L., and Guariglia, P. (2015). The meditative mind: a comprehensive meta-analysis of MRI studies. *BioMed Research International* 2015: 419808.

87 Bishop, S.R., Lau, M.A., and Shapiro, S.L. (2004). Mindfulness: a proposed operational definition. *Clinical Psychology: Science and Practice* 11: 230–242.

88 Kabat-Zinn, J. (2003). Mindfulness-based interventions in context: past, present, and future. *Clinical Psychology: Science and Practice* 10: 144–156.

89 Warnecke, E., Ogden, K., Bentley, M., and Nelson, M.R. (2017). 5-year follow-up of a randomised controlled trial of the effects of mindfulness practice on medical practitioners' stress. *MedEdPublish* pp. 11–19. ISSN 2312-7996.

90 Britta, K., Sara, W., Lazar, W. et al. (2011). Mindfulness practice leads to increases in regional brain gray matter density. *Psychiatry Research* 191 (1): 36–43.

91 Krasner, M.S., Epstein, R.M., Beckman, H. et al. (2009). Association of an educational program in mindful communication with burnout, empathy, and attitudes among primary care physicians. *Journal of the American Medical Association* 302 (12): 1284–1293.

92 Frankl, V. (1959). *Man's Search for Meaning: An Introduction to Logotherapy*. Beacon Press.

Further Reading

Dyrbye, L. and Shanafelt, T. (2016). A narrative review on burnout experienced by medical students and residents. *Medical Education* 50 (1): 132–149.

Enhancing Student Wellbeing. http://unistudentwellbeing.edu.au (accessed 28 June 2018).

Slavin, S.J. (2016). Medical student mental health. Culture, environment and the need for change. *Journal of the American Medical Association* 316 (21): 2195–2196.

The LIME website includes many resources, including good practice case studies on student recruitment and support. www.limenetwork.net.au/resources/good-practice-case-studies (accessed 28 June 2018).

Vogan, C.L., McKimm, J., Da Silva, A.L. and Grant, A. (2014). Twelve tips for providing effective student support in undergraduate medical education. *Medical Teacher* 36: 480–485.

Wasson, L.T., Cusmano, A., Meli, L. et al. (2016). Association between learning environment interventions and medical student well-being. *Journal of the American Medical Association* 316: 2237–2252.

34 Managing Remediation

Deborah Cohen[1], *Melody Rhydderch*[2], *and Ian Cooper*[3]

[1]Centre for Psychosocial Research, Occupational and Physician Health, School of Medicine, Cardiff University, Cardiff, UK
[2]Wales Deanery, Cardiff, UK
[3]School of Medicine, Cardiff University, Cardiff, UK

KEY MESSAGES

- Performance is a function of ability, motivation, personality, and organisational issues.
- Performance must be considered in the context of the individual's health and well-being, and the educational system in which they are situated.
- Moving into 'action' too early – providing solutions and fixing problems – may lead to disengagement, a loss of ownership, and reduced chance of successfully achieving change.
- High-quality records should be kept throughout the remediation process.

Introduction

Approaches and models to identify and assess performance concerns in doctors and medical students are well documented. However, views about 'how' and 'when' to deliver and manage remediation programmes remain varied and under discussion.

This chapter will explore how best to approach remediation given two key assumptions. First, that performance is a function not only of ability, but of motivation, personality, and organisational factors. Second, that performance is also a function of health, well-being, and the surrounding educational system. Given these assumptions, there is a clear argument that a holistic approach is most likely to be effective in developing engagement and long-term positive outcomes. Tutors who support medical students, supervisors who support doctors in training, and line managers who support doctors post-training all have a crucial role to play in helping the remediation process run smoothly. Moving into 'action', that is fixing problems and trying to find solutions too early, should be avoided. This can lead to a lack of engagement, resistance to the process, and reduced chance of successfully achieving change. Prescribed doses of training and didactic and generic teaching where the individual, their health, and their personal situation are not considered are unlikely to be successful.

Key principles to consider when entering any remediation process or programme include the following:

- getting the right system in place
- personalised remediation
- engagement and motivation
- taking account of organisational culture
- clarity of roles and boundaries.

We aim to provide a guide to how best support effective remediation across the continuum of medical education, training, and development, from medical students to senior consultants, but it should be kept in mind that every situation is different and that the principles and models described require thoughtful interpretation in relation to the individual entering remediation, the organisational or institutional context, and the surrounding legal or regulatory framework. The chapter will draw on evidence from the world of medical education, rehabilitation, and work performance and should be read in conjunction with other contributions to this book – particularly Chapter 17 (The Development of Professional Identity) and Chapter 33 (Supporting Learner Well-being) – in order to give the reader a broader perspective on the issues raised.

Drawing on our experience at Cardiff University, where work has been undertaken since 2001 to support doctors and medical students who are struggling, we will take the reader through the journey of remediation to include processes and systems, dialogue with the struggling doctor or student, how to explore problems in a sensitive way (including health and disability), and note-keeping. We offer advice on how to navigate and manage the process of the provision and management of remediation and a case study using a doctor identified with a performance concern is used to explore possible pathways for remediation. Towards the end of the chapter, we also signal some common problems and pitfalls.

Performance and Remediation

The management of performance issues in doctors and medical students continues to pose considerable challenges to employing organisations, educational bodies, and regulatory bodies. The problem is complex and goes beyond a

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

simple question of *ability*, since personality, motivation, and organisational factors all impact on individual performance [1]. Models of performance assessment are now well researched and have evolved with a consistent approach worldwide. Models of remediation, however, do not yet have that consistency. Examples of good practice do exist [2–9], but evidence of long-term outcomes is harder to establish [9–11].

The early literature and methods of remediation often considered performance separately from an individual's health and well-being. More recently, a broader and more holistic approach has been recognised as most effective, where health and the quality of the learning or organisational environment are considered alongside specific performance indicators.

A more proactive approach to remediation is required if organisations and educational institutions are to bring about change. A culture and attitude that encourages medical students and doctors to seek help early before health, well-being, or educational issues impact on performance is required. This must be accompanied by support structures and interventions that promote such an ethos. Integrating well-being and an openness about mental ill health in the profession into undergraduate and postgraduate curricula is necessary. This is now being addressed internationally, with many advances in thinking about how this might best be achieved [12, 13]. The stigma and shame associated with failure, and particularly mental illness, still predominates across medicine [14]. Literature around 'disclosure' has improved and understanding the obstacles doctors and medical students face when making decisions about who and how to disclose is better understood. Issues around provision of support, how to access it, and uncertainty about confidentiality are still major factors that need to be overcome [15, 16].

From Assessment to Remediation

Historically, the assessment of performance was targeted at the small minority of doctors whose practice gave substantial rise for concern. In the UK, such doctors were typically referred directly to the regulatory body, the General Medical Council (GMC). However, the introduction and embedding of revalidation into UK medical practice has led to a growing recognition that performance concerns must be dealt with promptly and locally where possible [16, 17]; that is, when concerns are emerging, not when they have become of serious concern and are warranting investigation. This shift in approach is mirrored internationally [10, 11, 18] and is iterated in documents such as the GMC's *Good Medical Practice* [19], that lays out the duties and responsibilities of doctors that practise in the UK. Alongside regular appraisal and performance review, *Good Medical Practice* also emphasises the importance of good health and doctors managing their own health proactively. *Good Doctors, Safer Patients* [20] spells out the importance of physician health on performance and patient safety and Boorman [21], in his review of the UK National Health Service (NHS), highlights staff well-being as key to improved performance and patient safety. Routine performance assessments have therefore increased, paralleled by a decreased tolerance for poor performance. An important outcome of this change has been that those who manage or support doctors and/or medical students locally are now required to develop their own knowledge and skills around remediation and associated processes, and to respond effectively to the needs of those for whom they are responsible.

Remediation is an intervention, or suite of interventions, required in response to some form of assessment against threshold standards of performance. To deliver effective remediation, *tutors* who support medical students, *supervisors* that support doctors in training, and *line managers* who support doctors after completion of training must be sensitive to the assessment process and be able to provide a flexible response. They must be able to offer personalised support and direction to the individual, whilst being sensitive and responsive to the organisation(s) or educational institutions in which they work. The problem may often lie in the organisational or institutional structure and culture, rather than solely with the individual referred [13, 22, 23]. Once a concern is raised tutors, supervisors, and line managers must consider the individual's health and well-being, personality, and motivation to change alongside organisational and social issues [1, 23–25]. Rising to these challenges is no easy task. Identifying that someone may be struggling is the first step, how then to support them with their performance is complex. Some early workplace signs for doctors, but which are as applicable to medical students in training, are contained in Box 34.1.

What follows are some principles that we believe to be fundamental to successful remediation.

BOX 34.1 FOCUS ON: Identifying the doctor in difficulty

The doctor who is struggling has many different faces. Paice [26] describes certain 'early warning signs' of a trainee doctor in difficulty. Some of the attributes she describes include the following, and these are as applicable to medical students as they are to doctors:

- the doctor who is often difficult to find (*the disappearing act*)
- the doctor who is always at work but achieves less than their colleagues (*low work rate*)
- the doctor who is quick to lose their temper (*ward rage*)
- the doctor who is inflexible and has difficulty prioritising (*rigidity*).

Other doctors she describes as 'displaying problems with their career choice or having difficulty obtaining their required exams' (*career problems*), while others 'lack insight' and reject constructive criticism (*insight failure*). Finally, there are those who find ways of 'sidelining' the difficult doctor (*bypass syndrome*).

Principles to Guide Remediation

Getting the Right System in Place

This is perhaps the biggest challenge. As already alluded to, many performance concerns do not just relate to educational issues but are also a function of organisation, social, financial, and health issues [16, 23]. These must all be addressed if remediation is to be successful. The culture of the organisation and/or institution is fundamental to successful outcomes where performance and productivity are central to success of an organisation [21, 27]. Tutors, supervisors, and line managers are central in influencing the culture [16, 21, 27] and as such impact on an individual's perception of support and their help-seeking behaviours. The provision of proactive systems of support is powerful and encourages early help-seeking.

The key features of good systems include the following:

- well-defined pathways and systems – for help-seeking
- guidance for tutors, supervisors, and line managers – as to what to do if they suspect a performance issue
- thoughtful separation of roles – between those who identify or assess a student or doctor's performance and those that provide support or remediation
- transparency – particularly around information sharing and confidentiality as perceptions of what happens to personal information influences whether someone will actively seek help
- a proactive system – to encourage self-awareness, overcome a culture of stigma, and promote well-being.

A challenge, notably for doctors in training, is a lack of clarity around the interface between workplace procedures for support and those provided by their educational institution. This confusion as to 'how' and 'where' to access support and who 'needs to know' is well documented as an important obstacle to early help-seeking and effective support systems for both medical students and doctors [15].

Personalisation

There are different approaches to remediation internationally. Some call for a more 'medical' model to remediation [28] while others, driven by funding and resource issues, call for a standardised system with 'courses' focused on particular issues, which seems a more effective solution. In general, responses to low-level concerns should be considered differently to more serious concerns. For instance, a medical student who is struggling with new knowledge, in a new environment, or is under pressure of being 'the best' in a new cohort may be dealt with differently to those who have persistent and more serious concerns around behaviour or attendance. Interventions developed in our own institution for groups of students centred around, for instance, time management, are effective and bring a sense of cohesion and normality rather than isolation and stigma to students. For more serious concerns, evidence from the world of rehabilitation consistently shows that a case management model – a holistic approach to an individual's needs – is most effective in bringing about change and improved performance [29]. The purpose of personalised, as opposed to group, remediation is to match provision closely to individual need and to take health, personality, and motivation to change into account. It is not about a standard prescription and must be flexible to both need and resources available.

Personalised remediation can, and should, be tailored to different learning preferences and an integrated, timely, and practical approach to remediation that enables the individual to continue working or studying without feeling overwhelmed by additional requirements or training must be at the centre of a good remedial programme. Tutors, supervisors, and line managers need to be aware of the ways in which other disciplines such as occupational health, occupational psychology, and language experts can contribute to remediation. The challenge often is the availability of resources and how to access them. This might be either via the workplace, the educational institution, or specialist services nationally or locally. Across the UK there are now specialist support services for doctors with complex mental health issues, provision that has grown over the last decade prompted by the establishment of revalidation and appraisal processes. The GMC, who also oversee the educational requirements of undergraduate training in the UK, have also recognised the need for a focus on support and remediation for students, particularly for those with mental ill health and disabilities [30, 31].

Engagement and Motivation

Goulet et al.'s [3] work on personalised remediation within the Canadian health system suggested that a doctor's motivation may initially be low, but that the involvement of a licensing authority has a positive effect on their motivation and cooperation. However, whilst the high stakes may motivate a person to agree to *attend* for remediation, it is not always enough to stimulate genuine engagement. Engaging doctors and medical students with the process and motivating them towards change requires careful planning. Using well established models of engagement, such as shared decision-making [32] and behaviour change [33], are at the centre of planning remediation and case management. In Cardiff our own preferred evidence-based method for behaviour change is motivational interviewing [25, 33], an established method used across health care in supporting sustained change in behaviour [34]. For those involved in remediation, understanding how to best provide support and promote behaviour change using evidence-based interventions is key to a successful outcome. In addition, spending time finding out what an organisation or educational institution can offer and how to access services and training before embarking on support will reap rewards. This adds to the confidence of the individual entering remediation and of the tutor, supervisor, or line manager managing the process. It builds a stronger and more effective relationship, something that is well documented as central to successful rehabilitation and performance management [21, 27].

Organisational Issues

Organisational factors may both trigger and perpetuate individual poor performance. Unsupportive work environments in the form of high work demands, role ambiguity, poor team working, and punitive cultures can all serve to undermine a doctor striving to maintain newly acquired behaviours. This is no different for medical students who are struggling to sustain their new found identity. Students are strongly influenced by the 'hidden curriculum' [35]. They too struggle with work demands and what is often perceived as a culture of recurrent assessment and unsupportive teaching environments [36]. For both doctors in training and medical students, there is the added pressure of moving between placements and jobs and building new relationships every few months. This takes its toll, with feelings of isolation and reduced self-confidence adding to the potential risks related to performance [37].

There is also a recognised pattern of behaviour among a subset of doctors who demonstrate a tendency to 'rage against the machine'. This group are characterised by a continued need to fight and blame the whole system, rather than considering the relevant importance of each battle and indeed considering which battles are worth fighting. It is sometimes said of this group that they particularly lack insight into their own issues and performance needs. This brings added challenges as this group can often be very resistant to change.

Clarity of Roles and Boundaries

Finucane et al. [18] have previously argued that performance assessment impacts upon three distinct groups: patients, doctors, and employers. The authors also argued that whilst these groups may have conflicting beliefs and expectations of assessment, the process must be acceptable for all. Transparency of the process is an important prerequisite to the doctor or medical student engaging. Tutors, supervisors, and line managers cannot consider themselves to be entirely independent in their role. They are in effect an instrument of their educational or employing institution. They must be explicit in their role and the boundaries that they work within. They must take care not to stray beyond those boundaries and roles. There is often an initial feeling of mistrust by those entering remediation. This is commonly associated with concerns about confidentiality and a feeling that they lack control over the process. For medical students, there is a profound concern around fitness to practise and being able to complete their course. Recent work about providing support for medical students with mental health issues highlights the importance of separation of support from assessment [30, 38]. Clarity at the outset about roles and boundaries for all involved helps overcome some of this mistrust. Clarity must extend to communication pathways and confidentiality; it must be made explicit where information is stored and who has access to it. For medical students addressing fears and often myths about fitness to practise, these should be discussed and explored at the outset.

Getting Started

Providing support to a doctor or medical student who presents with performance concerns requires careful consideration. They may feel their career is at risk, and often be resentful of the predicament they find themselves in. They may lack insight into the issues raised or have underlying health, personality, or social issues that may have precipitated performance issues. The aim is to help guide the individual through what can often be a painful and stressful process. Tutors, supervisors, and line managers must recognise that they will not have all the answers or skills required for every individual, but they can guide and signpost people to appropriate support where needed. In essence, they must also have insight into their own limitations as well as the boundaries within which they work. Providing an opportunity for joint solutions through effective conversations should not be underestimated. The relationship you build as their supervisor or tutor will be instrumental in the individual building trust and engagement with the process.

What Type of Conversation?

Several factors will impact upon the conversation. These include the following:

- The relationship between the tutor, supervisor, or line manager and doctor or medical student.
- The processes within the institution or workplace that determine how, when, and where an issue should be managed. Should this be within the organisation or external to it (e.g. locally or nationally)?
- The extent to which the conversation is considered to be formal or informal.
- The nature of the performance concern itself.

The relationship between the individual and their tutor, supervisor, or line manager is key to a good outcome and, as described above, should not be underestimated. This has been shown in many similar workplace settings [21, 27]. Understanding the process for managing performance concerns will enable more effective conversations from the outset. It provides a sound basis for starting a conversation which links to the 'transparency', 'equity', and 'clarity' of a process that sits at the heart of a good organisational process [39]. Tutors, supervisors, or line mangers must understand the triggers and processes for performance review. This will vary across postgraduate and undergraduate education and those who are employed and not in an educational environment. This will help frame the first and, to an extent, the most important conversation. Keeping written records of any dialogue, however informal, is important. Clarifying the recording of information and where information is stored helps with transparency and therefore engagement. The formal processes around meetings and information-recording, although at times difficult, are important to an overall robust and fair process.

There is often a tension for the doctor or medical student about how much information is shared and by whom. Supervisors often complain that they are not given enough information about an individual's previous performance concerns. However, doctors who have experienced performance investigations and remediation argue that unnecessarily sharing information about themselves can cause bias or mistrust. The concerns are similar for medical schools and medical students. Students describe being judged before they even enter a new placement or year group. This then leads to anxiety and stress as they feel they are being constantly watched as they try to learn. Patient safety and probity is of prime importance and must take precedent over confidentiality for those overseeing performance management. It is the clarity of the information exchange, how it is managed, and the extent of it that must be open and honest at the outset of any conversation if engagement and support is to be effective.

Planning the First Discussion

Careful planning reaps benefits and it is important to give this time before meeting with a doctor or medical student for the first time. Basic preliminary information is required, including a summary of the exact nature of the concern and evidence of significant events or complaints. Background information (as described below) is also helpful where it can be obtained prior to the meeting. General concerns about performance but without details are unhelpful. Documentation should be held together in a format that will allow the tutor, supervisor, or line manager to refer to during the meeting and, where appropriate, share some of the information with the doctor or student. Being open and transparent about information helps reduce resistance and mistrust. This will also help the structure of the meeting and the objectives that might need to be set. Setting the agenda at the beginning of the meeting will reduce misunderstandings and build engagement. It is often helpful to begin the meeting with an outline of why the individual is being seen, the way the meeting will be conducted, and any regulatory or procedural issues that need to be covered, and allow the individual time at this early stage to ask questions. Confidentiality and note keeping must also be discussed, as these are major obstacles to engagement for doctors and students under review. This is discussed more fully below in 'Exploring the Problem'. The discussion should aim to cover the presenting problem, factors leading up to the problem, a review of any significant event(s), personal or health-related issues, and, for doctors, relevant past workplace experience. Tutors, supervisors, and line managers must be prepared to discuss and be questioned (or challenged) about the process. Certain actions or protocols dependent on the institution or organisational processes and procedures may need to be negotiated (within certain limits), e.g. consideration as to whether a referral for an occupational health assessment is necessary. This will give the individual a sense of control and share in the decision-making rather than 'enforcing' a decision without discussion.

Obtaining Background Information

Background information informs and shapes the conversation with the individual presenting with a performance concern. Information required includes complete documentation of the concerns raised, by whom, and an idea of the timescale of events. Multiple sources are useful as they help triangulate the information and give a broad picture of what has transpired. Sources of information commonly used in the UK for doctors include significant event reviews and records from the appraisal and revalidation process. Other sources of information include 360° (multi-source) feedback and information from previous employment. For those in training, information available through e-portfolios or placements is also helpful. For medical students, their regular placement reviews, assessments, and exams provide a wealth of information. Most medical students now have academic mentors who follow students through their training. They also act as a source of information as well as providing ongoing support for the students.

Seeking 'off the record' uncorroborated information is neither helpful nor useful to the conversation.

Exploring the Problem

It is important that the individual is given time to discuss their perspective of the problem and of how such problems might have arisen. Providing a space to speak openly and without prejudice is an important part of the engagement process. Discussion about details of critical incidents or performance concerns and their views on events starts to allow ownership of a process at which they may often feel at odds. Listening is critical and checking understanding even more so. Problems that lead to underperformance are often multifaceted and so it is unlikely that the issues will be simple. It is not uncommon that health and personal issues come to the forefront as contributing to concerns and these must be dealt with sensitively. Building in enough time (or the ability to return for a second meeting promptly) is important. Our experience is that in some complex cases the first meeting may take two hours or more.

Allowing the individual to tell their story is often the best starting point for any conversation about performance. Some will often describe strong feelings of injustice and misunderstandings. They often have their own explanations and views as to what has happened and why. Doctors and medical students often describe being oblivious to their performance being under scrutiny. This can fuel their feelings of injustice. Others will seem bewildered by the process and have a deep sense of shame and self-stigmatisation. Active listening (see Box 34.2) will guide individuals through what is at times a personal and painful process. Using active listening can help and build engagement [40]. This is the first step towards a joint understanding of the problem and a move towards behaviour change [33].

Knowing when to respect a doctor's or medical student's privacy is equally as important as encouraging them to talk. There may come a point in the conversation where it becomes clear that there is a personal or health issue compounding or contributing to the performance concern. It is here that the tutor, supervisor, or line manager must remember their role and boundaries and guide the doctor or medical student appropriately. Their role is not to provide medical advice, treatment, or counselling. It is important to acknowledge there may be other issues related to the performance concerns and a general understanding of those issues might help direct the individual to the correct support. Seeking help must be encouraged. Understanding the organisational or educational process is also crucial at this stage. It is important to ensure that, where needed (or required) for a health issue, the individual should be referred or seek self-referral to occupational health for an expert opinion. This will be dependent on the health or personal issues raised and the performance concern. In some cases, at both undergraduate and postgraduate levels, an

BOX 34.2 FOCUS ON: Active listening

Active listening is a client-centred technique expounded and developed by Carl Rogers [40], ensuring that cues are recognised and explored. Rollnick, Butler, and McCambridge [33] describe 'three styles' of communication that can enhance this process.

Instruct: Give information or advice. Other activities associated with this style include directing, informing, leading, educating, telling, and using one's expertise. It is used when there is information one wants to provide that hopefully the person wants to receive.

Guide: Encourage the person to set his or her own goals and find ways of achieving them. Other activities associated with this style include coaching, negotiating, encouraging, mobilising, and motivating. It is used when the person is facing change, having to make decisions and to act upon them.

Follow: Understand the person's experience. Other activities used include gathering information, eliciting, attending, and empathising. It is used when one wishes to understand how the person feels or what has happened to him or her.

occupational health assessment is a necessary requirement for the process of remediation and is part of a regulatory process. One point of concern often voiced and often seen as a frustration is that the decision to seek support falls with the individual and you 'can't make them go'. This is correct and whether they need to go to occupational health, their family physician, or a specialist, the decision ultimately rests with the individual. However, raising the concern for someone's health in a supportive non-judgemental environment may help. It can start the conversation and allow the individual to reflect on the pros and cons of help-seeking and so arrive at their own solution. This fits with the model of behaviour change where individuals can be guided and supported to a decision rather than being 'told' what they must do. This approach builds engagement and supports behaviour change.

The supervisor, tutor, or line manager must document that the suggestion was made and perhaps revisit this at a later stage.

A framework for the first conversation as alluded to above is often helpful. This should be based on what are likely to be the individual's mains concerns and should offer them the opportunity to ask questions before entering into a detailed conversation. Areas to cover would include the following:

- what information has already been collected
- what will happen to previous and new information collected
- how it might be used
- where it will be stored
- who might have access to it, now or in the future.

This then provides the backdrop to start a more detailed conversation about the concerns raised. Documentation of significant events should be detailed, continually checking understanding of the doctor or medical student's perspective. Move away from confrontation. Challenging the individual with discrepancies in information recorded at this stage can lead to resistance and difficulty with engagement. Exploring previous significant events is helpful. For doctors, this should extend to previous employment and their time as a student. For medical students, it is often helpful to ask about potential issues whilst at school as well as health issues that may have arisen prior to entering training. Do not focus just on the present situation. This often will help elicit an emerging pattern of behaviour. Useful questions to guide the review include the following:

- What led up to the event?
- What actually happened?
- Was the way you managed the interaction effective or ineffective and why?
- On reflection, how else could it have been handled?

Taking an evidence-based approach to guide these types of conversations is invaluable. Behaviour change counselling sits at the heart of the description above. This approach evolved from the work of Carl Rogers [40] and was defined first by Miller and Rollnick [25]. Behaviour change counselling is a client-centred approach to discussing lifestyle changes. Rollnick describes it as: 'Ways of structuring a conversation to maximise the individual's freedom to talk and think about change in an atmosphere free of coercion and the provision of premature solutions.' By working carefully through a detailed history of events, a picture of behavioural patterns often starts to emerge. By using these principles and allowing the individual ownership of the story, it is our experience that they will often be quite open in providing information about difficulties and problems they have encountered in the past. The first discussion can provide great insights into how and why they might be struggling.

Where Next?

The first discussion is likely to provide a wealth of information about the performance concerns and underlying issues. If conducted effectively, from a neutral position, then the groundwork has been laid towards effective engagement. At this stage, it may be clear that further, more detailed information might be needed. It is important for the tutor, supervisor, or line manager to recognise and respect the boundaries of their role, whatever their background expertise. It may be the case that their role is to act solely as a case manager and be the coordinating central point of support. Such a role is to understand in depth the varying issues that might be impacting on performance and signpost and liaise with those who can provide treatment, support, or remediation. It may be the case in some organisations and institutions that the role is to deliver the remediation. A key message here is that it is not appropriate to treat the individual or give clinical advice. The tutor, supervisor, or line manager must and should help manage these boundaries. They must recognise that it is often not appropriate nor needed for them to have sensitive personal information about the person they are managing [30, 38].

The skill for the tutor, supervisor, or line manager is to know 'when', 'how', and 'who' to refer to for support, guidance, and/or expert opinions. They must develop robust communication channels. This will help give the medical student or doctor confidence in the process and remain engaged. It is often the case with complex performance problems that issues are multifactorial. A 'biopsychosocial approach' to understanding the performance concerns will enable more effective management.

A now well-accepted model to a holistic approach to performance concerns is to consider the impact and relevance of the following factors:

- past medical history (physical and mental health, the presence of long-term conditions)
- personal issues (family, relationship issues, financial concerns)
- personality
- language and communication skills
- cultural issues.

However, a key message at this point is to understand that there are experts whose legitimate professional role is to assess the potential relationship between each of the above factors and work performance. The 'Chinese wall' created by referring an individual to other health professionals to discuss issues associated with health, personality, language, and culture, which may or may not be related to performance, ensures that potentially deeply personal information can remain private to the individual where appropriate. There is a word of caution here. There are times when patient safety or probity becomes a significant issue. The tension is to understand with whom information should then be shared. This must be discussed with the individual at the outset and not when such a concern is raised. Agreeing a form of words and providing a document for the individual to sign when they enter into a process of assessment for remediation provides transparency and clarity for both parties and can help overcome possible difficult issues later.

The first discussion (and perhaps subsequent discussions) will provide an understanding of what factors might be impacting on the individual's performance. To build an effective relationship between the two parties requires clarity but also openness. The ability to be honest yet sensitive and not collude with the doctor or student when difficult issues are raised is not easy. The discussion should hinge around being able to provide access to experts to help ensure the individual receives the best possible help and support. This must range from expertise in physical and mental health as well as behavioural and educational needs.

Doctors and medical students are known to have significantly higher rates of common mental health disorders than the general population, so being aware of this throughout discussions is helpful [16]. What at face value might present as aggressive or bullying behaviour might on careful evaluation suggest underlying depression, anxiety, or more serious conditions such as bipolar disorder or cognitive impairment. Research has shown that doctors and students often fail to disclose mental ill health until they are desperate for help [15]. Physical ill health may also be an issue, for instance excessive fatigue linked to some long-term conditions may impact on performance. Finally, developing a clear understanding of the individual's social environment and placing this in the context of the problems that they are now facing is important in understanding *why* the individual has started to develop difficulties at this point. Box 34.3 provides a framework for understanding how these potential factors may lead up to and perpetuate performance concerns.

To conclude, the challenge for those supporting an individual with a performance concern is to determine the nature of the problem and the possible causes, whether more information is required from other sources, and how to access those resources. Coordinating support and remediation, maintaining dialogue, but not allowing 'silo working' is of prime importance if remediation is to be effective. What follows is a description of the kinds of information that other professionals should be able to provide.

Digging Deeper

Occupational Health Assessments

It is now generally accepted that doctors do present a special case in relation to the need for workplace support [16].

BOX 34.3 Biopsychosocial factors and stages in a doctor's underperformance

Factors	Biological	Psychosocial	Social
Predisposing	Underlying mental or physical disease	Personality Family	Cultural Family
Precipitating	Acute ill health events	Interactions at work	Economic factors Social isolation The culture of the organisation
Perpetuating	Chronic disease	Lack of insight by organisation or individual	Economic Cultural Organisational

Source: Adapted from Sharpe and Wilks [44].

For medical students, there is growing evidence that this is also the case [30, 31]. The major health concerns for doctors are common mental health problems, addictions, burnout, and a high rate of suicide [41, 42]. For medical students, mental ill health also predominates, with high levels of anxiety, depression, and eating disorders [43]. It is the timely access to specialist mental health providers that is required.

Employers and educational institutions should work in close collaboration with occupational health providers for advice and support in the assessment and possible management of doctors and medical students who present with a performance concern. An occupational health physician provides an independent assessment of an individual's health in relation to their work and workplace. They make assessments about how the workplace may impact on an individual's health and vice versa. An assessment and independent opinion from the occupational health provider can help provide timely and appropriate advice from specialists where needed and help liaise closely with the general practitioner or family physician who would provide the ongoing care for the individual.

For instance, a doctor who presented with poor team management and prioritisation might on detailed questioning admit to a previous head injury or cerebrovascular accident that could indicate a neuropsychological impairment. A presentation of aggressive or inappropriate behaviour might display features suggestive of underlying mental health issues such as depression, addiction, or personality disorder. Once a problem has been identified, onward referral for specialist opinion can be made, where appropriate.

Some doctors may have a long-term condition that impacts on their performance that falls within the legal definition of a disability. Relevant legislation must be considered in supporting them back into the workplace [24]. Once again, an occupational health physician can provide advice about elements of a remediation programme that would support the doctor or medical student whilst at the same time provide the organisation or educational institution with the legislative knowledge and advice they must follow, e.g. in relation to equalities. In the UK this is known as provision of 'reasonable adjustments' [45]. So, when constructing a programme of support for a doctor with health issues, advice from an occupational health physician is invaluable.

Behavioural Assessments

Understanding behavioural habits can help to establish effective individualised remediation. Behaviour is a function of personality, learnt behaviour, and situational drivers in the organisation. The purpose of this section is to indicate the kind of information an occupational psychologist could bring to understanding the relationship between personality and performance. It also distils some key messages about the impact of personality on remediation.

The link between personality and performance has been studied for many years, and research has mostly been conducted in an attempt to help the selection procedures. Research up to the late 1980s suggested that there was little, if any, association between personality and performance. However, this theory was dismissed with the emergence of the five-factor model of personality (see Box 34.4) [47–50].

Competency and curricula frameworks worldwide stress a number of different aspects that are required of a doctor, including the importance of good communication skills, being able to work well with colleagues and patients, technical competence, and general skills [51–53]. General skills include managing one's time, prioritising, being self-critical, problem solving, and analysing numerical data. The strategies individuals adopt to manage these skills are personality dependent.

One popular personality indicator that is widely used in health care organisations for personal development and growth is the Myers–Briggs Type Indicator (MBTI) [54]. If used correctly – and one reason for the MBTI's popularity is that anyone with an interest in coaching, whatever their background, can be trained to use it – MBTI has the ability to help individuals understand why they behave the way they do, and why they find certain situations difficult. Most importantly, it helps them explore a range of behavioural choices for different situations. Another useful psychometric tool is the Hogan Development Survey, which identifies personality-based performance risks and 'derailers' of interpersonal behaviour [55] and can be extremely useful in highlighting to individuals how they will tend to respond to pressure and stress.

Our experience is that it is important to work with the 'grain' of an individual's personality; using behavioural assessments helps remind the individual and those providing support that 'one size does not fit all'. Any suggested strategies for change are best sustained if they are close to the way an individual tends to normally do things. Small but sustainable shifts in behaviour are powerful.

BOX 34.4 WHERE'S THE EVIDENCE: Personality and performance

Since the early 1990s many different personality tests and performance in many different jobs have been investigated. Overall, it has been concluded that personality can be reliably measured using appropriate tests, and for most jobs, certain personality traits can predict job performance [46, 47]. Research has shown that in a large number of occupations job performance, and indeed success at medical school, can be predicted by the big five personality factors [48, 49]:

- extraversion
- agreeableness
- emotional stability
- openness
- conscientiousness.

Barrick et al. [50] suggest that although not all of the big five personality factors predict generalisable job performance, they do predict success in specific occupations.

Communication and Language Assessments

Assessment of communication sits at the heart of good remediation programmes and assessments are widely available, both nationally and internationally. However, the same cannot yet be said of language assessments. Yet this type of information can provide an invaluable baseline for working with individuals. Most educational institutions have experts in communication skills who can help provide assessments. Recognition of the importance of language and culture should not be underestimated. Extending assessments to language and seeing this as an assessment separate from clinical communication should be considered where appropriate. Simply observing a doctor consult can provide many insights for those who are trying to provide remediation.

Clinical Communication

Poor communication with patients and/or colleagues is frequently cited as a core indicator of performance problems. There are many methods and approaches to assessment of communication. Below is offered a method that, in our own institution, has helped engage individuals with the process and enable them to build insight into their own strengths and weaknesses. Using observed simulated patient consultations offers the opportunity for the individual to evaluate their own consulting style and behaviour alongside models of good practice in a non-judgemental manner. It helps the individual develop some insight into their own practice [56]. It can provide a good baseline indicator of their active listening and shared decision-making skills, as well as being an indicator of their likely approach to communicating with colleagues.

It is important to note that the object of the exercise is to continue to engage the individual to feel 'part of the process' whilst assessing their 'needs'. This exercise is not meant to serve as a challenge. Often, asking the individual to choose a scenario or interaction that they feel comfortable with is a good starting point. The use of challenging scenarios, for instance, breaking bad news or dealing with someone who is angry, may not be the best way to begin this type of assessment. It is intended to provide a baseline only. In other words, the activity here is to understand which skills the doctor relies on in everyday situations.

Self-evaluation is intended to be descriptive and diagnostic, rather than summative. Providing the doctor or medical student with some sort of structure for this is useful. This could, for instance, be a rating scale adapted from standardised forms used in an Objective Structured Clinical Examination. Using something standardised also provides the opportunity to reflect on progress at a later stage by repeating the exercise.

Language

The assessment of an individual's language skills is not a novel undertaking. In many countries it has been a requirement for international medical graduates for some years. However, understanding how an individual's language skills impact on performance remains a relatively new area.

In the UK, doctors from overseas may have had to complete specific tests before being able to practise. These include the International English Language Testing System (IELTS), and Professional and Linguistic Assessment Board (PLAB) to demonstrate competence in communicating in English. International students applying to study medicine in the UK are also required to reach certain levels of competence in English assessed thorough IELTS. It is good practice for tutors, supervisors, or line managers to be aware of the requirements of their own organisation and/or regulatory body when supporting students and doctors with English as a second language.

For some, one of the factors linked to performance problems may be their use of language. Use should be understood differently from usage. The latter relates more to 'correctness': accuracy of grammar, spelling, and pronunciation. 'Use' relates more to the ability to communicate clearly, coherently, and appropriately with another person with the range of language knowledge and skill at one's disposal. Of course, there may be some overlap between use and usage, for instance, in the selection of vocabulary.

It is important to look at language use in context: in the workplace, socially, and in relation to country of origin. To this end, the assessment needs to be open-ended, qualitative, and exploratory. As with other areas of the assessment, the objective is to agree with the individual possible areas for remediation, and the assessor's role is to provide expertise in helping them to identify and address those areas. This then can be fed into the remediation process.

Assessment of language is a highly specialised field but adds significant value to the assessment and the remedial plan. The assessment should comprise of two broad parts: a semi-structured interview and a task. The focus of the interview is the individual's experience of language learning and language use. Here, it is important to explore their own perception of their language skills and language use. The tasks could involve performing some form of extended speaking, for example, explaining something or giving a presentation. On occasions, the issues may be to do with written language. In this case, the individual should bring samples of reports or correspondence or an essay.

During the assessment, the assessor can observe the individual's language use. The aim is to highlight evidence of fluency and 'disfluency', to explore the range and appropriateness of lexis used by the doctor or student. This assessment is also interested in phonological features (pronunciation, stress, rhythm and intonation, discourse management, grammatical range, and accuracy). In particular, the purpose of the tasks is to enable an exploration of the individual's ability to convey meaning clearly, to present complex information with suitable guidance following rules of coherence and cohesion, to guide the listener, and to use appropriate rhetorical devices.

Delivering Remediation

Whatever the methods and services employed to undertake the assessment, the next step is to design and deliver the remedial programme. The major factors for the individual in question should now be evident. This next phase will be dependent on not only resources but will be dictated by the institution or organisation. However, 'doses' of training that are generic and one off are less likely to succeed in change, compared to carefully constructed and at times individually tailored programmes.

A key challenge for the tutor, supervisor, or line manager will be to consider their role in the remedial process. Some may have to undertake a combination of roles. During the assessment phase this might be acting as the case manager. That is, initiating the dialogue, providing a framework for the remedial need and referral on for further assessment. The ability to develop a trusting and open relationship with the individual at the outset must not be underestimated if the remedial process is to be successful. Later in the process, the role might shift to provide some remediation, such as review and coaching around specific educational issues. Finally, the role might move to provide an assessment of the individual's progress or difficulties they might continue to encounter. Being open and honest with the individual at the outset to explain these roles and signpost when and how they might provide them should be given time. The next section has been written recognising that many are likely to take all three roles in some form or other during the remediation process.

Case Review

The remedial plan will be dependent on the complexity of the situation. A complex case may require bringing together the individuals who have supported the assessment activity. Within a postgraduate training institution this might be managed through a 'performance support unit'. At undergraduate level, it would be managed by the medical school. This approach will bring robustness to the review process. This is important if decisions are challenged at a later stage. It may be, for instance, that an opinion from a specialist, such as a psychiatrist, is requested by occupational health. This will add to the robust process of assessment and remediation planning and delivery. Recording of all documentation should be rigorous. Once all the documentation is collated, the next stage is to discuss the evidence with the team and decide what the remedial plan should look like. We call this activity 'case review'. The exact form that this activity takes will depend on local processes and infrastructures. In essence, the questions should be: What interventions are required? In what order should they be delivered? Who should deliver them? What is the desired outcome? Finally: What are the most appropriate time lines for review? In Box 34.5 we describe such a process with reference to the case of a doctor Dr J.

The Remedial Plan

The case review should give the tutor, supervisor, or line manager an overall picture of the major issues that need to be addressed. In relation to Dr J. the team came to the conclusion that he was quite extrovert in nature, so 'spoke before he thought'. He had strongly held beliefs about what was right and wrong in terms of the quality of medical practice and felt he was justified in airing his views. However, he seemed to lack insight into organisational issues around his team and the hospital in which he worked. Finally, underlying this there were issues around lack of self-confidence in both the workplace and being able to support his wife with their son's disability.

For Dr J., the issues that required support were around learning to reflect on his practice, developing his negotiating skills, and work on organisational issues. These three threads would aim to develop more insight into his behaviour and provide him with some skills and strategies to manage more challenging situations. Thus, the interventions used are likely to require a blend of behavioural and cognitive approaches. This in turn might help him develop more confidence in thinking about how he might manage his personal situation.

BOX 34.5 Case study

Dr J. is in his fourth year of specialist training in obstetrics and gynaecology and has been referred following his annual assessment. He was noted to have somewhat erratic and unpredictable behaviour. His referral described someone who shouted at the nurses and the junior staff when under stress and who most recently had thrown equipment around the operating theatre. On a day-to-day basis he was known for being a social and friendly character, although in his 360° appraisal he has been criticised for being overly jocular with staff.

An initial assessment was conducted. This reveals that Dr J. is married with one child, a boy aged 5 who suffers with autism. Dr J. moved schools five times as a child. He participated in team sports at medical school and was a high achiever. His father died when he was 16. When asked about his strategy for getting things done, his belief is that it is 'OK to tell it as it is'. His consultation video revealed a tendency not to listen to patients, and there was also a lack of clarity in the way he gave and explained information. During his videoed doctor–doctor consultation, where he was asked to provide feedback to a more junior trainee who had persistent performance problems, he became quite agitated when the junior doctor questioned his judgement.

He gave an indication during all three assessments that he was angry to have been referred, as he felt strongly that he was a good doctor, cared about his patients, and had no problems with his clinical skills. He felt he had to contend with a very difficult work environment that was riddled with poor practice.

Mustering Resources

A key decision at this point is to consider who will be involved in the remediation process and how that will be resourced. Those institutions responsible for medical education vary about the specific arrangements that exist to undertake remedial work. What follows is a general description of remediation activity. One important role for the tutor, supervisor, or line manager is to ensure there is a coordination of remedial activities and there is effective (both written and oral) communication between all those involved in the process. In each of the activities described below, we use the term 'remediator' to cover the individual who is working with the doctor or student, acknowledging that a combination of different remediators may be used.

Structuring the Interventions

Once the remedial activities have been agreed, the order of delivering the specific interventions requires careful consideration. The correct order will allow the individual to move forward in a manner that provides a sense of autonomy over the process. The order can also, in some cases, help the individual to develop insight into their own performance and so decreases resistance to the process which they often feel at odds with. By allowing the individual a sense of control and the ability to work initially in their comfort zone the process starts to move them from engagement into action.

Starting Out

Understanding the individual's personality and what is of 'high importance' suggests pathways for initiating the process. For Dr J., starting with a session that explores real events in a way that might use role-play and video feedback would probably feel acceptable given his preference for practical, action-based learning. Using actors who are trained in giving feedback is especially helpful in this role. Thus, setting up the actor as someone with a similar personality to Dr J. would be a useful starting point. Dr J. would probably not find this too stressful as like personalities often respond well to each other. The remediator can then coach Dr J. around 'what he did', 'how it felt', and what he 'might want to do better'. The remediator might anticipate that reflection is less comfortable for Dr J. This would be predicted by personality questionnaire feedback and reinforced by work on learning styles. The actor can respond in role about how they felt, what worked, and what did not work. This is often the first step for the doctor in gaining insight into their normal behavioural pattern. It starts the reflection into daily practice that might not come easily to someone like Dr J. Videoing the session with the actor and asking Dr J. to reflect on what he sees in himself and the actor is another useful way to help self-reflection and can be a further less daunting method for those who find such reflective exercises difficult.

Coaching could follow on from such an activity. This session could focus on some of the activities and responses discussed in the first role-play session. Starting with such a coaching session for Dr J. may have led to resistance, as this is not a natural path for him. Further sessions may include interactive role-play with a more difficult case, and sessions building on organisational issues. Leadership and organisational issues are often areas that junior doctors struggle with. Dealing with hierarchy is something that for Dr J. might not come naturally. These types of activities that can then be practised 'on the shop floor' can lead to increased self-confidence and change. Had Dr J. been a more introvert individual, starting with reflection and critical incident analysis with a coach might have been a more natural pathway.

Coaching

Coaching in this context describes an activity whereby remediation is based on a method or methods for stimulating change. It is important that the methods and approaches chosen have an evidence base. It is also important that the acquisition of practical behavioural skills takes place within a wider framework whereby the individual is encouraged to think about their beliefs and strategies. Remediation has behavioural and cognitive components. The purpose of coaching is to create a context whereby the individual can acquire new behavioural skills and cognitive strategies side by side. At this stage, active listening skills help guide the individual to discover insights into the conditions under which they might be prepared to change, and thus its emphasis is on the cognitive. Rooting the development of new skills in a context whereby the individual is also discovering a new way of interpreting what is going on around them prepares the individual for acquiring and embedding those new skills.

Role-playing Newly Learned Skills

A typical tipping point comes with the realisation by the individual that their current approach is not working or does not always work, or simply that they could become more effective. In the case of Dr J., if shouting at the nurses does not result in the correct equipment being given to him when he needs it, what else can he do? One common pathway is that a doctor such as Dr J., who is usually extroverted and expressive, becomes uncharacteristically quiet and unsure about how best to assert himself. Equally, he finds this strategy frustrating because it is not his preferred way of behaving. He feels it also achieves nothing, but at least he perceives that it keeps him out of trouble or beyond criticism.

Role-play creates a situation whereby the new skills can be practised and improved in a safe environment. Using constructive feedback from the remediator and actor, combined with video review and self-assessment, can be powerful and effective in engaging the doctor. Some doctors question the extent to which role-play in this context can really simulate the organisational context. The use of events from the individual's own experiences, particularly those significant events that contributed to the referral, can be effective in bringing validity to the work. The doctor or student can be asked to retell the story of the event or events in detail, and the sequencing of what occurred and behaviours of those involved checked and clarified along the way. The actor will often be present at this stage to allow them to construct a clear picture. This type of intervention is used to explore memories of interpersonal interactions. It provides a structure to examine not just actual communication and language skills, but also beliefs about influencing, negotiation, and leadership.

Setting Goals and Time Lines

Remediation should be set apart from counselling or other forms of therapy that the individual may require and seen as a clearly defined set of interventions that have a manageable time line. Counselling or therapy might sit alongside the remedial process, but they are not one and the same thing. Setting time lines is part of the agenda-setting process.

Consideration of how many sessions should be provided and when progress should be reviewed adds structure and helps guides both the remediator(s) and the individual. It should be recognised here that for doctors, who may have developed more ingrained behavioural patterns, the process of remediation and the complexity of the programme may be greater than that required for medical students. Review dates should be considered at the outset. There are also external constraints to consider when setting time lines, such as in-training assessments, hospital reviews, or professional regulatory hearings or exams. Sharing an understanding with the individual that the remedial process sits alongside these time lines helps engagement and reduces anxiety. Understanding a doctor's training programme, work environment, and social environment is also important. For instance, if Dr J. was a doctor in training working in a nearby hospital with a review of their progress due in four months, for example, then most of the work could be undertaken in several sessions lasting two or three hours each over a period of three to four months. This allows time for practising and embedding new skills. If, however, Dr J. was a consultant (senior physician) who has a long distance to travel, it might be more applicable to run a two-day intensive workshop for him with a review one month later. This could be accompanied by him keeping a reflective diary that could be emailed to his remediator every week and the opportunity to use Skype™ or other social media methods to keep in contact. For medical students on placements away from the central hub of teaching, social media such as Skype and FaceTime can again be invaluable. Clarity of goals and setting achievable objectives at this stage helps set out the plans and reviews that will guide the process at a later stage.

Reviewing Progress

The amount of time taken to see progress appears to vary as a function of individual differences. The proxy measures for progress in this context are as follows:

- engagement with remediation, that is, turning up for appointments
- willingness and ability to engage in reflective conversation
- insight into the nature of the problem(s) and their causes
- ability to work with the remediator to generate alternative strategies and skills
- active participation in practising new skills
- ability to discuss the use of acquired skills and strategies and how these have been applied.

Individuals will vary in terms of the amount and complexity of the remediation required, as described. The process of review will also require some flexibility and should not be prescriptive. However, bringing the remediation team together to routinely review progress is useful and again adds to the robustness of the process. The team can and should be sensitive to particular tipping points, such as reaching a judgement about:

- when enough progress has been made such that remediation can stop
- whether a plateau in progress may have been reached such that further remediation is unlikely to achieve additional improvement
- whether the individual is likely to benefit from individualised support at all.

Measuring Success

Success is hard to measure, and at present there are few validated tools that help provide hard evidence. Success might be seen as improved feedback from colleagues and the remediator(s). It may be passing exams that have been repeatedly failed. It might be achieving success at interview where the individual had failed for many months previously. It might also be an end to patient complaints. Success might also be the individual recognising their own limitations, whether due to behaviour or underlying health issues, and leaving their chosen career path. However, developing insight and changing behaviour takes time. The remedial process should give the individual space to move from a position of resistance to a place where they are ready to change. It should guide them to develop an understanding of how their behaviour has impacted on their performance and offer solutions in the form of strategies and skills.

The Role of Insight

In our experience, performance concerns are complex. It is not possible to conclude that any one personality profile is more at risk than any other. However, common patterns of behaviour do exist that appear to trigger and perpetuate episodes of underperformance. More importantly perhaps, with regard to remediation, some patterns of behaviour stand out as indicating the stage of development that an individual may be at.

We have recognised the following specific patterns.

- The individual recognises the problem behaviour as an area for development, but does not know how to develop an alternative set of skills.
- The individual recognises the problem, but focuses on justifying why that behaviour pattern is important.
- The individual does not recognise that a particular behaviour pattern is negative.

Kruger and Dunning [57] described how poor performers have less insight into their inadequacies than high performers. The more competent individuals become, the lower they self-rate their own performance. Kruger and Dunning conclude that perhaps insight in some doctors can be achieved with training.

Evaluation

As in other educational settings, such as assessment and development centres, evaluation is concerned to a large extent with the utility of the process. The reason evaluation in this context remains a challenge is that whilst multiple sources of evidence and/or methods are often used, they are typically fragmented and not always conducted in a pre-planned way in the workplace. Even where they are pre-planned, it can be difficult to bring together evidence conducted by different professionals, often over many months. Thus evaluation of remediation services is still in its infancy and needs to be further developed. However, educational institutions, workplaces, and those that provide remediation should strive to develop some system of evaluation that fits within their organisational structure.

Pitfalls and Problems

The process of remediation can be high stakes for both parties. Ensuring clarity of process, communication, and confidentiality at all stages is important. Remediation is still too often undertaken as a last resort. Those requiring remediation may often feel a strong sense of injustice. This is often focused at one or all of the following: the organisation, the educational institution, the remediators, and those individuals that have highlighted the performance concerns. One not uncommon outcome is that the individual may appeal against a decision that they perceive as negative. Such an outcome might be the need for a medical student to retake a year or be removed from training. For doctors, it might mean them engaging with formal re-training, undertaking extra training sessions, or being removed from practice. It may mean formal referral to the regulatory body. For both students and doctors, the implications are often more wide ranging than just their practice or training. It includes financial and family-related issues as well as personal self-worth and self-efficacy. Appeals can be stressful for all parties, take up considerable time, and often embed further resistance to change in the doctor or student in question. The appeal may often be based on poor (or lack of) process or transparency of process. Thus, ensuring that the initial referral is appropriate, checking legal and ethical issues relating to the process of referral and remediation, establishing boundaries of practice and roles, and maintaining good records are as important as the remediation programme itself.

Inappropriate Referrals

Initial referral for a performance concern, referral for expert advice, and onward referral to a regulatory body are all triggers that should be clearly defined and have well described criteria. Those criteria should be transparent and available for all doctors and medical students to review. Where and how they are accessed will depend on the regulatory body and training organisers involved in assessment, training, and performance review in different countries. In the UK, the GMC provides clear guidance for doctors and the Medical School Council for Medical Schools. The package of remediation offered should again have clear criteria and centre on the areas of expertise that are available locally or nationally. Setting referral criteria for onward referral is crucial to reducing the risk of inappropriate referrals.

Legal and Ethical Issues

The nature of the remediation service provided must be carefully considered prior to entering into any agreement with an individual or organisation. The ethical and legal implications of a training-led service providing support for a doctor in training will be different from a hospital offering independent services to an employed consultant or indeed a medical student engaged in medical training. Understanding where the responsibilities lie must be clear. The nature of the information that can be shared with third parties (and when), should be carefully considered, documented, and shared with the individual prior to commencing any remedial process. Overriding all of this is transparency about confidentiality and how it is maintained. Areas to review when setting up a service should include the following.

- *Written reports*:
  Who they may legally be shared with and how – this will be dependent on the country in question's own legal framework. In the UK this would fall within the Access to Medical Reports Act (1988).
- *Patient safety*:
  If there is evidence of risk to patient safety, then there must be a pathway in which serious concerns can be raised with the referring organisation.
- *Consent*:
  The individual doctor should be aware of the contract they are entering into when they attend for remediation.
- *Case notes*:
  It is important to be clear who 'owns' notes made, how they are stored, and to establish protocols around access to case files.

Blurring of Boundaries

Good communication between stakeholders at the outset is key to delivering a successful service. Managing the expectations of the educational institution, the workplace, and the individual reduces the risk of misunderstanding later. Such communication might include:

- outlining the roles of the remediation process at the outset
- specifying what the remediation can offer and its limitations
- clarity of what the content of any progress report might look like
- patient safety and duty of care of the organisation or educational institution and the role of the tutor/supervisor/line manager in maintaining this.

Record-keeping

It is advisable to keep records to the same standard as any clinical records. A good maxim is that 'if it isn't written down, it hasn't happened'.

Failure to Fail

Knowing when to stop trying in the face of no improvement is probably the most difficult decision of all [58]. This is one of the most frequently asked questions by remediators and educators. There is a sense that, at times, there is a 'failure to fail' those who do not respond to a remediation programme or show any consistent progress. Organisations and training bodies may be reluctant to put their head 'above the parapet' in deciding that the student or doctor should consider a different career pathway. Instead they may just order a further 'dose of training' or remedial programme, an act that may serve to undermine supervisors and educators in their relationships with these doctors or students. There is no one definitive right answer, but having a defined approach to assessment, a clear pathway for remediation, and the involvement of different individuals and experienced remediators, means that the decision about when to stop should be much more robust.

Conclusion

The primary purpose of remediation is to support an individual, at any level in their career, to move from ineffective to effective medical practice and/or learning. It is about building and sustaining change. To be effective and sustainable requires a holistic approach to understanding the nature of performance concerns, their triggers, and impacts. At the heart of this chapter is the evidence-based argument that, as for any complex condition, it is important for those involved in remediation to simultaneously take account of health, personal, and social factors. The individual's well-being should be at the centre of remediation. Tutors, supervisors, or line managers are central to this process. They must understand their role (or roles), their boundaries, and the process and systems they support. They must have a working knowledge of resources, how to access them, and the legal frameworks that drive remediation. Well-designed remediation can make a difference to not only the individual but, ultimately, patient care.

References

1 Cohen, D. and Rhydderch, M. (2006). Measuring a doctor's performance: personality, health and well-being. *Occupational Medicine* 56: 38–40.

2 Cohen, D., Rollnick, S., Smail, S. et al. (2005). Communication, stress and distress: evolution of an individual support programme for medical students. *Medical Education* 39 (5): 76–81.

3 Goulet, F., Jacques, A., and Gagnon, R. (2003). An innovative approach to remedial continuing medical education, 1992–2002. *Academic Medicine* 80: 553–540.

4 Joesbury, H., Mathers, N., and Lane, P. (2001). Supporting GPs whose performance gives cause for concern: the North Trent experience. *Family Practice* 18 (2): 123–130.

5 Bahrami, J. (1997). Remediation. *British Medical Journal* 315: 2.

6 Burrows, P., Khan, A., Bowden, R. et al. (2004). The fresh start simulated surgery. *Education for Primary Care* 15: 328–335.

7 Firth-Cozens, J. and King, J. (2006). The role of education and training. In: *Understanding Doctors' Performance* (ed. J. Cox, J. King, A. Hutchinson et al.), 61–77. Oxford: Radcliffe Publishing in association with National Clinical Assessment Service of the National Patient Safety Agency.

8 Paice, E. (2006). The role of education and training. In: *Understanding Doctors' Performance* (ed. J. Cox, J. King, A. Hutchinson et al.), 78–90. Oxford: Radcliffe Publishing in association with National Clinical Assessment Service of the National Patient Safety Agency.

9 Wu, J.S., Siewert, B., and Boiselle, P.M. (2010). Resident evaluation and remediation: a comprehensive approach. *Journal of Graduate Medical Education* 2 (2): 242–245.

10 Barrett, A., Galvin, R., Steinert, Y. et al. (2015). A BEME systematic review of the use of workplace-based assessment in identifying and remediating poor performance among postgraduate medical trainees. *Systematic Reviews* 4 (1): 65.

11 Hauer, K.E., Ciccone, A., Henzel, T.R. et al. (2009). Remediation of the deficiencies of physicians across the continuum from medical school to practice: a thematic review of the literature. *Academic Medicine* 84 (12): 1822–1832.

12 Dunn, L.B., Iglewicz, A., and Moutier, C. (2008). A conceptual model of medical student well-being: promoting resilience and preventing burnout. *Academic Psychiatry* 32 (1): 44–53.

13 Wallace, J.E., Lemaire, J.B., and Ghali, W.A. Physician wellness: a missing quality indicator. *The Lancet* 374: 1714–1721.

14 Henderson, M., Brooks, S.K., del Busso, L. et al. (2012). Shame! Self-stigmatisation as an obstacle to sick doctors returning to work: a qualitative study. *BMJ Open* 2: e001776.

15 Cohen, D., Winstanley, S.J., and Greene, G. (2016). Understanding doctors' attitudes towards self-disclosure of mental ill health. *Occupational Medicine* 66 (5): 383–389.

16 Department of Health (2010). *Invisible Patients, Report of the Working Group on the Health of Health Professionals*. London: Department of Health.

17 Swanwick, T. and Whiteman, J. (2013). Remediation: where does the responsibility lie? *Postgraduate Medical Journal* 89: 1–3.

18 Finucane, P.M., Barron, S.R., Davies, H.A. et al. (2002). Towards an acceptance of performance appraisal. *Medical Education* 36: 959–964.

19 General Medical Council (2014). *Good Medical Practice. Working with Doctors working for Patients*. London: General Medical Council.

20 Department of Health (2006). *Good Doctors, Safer Patients: Proposals to Strengthen the System to Assure and Improve the Performance of Doctors and to Protect the Safety of Patients*. London: Department of Health.

21 Boorman, S. (2009). *NHS Health and Well-being Review*. London: Crown Publication.

22 Murphy, K.R. and Cleveland, J.N. (1995). *Understanding Performance Appraisal: Social, Organisational and Goal-Based Perspectives*. Thousand Oaks: Sage Publications.

23 Leape, L. and Fromson, J.A. (2006). Problem doctors: is there a system level solution. *Annals of Internal Medicine* 144: 107–115.

24 Harrison, J. (2006). Illness in doctors and dentists and their fitness for work – are the cobbler's children getting their shoes at last? *Occupational Medicine* 56: 75–76.

25 Miller, W. and Rollnick, S. (2013). *Motivational Interviewing: Preparing People to Change*, 3e. New York: Guilford Press.

26 Paice, E. and Orton, V. (2004). Early signs of the trainee in difficulty. *Hospital Medicine* 65: 238–240.

27 Black, C. (2008). Working for a Healthier Tomorrow: Work and health in Britain. https://www.gov.uk/government/publications/working-for-a-healthier-tomorrow-work-and-health-in-britain (accessed 1 June 2017).

28 Kalet, A. and Chou, C.L. ed. (2014). *Remediation in Medical Education. A Mid Course Correction*. New York: Springer.

29 Waddell, G. and Burton, A.K. (2004). Concepts of Rehabilitation for the Management of Common Health Complaints. https://www.gov.uk/government/publications/concepts-of-rehabilitation-management-of-common-health-problems (accessed 1 June 2017).

30 General Medical Council (2013). *Supporting Medical Students with Mental Health Conditions*. London: General Medical Council.

31 General Medical Council (2014). *Gateways to the Professions: Advising Medical Schools: Encouraging Disabled Students*. London: General Medical Council.

32 Légaré, F., Stacey, D., Turcotte, S. et al. (2010). Interventions for improving the adoption of shared decision making by healthcare professionals. *Cochrane Database of Systematic Reviews* 12 (5): CD006732.

33 Rollnick, S., Butler, C.C., McCambridge, J. et al. (2005). Consultations about changing behaviour. *British Medical Journal* 331: 961–963.

34 Rollnick, S., Miller, W., and Butler, C. (2008). *Motivational Interviewing in Health Care*. New York: Guilford Press.
35 Lempp, H. and Seale, C. (2004). The hidden curriculum in undergraduate medical education: qualitative study of medical students' perceptions of teaching. *British Medical Journal* 329: 770–773.
36 Miles, S. and Leinster, S.J. (2007). Medical students' perceptions of their educational environment: expected versus actual perceptions. *Medical Education* 41 (3): 265–272.
37 Feeney, S., O'Brien, K., O'Keeffe, N. et al. (2016). Practise what you preach: health behaviours and stress among non-consultant hospital doctors. *Clinical Medicine* 16 (1): 12–18.
38 Cohen, D. and Rhydderch, M. (2013). Support for tomorrow's doctors: getting it right, meeting their needs. *Occupational Medicine* 6: 2–4.
39 Sensky, T. (2010). Chronic embitterment and organisational justice. *Psychotherapy and Psychosomatics* 79: 65–72.
40 Rogers, C.R. (1951). *Client Centered Therapy*. Boston: Houghton-Mifflin.
41 Harvey, S.B., Laird, B., Henderson, M., and Hotopf, M. (2009). *The Mental Health of Health Care Professionals: A Review for the Department of Health*. London: Institute of Psychiatry, King's College London.
42 Crawford, J.O., Shafir, A., Graveling, R. et al. (2009). A systematic review of the health of health practitioners: strategic consulting report. Institute of Medicine. http://www.champspublichealth.com/writedir/ee61A%20Systematic%20Review%20of%20the%20Health%20of%20Health%20Practitioners%20-%20June%202009.pdf (accessed 1 June 2017).
43 Tyssen, R. and Vaglum, P. (2002). Mental health problems among young doctors: an updated review of prospective studies. *Harvard Review of Psychiatry* 10 (3): 154–165.
44 Sharpe, M. and Wilks, D. (2002). Fatigue. ABC of psychological medicine. *British Medical Journal* 325: 480–483.
45 UK Government (2010). Equality Act 2010. www.legislation.gov.uk/ukpga/2010/15/contents (accessed 22 December 2012).
46 Firth-Cozens, J. (2006). A perspective on stress and depression. In: *Understanding Doctors' Performance* (ed. J. Cox, J. King, A. Hutchinson and P. McAvoy), 22–37. Oxford: Radcliffe Publishing in association with National Clinical Assessment Service of the National Patient Safety Agency.
47 Kierstead, J. (1998). *Personality and Job Performance: A Research Overview*. Ottawa, Canada: Public Service Commission.
48 Dudley, N., Orvis, K.A., Lebiecki, J.E. et al. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance. *Journal of Applied Psychology* 91: 40–57.
49 Ferguson, E., James, D., and Madeley, L. (2002). Learning in practice: factors associated with success in medical school. *British Medical Journal* 324: 952–957.
50 Barrick, M.R., Mount, M., and Judge, T. (2001). Personality and performance at the beginning of the new millennium. *International Journal of Selection and Assessment* 9: 9–30.
51 General Medical Council (2018). *Outcomes for Graduates*. London: General Medical Council.
52 Frank, J.R., Snell, L., and Sherbino, J. ed. (2015). *CanMEDS 2015 Physician Competency Framework*. Ottawa: Royal College of Physicians and Surgeons of Canada.
53 Association of American Medical Colleges Liaison Committee on Medical Education (2016). Functions and Structure of a Medical School: Standards for accreditation of medical education programs leading to the M.D. degree. https://med.virginia.edu/ume-curriculum/wp-content/uploads/sites/216/2016/07/2017-18_Functions-and-Structure_2016-03-24.pdf (accessed 1 June 2017).
54 Myers, I.B., McCaulley, M.H., Quenk, N.L. et al. (1998). *MBTI Manual. A Guide to the Development and Use of the Myers Briggs Type Indicator*, 3e. Menlo Park: Consulting Psychologists Press.
55 Hogan Assessment Systems (2012). Hogan Development Survey: The Dark Side of Personality. https://www.hoganassessments.com/assessment/hogan-development-survey/ (accessed 26 June 2018).
56 Lane, C. and Rollnick, S. (2007). The use of simulated patients and role-play in communication skills training: a review of the literature to August 2005. *Patient Education and Counselling* 67: 13–20.
57 Kruger, J. and Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognising one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology* 77: 1121–1134.
58 Yepes-Rios, M., Dudek, N., Duboyce, R. et al. (2016). The failure to fail underperforming trainees in health professions education: a BEME systematic review: BEME Guide No. 42. *Medical Teacher* 38 (11): 1092–1099.

Further Reading

Cox, J., King, J., Hutchinson, A. and McAvoy, P. (eds) (2006) *Understanding Doctors' Performance*. Oxford: Radcliffe Publishing in association with National Clinical Assessment Service of the National Patient Safety Agency.

Kalet, A. and Chou, C.L. ed. (2014). *Remediation in Medical Education. A Mid Course Correction*. New York: Springer.

35 Diversity in Medical Education

Nisha Dogra[1] and Olivia Carter-Pokras[2]

[1]Greenwood Institute of Child Health, Department of Neuroscience, Psychology and Behaviour, College of Life Sciences, University of Leicester, Leicester, UK

[2]Department of Epidemiology and Biostatistics, School of Public Health, University of Maryland, College Park, MD, USA

KEY MESSAGES

- Sociodemographic shifts, legislative and accreditation frameworks, social goals, and values all drive the need for diversity to be addressed by educational organisations.
- Diversity refers to differences among individuals – specifically in the aspects individuals use to define themselves and their sense of culture.
- Diversity education describes an approach to the development of current and future health professionals that is clinically relevant and based on person-centred principles such as self-awareness, empathy, respect, and non-judgmental practice.
- Diversity education has been on the medical education agenda for some time but there continues to be a lack of clarity about what should be taught and how best to evaluate it.
- Educators need to consider the merits of different educational approaches and how these help deliver the objectives of diversity education.
- Congruent behaviours, attitudes, and policies are needed within the system or organisation in which health professionals are educated and work.
- There is an urgent need for more and better scholarship and research around diversity education.

Introduction

Significant sociodemographic changes, prompted largely by the rapid rise in global migration, have stimulated greater attention to diversity [1]. Equality legislation introduced over the last 40 years signifies a societal shift towards recognising that the voices of minorities are as relevant as those belonging to majority groups. Diversity, therefore, becomes an important issue in medical education both in how institutions manage themselves and how they ensure that their graduates are appropriately prepared to deliver high quality care to diverse populations. A synergy between these two aspects – institutional and curricular – is more likely to succeed in meeting diversity objectives than a sole focus on one or the other.

In this chapter we address culture and diversity in the context of health professions education. Though much of the literature tends to hone in on race and ethnicity when discussing culture and diversity, we encourage a broader view that includes all forms of difference. After elaborating on our definition of key terms such as culture, diversity, and inclusion, we discuss the importance of and drivers for diversity education in health professions education. We then describe two approaches to diversity education, cultural expertise and cultural sensibility, and consider how these approaches differ with respect to educational philosophy, content, processes, and assessment. We review and summarise the evidence for diversity education and consider the ways in which educational institutions address diversity. We conclude by considering how stakeholders might come together to share and engage in educational scholarship around culture and diversity to improve both education and health care experiences.

Perspective and Context

Each of us brings certain values, perspectives, ideologies, and experiences to our thinking about culture and diversity. These, in turn, shape our thinking about curriculum and our approaches to teaching. Toohey [2] reminds us that there are different approaches to teaching, each influenced by a variety of factors and each taking distinctive perspectives on knowledge, learning processes, learning goals and how they are expressed, content selection and organisation, assessment purposes and processes, requisite resources, and infrastructure. We believe it is important for educators to reflect upon and be transparent about their perspectives, hence we offer such reflections in relation to diversity education in Box 35.1.

In this chapter, we do not prescribe a specific curriculum or model for diversity education as such prescriptions may

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

BOX 35.1 Who we are: authors' background and perspectives

Nisha Dogra

Throughout my adolescent and early adult life, I struggled to fit into the categories in which I was placed and this has influenced my view of diversity education and its teaching. I have strived to develop an educational model that can be applied across different contexts to all interactions and avoids categorising people on the basis of a single characteristic or assumes that we can know the experience of others without asking.

Olivia Carter-Pokras

The daughter of a bicultural union, I lived in both Latin America and the United States and spent more than a decade working on national health and data policy issues. These experiences have helped me understand that acculturation is a bidirectional and dynamic process, race/ethnicity is a social-cultural construct, and incorporating community/patient/participant input is important.

BOX 35.2 Some factors that may influence an interaction between a patient and a health professional

Every single factor that makes a person who they are may influence their interactions with others. There may be advantages or disadvantages to this. All of the factors below and possibly more influence the initial interaction. Neither party is a neutral presence in the encounter.

Age	Personal histories
Attitudes	Personality disposition
Colour	Politics
Decoration	Power
Disability	Previous experiences
Dress	Projection
Emotional state (anxiety, fear, positive)	Religion
Empathy	Space
Gender	Stereotypes each may hold
Greetings behaviour	The situation in which they meet and reason for meeting
Language	Touch
Level of knowledge about the other	Verbal behaviour
Non-verbal behaviour	

not apply across different contexts (see Chapter 5, Principles of Curriculum Design). However, evidence suggests that similar challenges in diversity education arise across different contexts [3, 4] so a broad model based on some core concepts may be generalisable. Such an approach considers educators' views on diversity and pedagogical implications rather than recommending specific instructional approaches that may not align with such views.

Definitions

Culture

There are many definitions of culture and diversity, but none are value free and all arise within a certain context. For the purposes of this chapter we use the definition of culture adopted by the Association of American Medical Colleges (AAMC) Task Force (1999):

> Culture is defined by each person in relationship to the group or groups with whom he or she identifies. An individual's cultural identity may be based on heritage as well as individual circumstances and personal choice. Cultural identity may be affected by such factors as race, ethnicity, age, language, country of origin, acculturation, sexual orientation, gender, socioeconomic status, religious/spiritual beliefs, physical abilities, occupation, *among others*. These factors may impact behaviours such as communication styles, diet preferences, health beliefs, family roles, lifestyle, rituals, and decision-making processes. All of these beliefs and practices, in turn can influence how patients and heath care professionals perceive health and illness and how they interact with one another [5, p. 25].

The key features of this definition are that:

- culture is defined by each individual person in relationship to the group or groups with whom he or she identifies
- cultural identity is not ascribed to individuals on the basis of certain characteristics
- cultural identity is based on heritage as well as individual circumstances and personal choice and may be affected by many factors, including some not mentioned in the definition above (hence 'among others' being italicised)
- cultural beliefs and practices influence how patients and health care professionals perceive health and illness and how they interact with one another.

This definition also recognises the patient as more than just a clinical presentation and as a person with many different layers. Seeing patients as people is a crucial factor in helping health providers to deliver high quality care that is personalised, coordinated, collaborative, and compassionate. Individuals construct their own sense of identity and culture based on language, country, ethnicity, religion, etc. and the AAMC Task Force definition provides a dynamic view of culture, allowing for change in how individuals view themselves in different contexts and at different life stages. The definition explicitly states that both the patient and the health professional have their own cultures and unique world views. When patients and professionals interact, the relevant parts of their separate worlds must come together to develop an outcome that suits all parties. The process of reaching that outcome is dependent on how the patient and health care professional develop their relationship and understanding of each other's perspectives. Box 35.2 highlights some of the factors that may influence an interaction.

Whilst designed to apply to the clinical context in teaching students about working with patients from diverse backgrounds, the AAMC definition of culture is equally relevant to other contexts, such as education and training, as it focuses on what individuals bring to an interaction. So 'patient' can be replaced by 'student' and 'health professional' by 'educator'; the principles remain the same. Both will bring their own perspectives to the educational process as the patient and health professional do to the health care context. This definition of culture has also proved a useful way to engage participants in diversity education, mostly because it is inclusive and explicitly acknowledges that we are all influenced by a variety of factors [6].

Diversity

We view diversity as different expressions of culture. Diversity exists because individuals have a unique sense of their own culture. However, the mere existence of diversity does not guarantee respect for or acceptance of diversity. The Global Diversity Practice (2017) definition adds respect and acceptance to the definition: 'Diversity is any dimension that can be used to differentiate groups and people from one another … Fundamentally, diversity means respect for and appreciation of differences in age, gender, ethnicity, religion, disability, sexual orientation, education, and national origin that are implemented by laws and policies' [7].

The University of Oregon offers a similar perspective on diversity: 'The concept of diversity encompasses acceptance and respect. It means understanding that each individual is unique, and recognizing our individual differences … It is the exploration of these differences in a safe, positive, and nurturing environment. It is about understanding each other and moving beyond simple tolerance to embracing and celebrating the rich dimensions of diversity contained within each individual' [8].

These definitions are positively framed and suggest that diversity itself means positive attitudes towards those who are different. Such assumptions are questionable, as irrespective of whether individuals respect or value diversity, diversity is still present. This perhaps highlights the increasing attention to diversity education and explicit discussions of inclusion.

Inclusion

While diversity focuses on differences among individuals, inclusion attends to the interactions and relationships among individuals and the extent to which people feel valued and respected for who they are and what they contribute to a group or organisation. The Global Diversity Practice (2017) defines inclusion as 'organisational efforts and practices in which different groups or individuals having different backgrounds are culturally and socially accepted and welcomed, and equally treated'. The description emphasises a 'sense of belonging' and notes that an inclusive culture often requires organisational changes in mind-set, practice, and physical space. For example, who participates in meetings and how meetings are structured may change; office space may be reconfigured; access to information and resources may be redistributed for greater equity. In inclusive cultures, people are engaged and feel valued as being essential to the success of the organisation. In short, 'diversity is the mix and inclusion is getting the mix to work well together' [7]. Both of these concepts are important to address in diversity education.

Diversity Education

There are multiple educational approaches to diversity and inclusion. Some focus on teaching about diversity through curricular content describing different cultures, belief systems, group characteristics, and various historical, political, economic, and other social structures that have differentially impacted certain groups [9–11]. Other educational approaches focus on identity, individual differences, and skills related to understanding self and others [12, 13]. Most approaches provide educational interventions designed for individual-level learning, which is the primary focus of this chapter. There are also institutional-level approaches to diversity that focus on educational systems and structures such as policies, missions, metrics, resource allocation, etc. These approaches are briefly discussed at the end of the chapter.

The social identity framework influenced much of the early thinking about diversity and educational interventions. Tajfel and Turner [14] proposed that the groups (e.g. social class, family, football team, etc.) to which people belong are an important source of pride and self-esteem. Groups give us a sense of social identity: a sense of belonging to the social world. To enhance the status of the groups with which we identify, we may overplay the positive characteristics of those groups and look negatively at those that differ from us [14]. This categorisation may create an 'us' and 'them' mind-set. However, in practice, this simple categorisation is more complex as we do not always identify with all the stereotypes of commonly stated characteristics of the groups to which we feel we have some belonging. The current world stage is ripe for debating the implications of social identity, both positive and negative. It is arguable that diversity education helps this debate and helps people realise that to treat someone who is potentially different from ourselves does not minimise us but may actually enrich us by hearing another perspective.

Of note, diversity education is not about 'political correctness', 'superficial tokenism', 'teaching stereotypical or categorical information', or 'forcing certain attitudes' and is not limited to ethnicity. Diversity education intersects with all parts of the medical education curriculum and is deeply embedded in ethics, professionalism, and clinical practice [13].

Drivers of Diversity Education

Multiple forces drive the need for diversity education in health professions education. From a sociodemographic perspective, the increasing diversity of patients and workforce arising through migration and travel has stimulated demand for diversity education as a means of serving the health needs of migrant populations [15]. The changing demographics of practice and societal expectations for patient-centred care that is sensitive to links between culture and health make diversity considerations more relevant than ever. Strong evidence linking culture and health

[16] indicates that clinicians need to incorporate patients' culture as an integral part of care.

Legislative frameworks, which vary from country to country, are also significant drivers of diversity education. In the UK, students can expect their educational institution to comply with equality and diversity legislation that specifies how their faculty and associated health care providers are expected to practise. In the US, some states have passed legislation that mandates or strongly recommends inclusion of diversity education in curricula for health professionals [3, 17]. These models are discussed more below. Accreditation standards and other policies have also influenced many leaders of educational institutions to include diversity education [3, 18]. Over the last two decades, largely out of concern for compliance, medical schools have worked to incorporate diversity into their curricula, but with little clarity about what this might entail.

Social goals and values are another key driver of diversity education. Many educators view diversity education as a way to help reduce disparities in access to care and outcomes across different groups. Increasing, albeit limited, high-quality research evidence supports this view. Studies have shown that diversity education can improve patient outcomes and contribute to safer care delivery [19, 20], improve patient experience [21], and ensure the provision of patient-centred care [22, 23]. Other studies suggest that taking a patient-centred approach improves health outcomes and that focusing on improving the quality of care for all patients helps reduce disparities [24]. Reducing disparities in access, quality of care, and health outcomes are major drivers of both diversity education and social accountability approaches. For the latter, the moral argument is clearly stated [25].

The drivers described above have shaped the nature of diversity education in ways that warrant reflection. But important though such drivers are, it is political priorities, programmatic needs assessments, and learner-generated goals that determine the parts of the curriculum that receive particular attention at any one moment in time. Some legislative drivers have fostered a narrow focus on ethnicity as the only relevant factor for diversity. Increasingly though, the literature shows that people may identify through more than one factor. Cultural sensibility models, now commonly referred to as 'intersectionality', take this multifactorial approach to identification and diversity into account [26, 27]. Similarly, drivers based on social goals and values might seem as though diversity education aims to create health professionals with a singular world view. Instead, we (and many others) see the goal of diversity education as helping health professionals and educators recognise their own perspectives and the impact these have on how they execute their roles as health and/or education professionals. This range of occasionally conflicting drivers has also led to a range of approaches.

Approaches to Diversity Education

Much of the literature in diversity education describes programmes, interventions, and models that address diversity in the context of patient care, with aims ranging from improvements in communication (e.g. agenda-setting, information gathering, negotiating treatment) [28–30] and relationships (e.g. patient experiences of care, empathy, self-awareness, knowledge/understanding of differences) between patients and health professionals [31–34] to reductions in health disparities [9, 35]. These efforts have been described in a variety of ways [3], including but not limited to: cultural competence [10, 29, 36, 37], cultural safety [38–40], cultural sensitivity [41–43], cultural sensibility [33, 34], cultural humility [44], multicultural training [32], structural competency [45], and critical consciousness [46]. There is limited evidence to suggest that any one of these efforts is 'better' than another; all have potential challenges, especially when there is a lack of coherence between the educational philosophy, content, process, and outcomes assessed.

Our review of diversity education in health professions education suggests two approaches: one based on a notion of *expertise*, the other on a notion of *sensibility* (Box 35.3). Many of the interventions and models described above align more with cultural expertise than with cultural sensibility in that they focus on ethnicity as the defining characteristic of patients [12]. In the sections below we describe these two approaches and discuss the differences between them with respect to educational philosophy, content, process, and assessment (Boxes 35.4–35.7).

> **BOX 35.3 Two approaches to diversity education**
>
> **Cultural expertise**
>
> An expert may be described as having special skill at a task or knowledge in a subject [47]. The notion that, through gaining knowledge about 'other' cultures, someone can develop cultural expertise has given rise to educational programmes that try to impart cultural competence, to create 'cultural experts' so that the health provider has expert knowledge to treat individuals belonging to groups about which they have been trained.
>
> **Cultural sensibility**
>
> Cultural sensibility should not be confused with the more common term 'cultural sensitivity'. In general usage, sensibility (openness to emotional impressions, susceptibility, and sensitiveness) [47] relates to a person's moral, emotional, or aesthetic ideas or standards. Thus, cultural sensibility is interactional: if one is open to outside experience, one might reflect and change because of that experience. This is not necessarily the case with cultural sensitivity, which focuses on the quality or degree of being aware of cultural issues. In cultural sensibility, there is no notion of acquiring expertise about others; rather, this approach recognises that we need to be aware of our own perspectives and how they affect our ability to view the perspectives of others with an open mind.

Cultural Expertise

The cultural expertise approach draws heavily on a 'cultural competence' model of diversity education, which Cross and colleagues explain as follows:

> The model called 'cultural competence' … involves systems, agencies and practitioners with the capacity to respond to

BOX 35.4 Two approaches to diversity education: Educational philosophy

Educational philosophy	Cultural expertise	Cultural sensibility
Epistemology (what constitutes knowledge)	• Knowledge exists independent of context • Positivism	• Knowledge depends on context • Constructivism
Nature of knowledge	• People are categorised into groups • Cultural competence is based on knowledge of key characteristic of these groups	• People are not categorised into groups • Cultural competence is based on knowledge of people as individuals
Use of categorisation	Categorisation is helpful	Categorisation may be unhelpful
Conception of reality	Objective reality to be revealed or discovered	No single objective reality to be discovered
Analytical perspective	Reductionist	Holistic
Historical connection	Rooted in historical context of minority disadvantage and white domination	Steps outside of the historical context of race
Politics of institutions	Improve competence of providers and/or users to improve access to care/services	Does not work on a competence level
Relation to inequalities	Attempts to change and reduce health care inequalities	Acknowledges inequalities but as such does not directly attempt to change them
Role of teacher	Teacher sets the agenda	Teacher introduces the agenda
Role of learner	Receive information	Contribute to dialogue and actively listen

BOX 35.5 Two approaches to diversity education: Educational content

Educational content	Cultural expertise	Cultural sensibility
Bernstein's curriculum type [54]	Collection type • knowledge is hierarchical such that new knowledge builds on prior and becomes increasingly complex and abstract • focus on depth	Integrated type • knowledge is laterally linked and interrelated for practical use or problem solving • focus on breadth
Nature of content	Parochial Specific	Global Non-specific
Organisation of content	To meet local needs/demands	To maximise student self-learning
Curriculum	Fact acquisition to gain body of knowledge	Self-reflection and self-awareness of students
Teaching focus	Groups (treats people as groups) More service-centred	Individuals (views individuals as potentially parts of different groups in different contexts) More patient-centred
Focus of content	Students learn about others	Students learn as much about others as themselves

the unique needs of populations whose cultures are different than that which might be called 'dominant' or 'mainstream' American. The word culture is used because it implies the integrated pattern of human behavior that includes communications, actions, customs, beliefs, values and institutions of a racial, ethnic, religious or social group. The word competence is used because it implies having the capacity to function in a particular way: the capacity to function within the context of culturally integrated patterns of human behaviour as defined by the group [36, p. 3].

Cultural competence as defined by Cross and colleagues [36] suggests a cultural expertise approach in that it focuses on learning about 'others' who are different and presents a reductionist notion of diversity. Although this definition does not emphasise working towards services that are sensitive to an individual patient's needs, it highlights the needs of groups that may or may not be as homogeneous as implied. Perhaps now dated, the concept of cultural competence provided an important foundation for current

BOX 35.6 Two approaches to diversity education: Educational processes

Educational processes	Cultural expertise	Cultural sensibility
Learning process	Acquisition of knowledge	Acquisition of principles (method)
Desired learning outcomes	Command of body of information and facts	Command of mode of respectful questioning
Learning goals expressed as	Skill and competence	Attitudes and self-reflection
Framing of content	Certain Dichotomous Right or wrong	Variable, uncertain Continuum or multifaceted Mostly grey areas
Cultural focus	Majority view of other cultures dominant Majority whites need to consider needs of minorities	No focus on particular groups – all individuals need to consider needs of others
Pedagogical approach	Didactic Teacher-directed	Directed self-learning Teacher and learner co-construct meaning
Role of experts	Experts understand cultural perspectives of certain groups	No one individual has ownership of expertise of others with respect to identification of cultural belonging

BOX 35.7 Two approaches to diversity education: Assessment

	Cultural expertise	Cultural sensibility
What purpose does the assessment serve?	Demonstrates knowledge of other cultures	Demonstrates some understanding of self and ability to assess own learning
What assessment methods are used?	Paper and pencil tests ranging from multiple-choice questions and short answers to long essays Checklists	Reflective journals, project work (usually experientially based) Self-assessments
Who are the assessors?	Teachers	Students
What outcomes are most valued?	Facts about other cultures	Methods of inquiry that demonstrate awareness and openness to diverse perspectives Ability to be critical and self-reflective Capacity for dialogue
How transferable are learning outcomes?	Content only pertains to *cultural issues*	Content can apply to any context in which there are differences between the doctor and patient be they cultural, gender, level of education
How is programme success defined?	Students gain competence in other cultures; bonus if students learn about themselves	Students learn about themselves and can appreciate other perspectives

developments in diversity education [48, 49]. Many medical educational programmes use the terminology to describe ways of educating trainees to work effectively with those who are different from them. More recent thinking incorporates the concept of structural competence, which focuses on 'systemic, institutional determinants of health inequalities' [45], and offers critical consciousness as a way of understanding how social contexts and power structures influence individuals, interpersonal interactions, and communities [46, 48].

In the expertise approach, culture is perceived as an external characteristic, something that others can see in what people do and how they behave. Race and ethnicity are often emphasised, in comparison with other aspects such as gender and social class. Differences between individuals are generalised and relationships in society are perceived to differ among groups. 'Cultural immersion' programmes exemplify this perspective. These programmes imply that learning about one ethnic family in depth provides a generalisable experience when encountering others from the same ethnic group [50, 51]. From the cultural expertise perspective, dialogue regarding culture takes place at a group level and the individual's identity is fixed irrespective of the context. For example, an Indian woman is an Indian first and foremost, irrespective of the context. Yet, if she is seeing a gynaecologist she might identify her sex as being the more pertinent issue.

The cultural expertise approach also reduces culture to specific traits and simplifies people's culture into items that can be observed. This approach does not recognise the meanings that individuals may assign to their sense of self and culture and instead assumes outsiders can learn to look for characteristic behaviours (often in the form of lists) to identify a person's culture. For example, Culhane-Pera and colleagues [52] found that doctors-in-training wanted concrete pieces of information from which they could generate 'do and don't' lists for use in clinical practice. The authors encouraged teachers to resist providing such lists. Gudykunst [53] hypothesised that people with a low tolerance for ambiguity try to gather information that supports their own beliefs, while those with a high tolerance for ambiguity seek more 'objective' information about others. Cultural expertise teaching may reinforce stereotypes, as health professionals try to resolve ambiguity by interpreting messages in ways that suit them rather than being open to what others say or mean due to discomfort with ambiguity.

Cultural Sensibility

The cultural sensibility approach has a different perspective on culture. It acknowledges that unless we ask others, we cannot know about them. The AAMC definition of culture presented at the beginning of this chapter aligns with the cultural sensibility approach [5]. The drive behind this approach is not so much that health care inequalities can be addressed by cultural sensibility, but that an understanding of how individuals see and understand themselves may help practitioners improve individuals' access to health care. This approach recognises that a range of factors, such as poverty and age, and their interplay may more significantly affect health than culture alone [54].

In contrast to Cross and colleagues [36], Kim [37] described a model aligned with cultural sensibility that emphasises the dynamic, interactive nature of communication processes between two or more individuals. The relationship between the individual communication system and the multi-person communication system is multidirectional and multilateral in causality. All parties involved in a given encounter, including the conditions of the social context in which the encounter takes place, co-determine the communication outcome. This means that no single element in a multi-person communication system can be separated out for being solely responsible for the outcomes. Each person has a reality of his/her own perspective and gives different meanings that make sense to them. An example of this might be the interaction between a medical student and his/her tutor. Both parties bring many factors into the interaction and their understanding of what was agreed upon may differ depending on how those factors have influenced the interaction.

This systems-based approach resulted from unease with the processes and outcomes of the cultural expertise model and was important to the development of cultural sensibility. Kirmayer [55] argued that cultural competence needed to be rethought as it was developed in a particular context (the US) with a specific way of looking at issues. He also raised concerns that cultural competence reifies culture by ascribing fixed characteristics to an individual without taking into account other relevant factors such as personal history and context.

The cultural sensibility approach [12] makes the terminology, educational philosophy, educational processes, content, and outcomes explicit. The key points of this approach are discussed in the next sections and summarised in Boxes 35.4–35.7. The principles of the cultural sensibility approach have been applied to teaching medical students, health care professionals, mentors, student support staff, and educators. Some of these interventions have been evaluated (see, for example [36]) and as the approach has been refined, diversity has been integrated with other curricular topics such as professionalism and communication skills.

Educational Philosophy (Box 35.4)

The educational philosophy of the cultural expertise approach is based on the epistemological position that knowledge exists independent of a context. From this perspective, culture can be categorised in the same way that medical disorders are categorised into underlying lesions that are indicated by signs and symptoms. A constellation of particular signs and symptoms lead to the diagnosis of a disorder that may or may not manifest in practice. The cultural expertise approach, in the extreme, treats culture in the same way in that particular signs of how people behave (e.g. the food they eat) or particular characteristics (e.g. skin colour) or beliefs (e.g. views about alternative medicine) are used to categorise people into cultural groups [30, 56].

In the cultural sensibility approach, educational philosophy is rooted in a wider social context and is located within a social constructionist perspective. Knowledge is seen as contextual and does not necessarily need to be categorised. Cultural sensibility does not use the medical approach as a metaphor and does not attempt to look for signs and symptoms, which can lead to a classification of an ethnic group or other social category. Cultural sensibility recognises that different people interpret the world differently, such that two individuals in the same group who experience the same event may take very different meanings from it. The philosophy behind this approach is that there is no single objective reality to be discovered. Individuals construct their own version of their culture dependent on the various social discourses of which they are aware or in which they participate. The different underlying philosophies of the cultural expertise and cultural sensibility approaches result in differing educational processes, contents, and assessments.

Content (Box 35.5)

The aim of the cultural expertise approach is to create 'experts' by giving them knowledge about specific groups. The focus is not necessarily on the learners challenging the knowledge or its presentation. The curricular content tends to address knowledge, skills/abilities, and attitudinal learning objectives.

Knowledge: describe a person's history and culture of country of origin; identify pertinent psychosocial stresses, family life, and intergenerational issues; know the difference between culturally acceptable behaviours and where the behaviours indicate potential illness; recognise the role of religion; be familiar with cultural beliefs about causes and treatments of disease; and explain differences in disease prevalence and response to medicine and other treatments.

Skills/Abilities: interview and assess patients in the target language (or via interpreter); communicate with sensitivity to cross-cultural issues; avoid under/over diagnosing disease states; understand the patient's perspective; formulate culturally sensitive treatment plans; effectively utilise community resources; and act as a role model and advocate for bilingual/bicultural staff and patients.

Attitudes: acknowledge the degree of difference between patient and physician; demonstrate empathy by recalling the patient's history of suffering; have patience in shifting away from the Western view of time and immediacy; respect the importance of culture as a determinant of health, the existence of other world views regarding health and illness, the adaptability and survival skills of patients, the influence of religious beliefs on health, and the role of bilingual/bicultural staff; and demonstrate humour by having the ability to laugh with oneself and others (Lee, as cited by AMA 1999: Section X : 17) [57].

The approach here focuses on differences and there is limited reflection and self-awareness as culture is largely externalised.

The cultural sensibility approach is built on a transformative learning approach. In this approach, curricular content and learning objectives tend to focus on a general approach that can be used in many contexts, rather than on knowledge of, skills/abilities for, and attitudes towards particular groups of people based on culture, ethnicity, gender, etc. Some examples of learning objectives consistent with the cultural sensibility approach are given here.

Knowledge: recognise broad psychosocial issues that can affect the way individuals perceive health and access health services; know the contexts that information is presented or received in.

Skills/Abilities: demonstrate a method for acknowledging difference; work with differences in a constructive and positive way; acknowledge that difference between doctor and patient is potentially present in all encounters, not just those where ethnicity differs.

Attitudes: engage in self-reflection and self-awareness; attend to interactions with others with awareness that dialogue has the potential to change either, both, or neither of the participants; strive to achieve shared understanding through interaction and dialogue.

The cultural sensibility approach cannot work if health care professionals do not recognise the need to be aware of their own biases and triggers and the impact these potentially have on the care they deliver. Health care professionals must be consciously aware of bias not only because it may lead to suboptimal care of individuals subject to prejudice, but also because there may be overcompensation from a sense of guilt. Students cannot predict what patients will bring into the room, so they are encouraged to have a good understanding of themselves and how their perspectives may influence their interactions with patients.

Cultural sensibility acknowledges that student understanding of other cultures is linked to their understanding of wider sociological debates, and their own meaning of culture and cultural belonging. Before students can, most effectively, make sense of the struggles that people have experienced in assigning certain meanings to their lives, students need to be aware of the meanings they assign to their own lives [58]. It is this awareness of self (or the lack thereof) that students take to a clinical consultation (and indeed to any work context). Whether or not they are aware of it, what they take and what the patient brings to the consultation are influenced by each other: the cultural sensibility approach emphasises that if a doctor has had no exposure to difference, when he/she comes across it in practice, he/she may feel uncomfortable. In turn, the patient may pick this up and interpret it in many ways, some of which may be negative and lead to a less effective dialogue between the two. If the doctor is aware of his/her own discomfort, he/she can be more attuned to the possibility that the patient's response may be equally related to the doctor's behaviour and to the patient's 'culture'. For example, if for whatever reason, the doctor is uncomfortable about the issue of domestic violence, the doctor may either not raise the issue, or raise it in such a way that tells the patient the subject is taboo. The patient may leave with their story unheard, whatever their 'ethnic' background.

Educational Processes (Box 35.6)

The two approaches differ in educational processes (e.g. pedagogical techniques and instructional methods designed to achieve certain outcomes). Since the cultural expertise approach is knowledge driven, instructional methods aim to support learners' acquisition of facts; for example, learners should know what the views of Native American Indians are towards homeopathic remedies and, more generally, have facts at their fingertips relating to different ethnic or cultural groups (terms often used interchangeably, though they are different). Learning outcomes are expressed in terms of skill and competence [52]. The cultural expertise approach tends to be dichotomous and implies that there is a wrong or right way to address issues. There is also a view that individuals might be experts about groups: i.e. individuals carry a body of knowledge that qualifies them to profess expertise in a particular culture [56]. The certainty of knowledge presented in the cultural expertise approach aims to give students a sense of comfort and confidence through the knowledge they learn. As such, the cultural expertise approach does not challenge students to consider or question the validity of this knowledge.

In contrast, cultural sensibility is attitudes driven. In this approach, educators ask students to evaluate their attitudes and consider the effect these might have on the care they deliver. The educational processes focus on learning principles, concepts, and skills applicable across a range of issues. The process is recursive and reflective. Emphasis is on a transformative approach to learning, similar to critical consciousness discussed by Halman and colleagues [46]. The learning enables the student to reason about applications beyond the scope of the immediate question, theorise about related issues, or reflect on his or her own actions and understanding.

The cultural sensibility approach is process oriented and, at best, involves dialogue between learners and teachers. Teachers are seen as facilitators in the student learning process, consistent with Vygotsky's zone of proximal development concept [59]. Expertise in the cultural sensibility approach is about understanding the influence of culture and the meaning of culture to each individual rather than specific, codifiable knowledge of groups. Cultural sensibility acknowledges uncertainty, tries to ensure learners become comfortable with not knowing, and prompts the realisation that there are few wrongs or rights. Students are expected to feel uncomfortable about some of the issues raised and to find the learning personally challenging. The approach seeks to support students through discomfort and challenges rather than avoiding difficult issues. It emphasises that culture is complex and that there are many grey areas. This approach enables students to move towards directed self-learning and tends to focus on the individual and the meanings the individual has about their sense of culture and cultural belonging.

Assessment (Box 35.7)

The cultural expertise approach assesses knowledge of other cultures typically through tests comprised of multiple-choice questions, short-answer questions, essays, or marks on a checklist. In a paper published in 2007, Kumas-Tan and colleagues [60] reviewed 10 widely used quantitative measures of cultural competence in health professions education and identified problematic assumptions associated with these measures. Most conceptualised culture and diversity in terms of race and/or ethnicity and attributed cultural incompetence to lack of familiarity with and/or discriminatory attitudes towards 'others' whose world views differ from those of the dominant group (typically white, Western). The authors emphasised the need for assessments of cultural competence that move beyond attribute-based knowledge of 'the other' and instead incorporate constructs such as ethical sensitivity, cultural humility, critical thinking, and changes in actual practice. The authors also recommended use of qualitative and mixed-methods of assessment [60]. In another paper published in the same year, Gozu and colleagues [61] reviewed all studies evaluating cultural competence curricula by using at least one self-administered measure of knowledge, skills, or attitudes associated with cultural competence. Among the 45 instruments included in their review, only six instruments had any published evidence of validity and reliability and most involved self-assessment of knowledge, skills, attitudes, or behaviours. Based on item-level review of 23 instruments, the authors raised many concerns similar to those of Kumas-Tan and colleagues [60], as well as additional concerns about social desirability bias in responses, potential reinforcement of stereotypes, questionable accuracy of self-assessment, and tenuous inferences about cultural competence based on reported behaviours [61].

Cultural sensibility encourages an openness to new possibilities by focusing on a willingness to accept not always knowing and developing a capacity to engage in dialogue with others and withholding a judgement. Correspondingly, assessments focus on students' demonstration of reflective skills, often through project work. The learning, as an ongoing process, is expected to continue beyond the lifespan of the teaching module or curriculum. Curcio and colleagues [62] developed an assessment instrument that is consistent with the cultural sensibility model in that it does not make assumptions about the individual completing the instrument and acknowledges that all participants have a culture. The instrument was designed to identify the need for diversity education and to inform curricular decisions. It can also be used to facilitate self-awareness and achievement of learning outcomes associated with cultural sensibility (e.g. ability to understand cultural forces that affect physicians and patients and to identify the interaction between these forces) [62].

Although it is generally agreed that assessment of diversity needs to be multifaceted [63], the complex nature of such a process often means it is tokenistic. Students may identify the 'diversity' OSCE station and thereby diminish its effectiveness. In an effort to assess diversity in ways that better align with diversity in clinical practice, some medical schools have shifted attention to diversity as part of patient-centred care. For example, two UK universities are piloting assessments that focus on how the student responds to a variety of specific patient needs. This moves away from diversity assessment based on students' ability to impart information in a set way based on knowledge of a particular cultural or ethic group and towards diversity assessments based on how students respond to what a patient says or does and how well students integrate all aspects of care.

Evaluation of Diversity Education: Does Diversity Education Work?

In 2003, Betancourt and colleagues [9] published one of the seminal reviews of the literature on efforts to address racial and ethnic disparities in health and health care. The review considered a wide range of interventions to address sociocultural barriers to care and proposed a framework to describe the broad cluster of cultural competence interventions based on the level of focus: institutional or organisational, structural, or clinical. The clinical cultural competence interventions category included education and training designed to 'equip health care providers with knowledge, tools, and skills to better understand and manage sociocultural issues in the clinical encounter' (p. 298), which is similar to the cultural expertise approach described above. The authors also acknowledged 'a newer approach', much like the cultural sensibility approach, that 'focuses on the process of communication and trains providers to be aware of certain cross-cutting cultural and social issues and health beliefs that are present in all cultures.

The focus is on the individual patient as teacher and on developing important attitudes and skills for providers' [9, p. 299]. The authors concluded that, 'some balance of cross-cultural knowledge and communication skills seems to be the best approach to cultural competence training', despite finding few studies evaluating any educational interventions [9, p. 299].

Since publication of this important framing work by Betancourt and colleagues [9], several additional reviews have been conducted [22, 23, 64–72] (see Box 35.8). Overall, these reviews suggest diversity education can enhance knowledge or awareness of cultural differences and improve skills and attitudes towards diversity [22, 64–67, 69, 71]. However, consistent with a 2005 review of the methodological rigour of 64 studies evaluating cultural competency training [73], current evidence is limited by lack of conceptual clarity around cultural competence, weak study designs, and few robust measures of outcomes

BOX 35.8 WHERE'S THE EVIDENCE: Evaluation of diversity education

This summary of key findings is based on 11 articles that review literature on the effectiveness of cultural competence training [22, 23, 64–72]. As noted in many of these articles, cultural competence is a broad term that includes many concepts such as cultural sensitivity, multiculturalism, cultural humility, cultural awareness, critical consciousness, anti-racist pedagogy, and implicit bias reduction. Specific outcomes and methods used to evaluate effectiveness vary considerably among studies, but can be clustered into the general outcome domains listed below.

General outcome domains used in evaluation of cultural competence training	Example indicators/measures	Review articles
Impact on health professionals		
Increase in cultural competence (general, multi-faceted)	Objective Structured Clinical Examination (OSCE) with performance-based scoring rubric Non-verbal communication Inventory to Assess the Process of Cultural Competence among Health care Professionals (IAPCC-R) [74]	[66, 67]
Increase in knowledge or awareness of cultural differences, disparities	Cultural Knowledge Scale [75] Written exams testing: • knowledge of general cultural concepts (e.g. impact of culture on patient-provider encounter) • culturally-specific knowledge (e.g. knowledge of disease prevalence in various populations) • understanding of cultural humility	[22, 65, 69, 71]
Improvement in skills/behaviours	OSCE or assessment of performance with real or simulated patient on skills such as: • patient-professional communication • patient-centred care • self-reported practices (e.g. use of the Listen, Elicit, Assess, Recommend, Negotiate (LEARN) framework [76])	[22, 64–66, 69, 71]
Change in attitudes	The Cultural Self-Efficacy Scale (CSES) [77] Interest in learning about patient and family backgrounds Confidence delivering culturally competent care	[22, 64, 65, 69]
Increase in self-awareness (including awareness of implicit biases)	Written reflections Self-Assessment of Cultural Awareness questionnaire [41] Implicit Association Test (IAT) with feedback [78]	[67, 72]
Reduction in implicit bias	Race Implicit Association Test (IAT) [79]	[72] Mixed evidence: [70]
Impact on patients		
Increased patient satisfaction	Patient-Reported Physician Cultural Competency (PRPCC) [80] Patient family satisfaction questionnaires	[67, 71] Limited evidence: [23, 63]
Increased equity in services across groups	Patient involvement in care	Limited evidence: [68]
Reduction in health disparities and other health outcomes	Patient adherence	Limited evidence: [23] (trend suggesting positive impact: [22])

[69]. Evidence of impact on clinical practice is even sparser, for similar reasons [23, 69, 70]. Box 35.8 shows some of the key systematic reviews of diversity-related educational interventions for health professionals.

The lack of conceptual clarity is a problem when attempting to evaluate the evidence for diversity education. In an effort to clarify the domains of cultural competence, the Association of American Medical Colleges (AAMC) developed the Tool for Assessing Cultural Competence Training (TACCT) [81]. This tool was 'designed for medical school leaders to examine all components of their cultural competence curricula, identify gaps and redundancies, and make the best use of opportunities and resources' [69]. Jernigan and colleagues used the tool to describe and evaluate cultural competence training in US medical schools [69]. Among the 18 schools included in the study, the authors found substantial variation in the nature and quality of the training as well as in the outcomes assessed. The authors concluded: 'The appropriateness of cultural competence training as a strategy to eliminate racial and ethnic disparities in health care remains poorly understood' [69].

There is clear room for improvement in evaluation of diversity education. In addition to concerns raised about lack of conceptual clarity, poor study design and absence of standardised outcome measures, we note that most studies of diversity education to date focus on cultural competence training, which fits with cultural expertise approaches. Few studies have evaluated cultural sensibility approaches. One explanation for this may be the preference in reviews and by funding agencies for positivist or experimental modes of evaluation, which may be less suitable for the goals of diversity education. Qualitative and narrative accounts from relevant stakeholders of the impact of diversity education must be recognised as valuable sources of evaluative evidence. As demonstrated in studies by Neff and colleagues [82] and by Nazar and colleagues [83], qualitative data yields insights about learners' experiences of diversity education that reveal both valuable shifts in relationships and perspectives that are difficult to quantify as well as areas in need of improvement, such as support when learners feel overwhelmed by awareness of structural barriers to patients' health. The preference for measurable outcomes may also reinforce use of instruments and data that equate culture with ethnicity and assume that 'culture' is possessed only by those belonging to minority groups. Evaluation of educational interventions that fit with a cultural sensibility approach must take a much broader, person-specific view of culture. Dogra and Vostanis [84] found that the majority of the 17 staff interviewed had a traditional positivist view of culture, with a greater emphasis on ethnicity over other factors. Many staff felt that previous training reinforced or created stereotypes and yet identified a need for more information about specific groups.

Developing Research and Scholarship

There is a need to develop a community of practice [85] for diversity education that can both support teachers as well as provide opportunities for reflection and critique of developments in diversity education (for more on communities of practice, see Chapters 4 and 12 of this book). Communities of practice can enable both research and scholarship. Such communities of practice need to actively engage all stakeholders in the educational and health care process to ensure that the approach used can be integrated into all aspects of the organisation's functioning. Through exploration of how different stakeholders understand diversity, an organisation can develop a more coherent approach. The findings from this exploration can provide a rationale and evidence for being inclusive, which can then be used to help create an institutional culture that respects difference but is also prepared to challenge difference when it contradicts legal and social expectations of the context (for example not segregating students to comply with religious practice when it is unlawful; not allowing students to 'miss' parts of the curriculum they feel are at odds with their own beliefs).

King and colleagues [86] argued that diversity training and education have developed with limited connection to scholarship, which has resulted in slow development with narrow focus. The reviews discussed above have noted the paucity of high-quality research [23]. Unless we are clearer about the educational approaches used and identify appropriate assessment methods, it is difficult to know whether the lack of success lies with the educational philosophy, process, or expected outcomes or within educational organisations themselves. Rigorous design and evaluation of diversity education programmes are therefore needed. Over the past few years, scholars have convened health professional educators, experts in cultural diversity education, and other stakeholders to discuss opportunities to strengthen cultural diversity education. Specific research questions include [87, p. 19]:

- How can we link training to improved health outcomes using a universal cross-cutting approach to improve culturally competent care delivery that also addresses limited health literacy? [88]
- Can communication training focused on understanding underserved populations enhance patient comprehension and change behaviour?
- Do programmes that aim to change attitudes have different impact than programmes that focus primarily on behaviour change?
- How can we effectively include community stakeholders for health professional training and what is the added value of using patients as teaching partners?

There is also a need to develop scholarship regarding:

- Effective faculty development that helps educators develop curricula to address diversity, model culturally sensitive approaches, and attend to undesirable aspects of the hidden curriculum.
- Sustainability of attitudinal changes across educational contexts and into practice.
- Relationship between attitudinal changes and patient outcomes.

However, we need to ensure future research is more rigorous and of a higher quality than we have managed to date. Cross-institutional research with multidisciplinary staff may help us better answer these questions.

How Institutions Address Diversity

Many institutions may assume there is a 'correct' position on diversity and a single correct way to approach diversity education. This view may lead to the perception that diversity education is prescriptive and not respectful of different perspectives. Consequently, learners may not be encouraged to reflect and challenge their own perspectives, thereby undermining principles taught in the cultural sensibility approach. Additionally, health professionals and educators (including patient educators) may not model the reflection expected of students [89], thus further contributing to a 'hidden curriculum' [90]. These circumstances support the view that institutions need to consider how they address diversity as a whole [91]. For example, Smith recommends 'a systems approach' to diversity and offers a framework for building institutional capacity for diversity and inclusion in academic medicine. Her work identifies key elements such as: a mission that 'considers diversity as core to excellence'; institution-wide understanding of diversity as 'both inclusive and differentiated' so that a wide range of perspectives are included (e.g. race, class, disability, religion, sexual orientation, etc.); alignment of intention and actions throughout key parts of the institution; selection and monitoring of key metrics associated with progress; and development and placement of diverse leadership across all levels of the institution [91].

Here we consider some strategies that may be adopted within institutions alongside some of the external influences to which they must respond.

Internal Strategies

Alignment of Intention and Action

Managing diversity in clinical contexts requires more than diversity education of individual health professionals. Congruent attitudes, behaviours, and policies are also needed within the system or agency in which health professionals are educated and work [36, 92]. Institutions need to effectively model what is expected from their staff and learners. Students' experiences outside the classroom also impact their knowledge and understanding of culture and diversity. In one study of seven medical schools, medical students in the clinical phase of training reported having received more diversity content than reported by their course or clinical clerkship directors [93]. Students' experiences with diversity education content in the 'informal or hidden curriculum' can help explain these findings.

The 'hidden' curriculum during clinical years can have powerful effects on student attitudes. Attitudes, practices, and reactions of staff role models and peers can be good or bad (e.g. offhand comments behind people's back, preferential treatment of certain patient types). This hidden curriculum can 'reinforce or undervalue those aspects of practice we would like students to acquire' [94, p. 802]. For example, a qualitative study of participant observations and semi-structured interviews in two internal medicine wards in a teaching hospital in the Netherlands found multiple examples of medical students observing 'residents, nurses, and attending physicians using disparaging names or remarks, jokes, and generalizations' regarding elderly patients [95, p. 37]. These observations can have a very negative impact on subsequent attitudes and behaviours of trainees. In a qualitative study of three Canadian medical schools, medical students shared examples of role-modelling at odds with institutional values of equality (e.g. words, attitudes or behaviours reflecting the view that personal characteristics such as obesity or poverty are signs of weakness or moral failure) [96]. The authors noted that when faced with discordant role modelling, students reacted by 'challenging, dissociating themselves, with silence, or with confusion followed by attempts at individual transformation to realign careers and behaviours with those of teachers' [96, p. 887]. Cultural sensibility makes these challenges more explicit so that students are better prepared to recognise and address them.

Selecting and Monitoring Metrics

Recognising exemplary professional behaviour and addressing unprofessional behaviour that learners encounter in clinical settings is a critically important part of diversity education, since the most effective techniques for developing professionalism are role modelling and mentoring [97]. Facilitated conversations and writing exercises (reflective learning) have also been shown to contribute to improvements in professionalism and practice [98]. One proactive approach to address unprofessional behaviours by specific staff members in clinical settings is the use of a web-based reporting system for any member of the health care team to report a staff member's unprofessional behaviour [99]. The severity and pattern of the reports is then assessed by a council of peers, and action is taken if the reports reflect a single egregious act or repeated unprofessional behaviour. Another proactive approach used by several schools is to ask students to submit professional and unprofessional stories online and then reflect on these clerkship experiences with 'a faculty member in a confidential and non-threatening environment' [100].

Faculty Development

For diversity education to be coherent, educators in the health professions need to model the principles discussed above. Faculty cannot integrate diversity into the curriculum if the topic is viewed as only relevant to individuals belonging to minority groups. Faculty cannot expect students to adhere to principles that only apply in certain contexts; they also cannot expect students to have potentially difficult conversations with patients if faculty avoid difficult conversations with students. Acosta and Ackerman-Barger [101] argue for faculty to be appropriately trained on 'how to conduct interracial dialogues on race, racism, oppression, and the invisibility of privilege' (p. 285).

Faculty need to be trained and supported as they can be as anxious as students about managing diversity appropriately. Similarly, the training needs of patient educators and staff must be considered. If we see patients as whole people we can then translate this approach to our peers and students – none of us are unidimensional and the strengths of diversity may be uncovered when we are prepared to be curious and willing to address our own potential fears and doubts. More and more faculty development for diversity is occurring, but it is often not formally evaluated and may reflect the fragmentation and lack of coherence seen within health professions education generally [102].

The refined model of cultural sensibility outlined by Karnik and Dogra [34] was originally developed for medical undergraduate education and has been applied for other interactions (peer relationships, student support, trainers). The model could also be used for faculty development, so there is a coherent model for diversity throughout the institution. Having an institution-wide model for diversity and diversity education helps ensure that what applies to the curriculum applies to student support and other interactions across the board. Institutions can then begin to address some of the issues raised above, such as contradictions between the formal (explicitly taught) and hidden curriculum and be more transparent about the value of diversity in the institution. Students can also carry this forward once they are practising and teaching beyond medical school.

BOX 35.9 FOCUS ON: National Culturally and Linguistically Appropriate Services (CLAS) Standards in Health and Health Care [17]

Published by the US Department of Health and Human Services, the National CLAS Standards are 'intended to advance health equity, improve quality, and help eliminate health care disparities by establishing a blueprint for health and health care organizations to':

Principal Standard

1 Provide effective, equitable, understandable, and respectful quality care and services that are responsive to diverse cultural health beliefs and practices, preferred languages, health literacy, and other communication needs.

Governance, Leadership, and Workforce

2 Advance and sustain organizational governance and leadership that promotes CLAS and health equity through policy, practices, and allocated resources.

3 Recruit, promote, and support a culturally and linguistically diverse governance, leadership, and workforce that are responsive to the population in the service area.

4 Educate and train governance, leadership, and workforce in culturally and linguistically appropriate policies and practices on an ongoing basis.

Communication and Language Assistance

5 Offer language assistance to individuals who have limited English proficiency and/or other communication needs, at no cost to them, to facilitate timely access to all health care and services.

6 Inform all individuals of the availability of language assistance services clearly and in their preferred language, verbally and in writing.

7 Ensure the competence of individuals providing language assistance, recognising that the use of untrained individuals and/or minors as interpreters should be avoided.

8 Provide easy-to-understand print and multimedia materials and signage in the languages commonly used by the populations in the service area.

Engagement, Continuous Improvement, and Accountability

9 Establish culturally and linguistically appropriate goals, policies, and management accountability, and infuse them throughout the organization's planning and operations.

10 Conduct ongoing assessments of the organization's CLAS-related activities and integrate CLAS-related measures into measurement and continuous quality improvement activities.

11 Collect and maintain accurate and reliable demographic data to monitor and evaluate the impact of CLAS on health equity and outcomes and to inform service delivery.

12 Conduct regular assessments of community health assets and needs and use the results to plan and implement services that respond to the cultural and linguistic diversity of populations in the service area.

13 Partner with the community to design, implement, and evaluate policies, practices, and services to ensure cultural and linguistic appropriateness.

14 Create conflict and grievance resolution processes that are culturally and linguistically appropriate to identify, prevent, and resolve conflicts or complaints.

15 Communicate the organization's progress in implementing and sustaining CLAS to all stakeholders, constituents, and the general public.

These standards can be found at https://www.thinkculturalhealth.hhs.gov/clas/standards

External Influences

In addition to these strategies within institutions, there are external influences to which institutions must respond. Differences in legislation, accreditation standards, and other policies mean that institutions have different levels of responsibility and obligations to ensure a non-discriminatory and safe environment for students, faculty, and staff. Various experts as well as governmental and accreditation agencies have identified cultural diversity education as essential for eliminating disparities related to patients' health status and access to health care [3, 19]. For example, in 2000, the accrediting body for US and Canadian medical schools required that faculty and students 'must demonstrate an understanding of the manner in which people of diverse cultures and belief systems perceive health and illness and respond to various symptoms, diseases, and treatments' and 'recognize and appropriately address gender and cultural biases in themselves … others, and in the … health care delivery' [103, p. 10]. Ten years later, the Affordable Care Act called for investment in diversity education research and curriculum in the US [104].

Several states in the US have passed legislation that mandates or recommends diversity education and training of physicians and other health professionals. Unfortunately, legislative and other administrative policies to mandate or recommend diversity education usually have not been accompanied with funding. For example, no funding was provided by the US Congress to support the diversity education research and curriculum provisions of the Affordable Care Act [104]. In the UK, educational institutions have statutory responsibilities to ensure they comply with the Equality Act 2010 [105]. However, the focus can be on equality rather than diversity [13].

In order to provide a more supportive environment for diversity education and related efforts, the US Task Force on Community Preventive Services called for programmes to recruit and retain staff who reflect the cultural diversity of the community. Use of interpreter services or bilingual providers, use of linguistically and culturally appropriate health education materials, and culturally specific health care settings (e.g. signage in different languages, location near community needs) are also needed [63]. Fifteen specific action steps have been identified for individuals and health care organisations that address governance, leadership, and workforce; communication and language assistance; and engagement, continuous improvement, and accountability [17]. These revised National Culturally and Linguistically Appropriate Services (CLAS) Standards in Health and Health Care (Box 35.9) now acknowledge the importance of health literacy in providing CLAS [17]. Low health literacy, or the lack of skills to obtain, process, and understand health information needed to make informed health decisions [106] is associated with poor health outcomes and increased costs [107]. A competent health care professional must not only respect consumers' and patients' culture and beliefs, but also their ability to understand health information and put it to use in their daily lives (health literacy). Health care interventions at multiple levels are therefore needed.

Conclusion

In this chapter, we defined culture and diversity and compared educational models designed to deliver diversity education in a coherent and relevant way. We discussed how teaching diversity so that health professionals deliver better care and manage their interactions with peers, colleagues, and patients is important but also requires institutions themselves to model what they expect of their learners. We found a paucity of high-quality research, which tells us the field is wide open for institutional and international collaboration. As educators finish reading this chapter they might want to challenge themselves and ask how their own perspectives influence how they aid or abet curricular development and implementation at their institution and what changes they might make.

References

1 Pew Charitable Trust (2016). Three Demographic Trends Changing Our World: Rise of migration, the middle class, and African faith trigger global shifts (19 July 2016). http://www.pewtrusts.org/en/research-and-analysis/analysis/2016/07/19/3-demographic-trends-changing-our-world

2 Toohey, S. (1999). *Designing Courses for Higher Education*, 44–49. Buckingham: The Society for Research into Higher Education and Open University Press.

3 Dogra, N., Reitmanova, S., and Carter-Pokras, O. (2010). Teaching cultural diversity: current status in U.K., U.S., and Canadian medical schools. *Journal of General Internal Medicine* 25 (Suppl): 164–168.

4 Hudelson, P., Dogra, N., Hendricks, V.P, et al. (2016). The challenges of integrating cultural competence into undergraduate medical curricula across Europe: experience from the C2ME, 'Culturally competent in medical education' project, MedEdPublish. http://www.mededpublish.org/manuscripts/365

5 Association of American Medical Colleges (1999). Report III Contemporary Issues. In: *Medicine, Communication in Medicine: Spirituality, cultural issues and end of life care*. Medical School Objectives Project, Association of American Medical Colleges. https://members.aamc.org/eweb/upload/Contemporary%20Issues%20In%20Med%20Commun%20in%20Medicine%20Report%20III%20.pdf

6 Dogra, N. (2004). The learning and teaching of cultural diversity in undergraduate medical education in the UK. PhD thesis. University of Leicester.

7 Global Diversity Practice (2017). www.globaldiversitypractice.co.uk/what-is-diversity-inclusion (accessed 14 October 2017).

8 University of Oregon (1999). http://gladstone.uoregon.edu/~asuomca/diversityinit/definition.html (accessed 14 April 2017).

9 Betancourt, J.R., Green, A.R., Carillo, E.J., and Ananeh-Firempong, A. (2003). Defining cultural competence: a practical framework for addressing racial/ethnic disparities in health and health care. *Public Health Reports* 118 (4): 293–302.

10 Seeleman, C., Suurmond, J., and Stronks (2009). Cultural competence: a conceptual framework for teaching and learning. *Medical Education* 43 (3): 229–237.

11 Metzl, J.M. and Hansen, H. (2014). Structural competency: theorizing a new medical engagement with stigma and inequality. *Social Science and Medicine* 103: 126–133.

12 Dogra, N. (2003). Cultural competence or cultural sensibility? A comparison of two ideal type models to teach cultural diversity to medical students. *International Journal of Medicine* 5: 223–231.

13 Dogra, N., Bhatti, F., Ertubey, C. et al. (2016). Teaching diversity to medical undergraduates: curriculum development, delivery and assessment. AMEE GUIDE no. 103. *Medical Teacher* 38 (4): 323–337.
14 Tajfel, H. and Turner, J.C. (1979). An integrative theory of intergroup conflict. In: *The Social Psychology of Intergroup Relations*, vol. 33 (ed. W. Austin and S. Worchel), 47. Monterey: Brooks/Cole.
15 Knipper, M., Seeleman, I., and Essink, M. (2010). How should ethnic diversity be represented in medical curricula? A plea for systematic training in cultural competence. *Tijdschrift voor Medisch Onderwijs* 29: 54–60.
16 Napier, A.D., Ancarno, C., Butler, B. et al. (2014). Culture and health. *Lancet* 384: 1607–1639.
17 Office of Minority Health. Think Cultural Health. U.S. Department of Health and Human Services. https://www.thinkculturalhealth.hhs.gov/clas/clas-tracking-map (accessed 20 October 2017).
18 Salas-Lopez, D., Holmes, L.J., Mouzon, D.M., and Soto-Greene, M. (2007). Cultural competency in New Jersey: evolution from planning to law. *Journal of Health Care for the Poor and Underserved* 18 (1): 35–43.
19 Smedley, B.D., Stith, A.Y., and Nelson, A.R. (2003). *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care (Committee on Understanding and Eliminating Racial and Ethnic Disparities in Health Care, Board on Health Sciences Policy)*. Washington DC: Institute of Medicine—National Academy Press.
20 Hesselink, G., Flink, M., Olsson, M. et al. (2012). Are patients discharged with care? A qualitative study of perceptions and experiences of patients, family members and care providers. *BMJ Quality and Safety* 21 (Suppl 1): i39–i49.
21 McKenzie, F., Joel, K., Williams, C., and Pritchard-Jones, K. (2014). Learning what high quality compassionate care means for cancer patients and translating that into practice. *Patient Experience Journal* 1: 2. http://pxjournal.org/journal/vol1/iss2/18.
22 Beach, M.C., Price, E.G., Gary, T.L. et al. (2005). Cultural competence: a systematic review of health care provider educational interventions. *Medical Care* 43 (4): 356–373.
23 Lie, D.A., Lee-Rey, E., Gomez, A. et al. (2011). Does cultural competency training of health professionals improve patient outcomes? A systematic review and proposed algorithm for future research. *Journal of General Internal Medicine* 26: 317–325.
24 Beal, A. (2004). Policies to reduce racial and ethnic disparities in child health and health care. *Health Affairs* 23: 171–179.
25 Lindgren, S. and Karle, H. (2011). Social Accountability of Medical Education: Aspects on Global Accreditation, pp. 667–672. http://www.tandfonline.com/doi/abs/10.3109/0142159X.2011.590246 (accessed 16 April 2017).
26 Powell-Sears, S.K. (2012). Improving cultural competence education: the utility of an intersectional framework. *Medical Education* 46 (6): 545–551.
27 Ewen, S., Barrett, J., and Howell-Meurs, S. (2016). Disparity and health professional education: a new approach. *Medical Science Educator* 26 (2): 247.
28 Berlin, E.A. and Fowkes, W.C. Jr. (1983). A teaching framework for cross-cultural health care: application in family practice. *Western Journal of Medicine* 139 (6): 934–938.
29 Leininger, M. (1991). *Culture Care, Diversity and Universality: A Theory of Nursing*. New York: National League for Nursing Press.
30 Kleinman, A. and Benson, P. (2006). Anthropology in the clinic: the problem of cultural competency and how to fix it. *PLoS Medicine* 3 (10): 1673–1676.
31 Duffy, M.E. (2001). A critique of cultural education in nursing. *Journal of Advanced Nursing* 36 (4): 487–495.
32 Sue, D. (1997). Multicultural training. *International Journal of Cultural Relations* 21: 175–193.
33 Dogra, N. and Karnik, N. (2003). First-year medical students' attitudes toward diversity and its teaching: an investigation at one US medical school. *Academic Medicine* 78 (11): 1–10.
34 Karnik, N. and Dogra, N. (2010). The cultural sensibility model for children and adolescents: a process oriented approach. *Child and Adolescent Psychiatric Clinics of North America* 19 (4): 719–738.
35 Betancourt, J. (2006). Cross-cultural medical education: conceptual approaches and frameworks for evaluation. *Academic Medicine* 78: 560–569.
36 Cross, T., Bazron, B., Dennis, K.W., and Isaacs, M.R. (1989). Towards a Culturally Competent System of Care. Georgetown University Child Development Center. p. 3. https://nccc-portal.techlab.works/uploads/publication/file/22/Towards_a_Culturally_Competent_System_of_Care.pdf (accessed 20 October 2017).
37 Kim, Y.Y. (1992). Intercultural communication competence: a systems-theoretic view. In: *Promoting Cultural Diversity: Strategies for Health Care Professionals* (ed. K.H. Kavanagh and P.H. Kennedy), 371. Newbury Park: Sage.
38 Polaschek, N.R. (1998). Cultural safety: a new concept in nursing people of different ethnicities. *Journal of Advanced Nursing* 27: 452–457.
39 Carberry, C. (1998). Contesting competency: cultural safety in advanced nursing practice. *Journal of the Royal College of Nursing, Australia* 5 (4): 9–13.
40 Richardson, S. and Williams, T. (2007). Why is cultural safety essential in health care. *Medicine and Law* 26: 699–707.
41 Majumdar, B., Browne, G., Roberts, J., and Carpio, B. (2004). Effects of cultural sensitivity training on health care provider attitudes and patient outcomes. *Journal of Nursing Scholarship* 36 (2): 161–166.
42 Majumdar, B. (1999). Cultural sensitivity training among foreign medical graduates. *Medical Education* 33: 177–184.
43 Lum, C.K. and Korenman, S.G. (1994). Cultural sensitivity training in US medical schools. *Academic Medicine* 69 (3): 239–241.
44 Tervalon, M. and Murray-Garcia, J. (1998). Cultural humility versus cultural competence: a critical distinction in defining physician-training outcomes in multicultural education. *Journal of Health Care for the Poor and Underserved* 9 (2): 117–125.
45 Hansen, H. and Metzl, J.M. (2017). New medicine for the U.S. health care system: training physicians for structural interventions. *Academic Medicine* 92 (3): 279–281.
46 Halman, M., Baker, L., and Ng, S. (2017). Using critical consciousness to inform health professions education: a literature review. *Perspectives on Medical Education* 6 (1): 12–20.
47 Thompson, D. ed. (1995). *The Concise Oxford Dictionary of Current English*. Oxford: Clarendon Press.
48 Dao, D.K., Goss, A.L., Hoekzema, A.S. et al. (2017). Integrating theory, content, and method to foster critical consciousness in medical students: a comprehensive model for cultural competence training. *Academic Medicine* 92 (3): 335–344.
49 Jenks, A.C. (2011). From "lists of traits" to "open-mindedness": emerging issues in cultural competence education. *Culture, Medicine and Psychiatry* 35: 209–235.
50 Loudon, R. and Greenfield, S. (1998). Undergraduate medical education must include improving health of minority ethnic communities. *British Medical Journal* 317: 1660.
51 Godkin, M. and Weinreb, L. (2001). A pathway on serving multicultural and underserved populations. *Academic Medicine* 76: 513–514.
52 Culhane-Pera, K.A., Reif, C., Egli, E. et al. (1997). A curriculum for multicultural education in family medicine. *Family Medicine* 29: 719–723.
53 Gudykunst, W.B. (1992). Being perceived as a competent communicator. In: *Promoting Cultural Diversity: Strategies for Health Care Professionals* (ed. K.H. Kavanagh and P.H. Kennedy), 388. Newbury Park: Sage.
54 Logan, S. and Spencer, N. (2000). Inequality and children's health. *Child: Care, Health and Development* 26 (1): 1–3.
55 Kirmayer, L. (2012). Rethinking cultural competence. *Transcultural Psychiatry* 49 (2): 149–164.

56 Deloney, L.A., Graham, C.J., and Erwin, D.O. (2000). Presenting cultural diversity and spirituality to first-year medical students. *Academic Medicine* 75: 513–514.
57 American Medical Association (1999). Enhancing the 'Cultural Competence' of Physicians. Council on Medical Education Report 5-A-98. In: *Cultural Competence Compendium*. American Medical Association.
58 Dogra, N. and Wass, V. (2006). Can we assess students' awareness of 'cultural diversity'? A qualitative study of stakeholders' views. *Medical Education* 40: 682–690.
59 Wertsch, J.V. (1991). *Voices of the Mind: A Sociocultural Approach to Mediated Action*. Cambridge, MA: Harvard University Press.
60 Kumaş-Tan, Z., Beagan, B., Loppie, C. et al. (2008). Measures of cultural competence: examining hidden assumptions. *Academic Medicine* 82: 548–557.
61 Gozu, A., Beach, M.C., Price, E.G. et al. (2007). Self-administered instruments to measure cultural competence of health professionals: a systematic review. *Teaching and Learning in Medicine* 19 (2): 180–190.
62 Curcio, A., Ward, T., and Dogra, N. (2012). Educating culturally sensible lawyers: a study of student attitudes about the role culture plays in the lawyering process. *University of Western Sydney Law Review* 16: 98–126.
63 Anderson, L.M., Scrimshaw, S.C., Fullilove, M.T. et al. (2003). Culturally competent healthcare systems: a systematic review. *American Journal of Preventative Medicine* 24 (3S): 68–79.
64 Bhui, K., Warfa, N., Edonya, P. et al. (2007). Cultural competence in mental health care: a review of model evaluations. *BMC Health Services Research* 7: 15.
65 Forsyth, C.J., Irving, M.J., Tennant, M. et al. (2017). Teaching cultural competence in dental education: a systematic review and exploration of implications for indigenous populations in Australia. *Journal of Dental Education* 81 (8): 956–968.
66 Gallagher, R.W. and Polanin, J.R. (2015). A meta-analysis of educational interventions designed to enhance cultural competence in professional nurses and nursing students. *Nurse Education Today* 35: 333–340.
67 Govere, L. and Govere, E.M. (2016). How effective is cultural competence training of healthcare providers on improving patient satisfaction of minority groups? A systematic review of literature. *Worldviews on Evidence-Based Nursing* 13 (6): 402–410.
68 Horvat, L., Horey, D., Romios, P., and Kis-Rigo, J. (2014). Cultural competence education for health professionals. *Cochrane Database of Systematic Reviews* 5 (Art. No.: CD009405). doi: 10.1002/14651858.CD009405.pub2.
69 Jernigan, V.B.B., Hearod, J.B., Tran, K. et al. (2016). An examination of cultural competence training in US medical education guided by the Tool for Assessing Cultural Competence Training. *Journal of Health Disparities Research and Practice* 9 (3): 150–167.
70 Maina, I.W., Belton, T.D., Ginzberg, S. et al. (2018). A decade of studying implicit racial/ethnic bias in healthcare providers using the implicit association test. *Social Science and Medicine* 199: 219–229.
71 Renzaho, A.M.N., Romios, P., Crock, C., and Sonderlund, A.L. (2013). The effectiveness of cultural competence programs in ethnic minority patient-centered health care – a systematic review of the literature. *International Journal for Quality in Health Care* 25 (3): 261–269.
72 Zestcott, C.A., Blair, I.V., and Stone, J. (2016). Examining the presence, consequences, and reduction of implicit bias in health care: a narrative review. *Group Processes & Intergroup Relations* 19 (4): 528–542.
73 Price, E.G., Beach, M.C., Gary, T.L. et al. (2005). A systematic review of the methodological rigor of studies evaluating cultural competence training of health professionals. *Academic Medicine* 80 (6): 578–586.
74 Campinha-Bacote, J. (2007). Inventory for Assessing the Process of Cultural Competence Among Healthcare Professionals – Revised (IAPCC-R). http://transculturalcare.net/iapcc-r (accessed 27 June 2018).
75 Brathwaite, A.C. and Majumdar, B. (2006). Evaluation of a cultural competence educational programme. *Journal of Advanced Nursing* 53: 470–479.
76 Ferguson, J., Keller, D., Haley, H., and Quirk, M. (2003). Developing culturally competent community faculty: a model program. *Academic Medicine* 78 (12): 1221–1228.
77 Bernal, H. and Froman, R. (1987). The confidence of community health nurses caring for ethnically diverse populations. *Journal of Nursing Scholarship* 19: 201–203.
78 van Ryn, M., Hardeman, R., Phelan, S.M. et al. (2015). Medical school experiences associated with change in implicit racial bias among 3457 students: a medical student CHANGES study report. *Journal of General Internal Medicine* 30: 1748–1756.
79 Castillo, L.G., Reyes, C.J., Brossart, D.F. et al. (2007). The influence of multicultural training on perceived multicultural counseling competencies and implicit racial prejudice. *Journal of Multicultural Counseling and Development* 35 (4): 243–254.
80 Thom, D.H., Tirado, M.D., Woon, T.L., and McBride, M.R. (2006). Development and evaluation of a cultural competency training curriculum. *BMC Medical Education* 6 (1): 6–38.
81 Association of American Medical Colleges (2006). Tool for Assessing Cultural Competence Training (TACCT) Resource Guide. Association of American Medical Colleges.
82 Neff, J., Knight, K.R., Satterwhite, S. et al. (2017). Teaching structure: a qualitative evaluation of a structural competency training for resident physicians. *Journal of General Internal Medicine* 32 (4): 430–433.
83 Nazar, M., Kendall, K., Day, L., and Nazar, H. (2015). Decolonising medical curricula through diversity education: lessons from students. *Medical Teacher* 37 (4): 385–393.
84 Dogra, N. and Vostanis, P. (2007). Providing clinical services for a diverse population: views on training of child and adolescent mental health practitioners. *Journal of Interprofessional Care* 21 (6): 645–655.
85 Lave, J. and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge, UK: Cambridge University Press.
86 King, E.B., Dawson, J.F., Kravitz, D.A., and Gulick, L.M.V. (2012). A multilevel study of the relationships between diversity training, ethnic discrimination and satisfaction in organizations. *Journal of Organizational Behavior* 33 (1): 5–20.
87 Lie, D., Carter-Pokras, O., Braun, B., and Coleman, C. (2012). What do health literacy and cultural competence have in common? Calling for a Collaborative Health Professional Pedagogy. *Journal of Health Communication: International Perspectives.* 17: 13–22.
88 Fortier, J.P. and Bishop, D. (2003). Setting the agenda for research on cultural competence in health care: final report (ed. C. Brach). Rockville, MD: U.S. Department of Health and Human Services Office of Minority Health and Agency for Healthcare Research and Quality.
89 Beagan, B.L. (2003). Teaching social and cultural awareness to medical students: it's all very nice to talk about it in theory but ultimately it makes no difference. *Academic Medicine* 78: 605–614.
90 Hafferty, F.W. and Franks, R. (1994). The hidden curriculum, ethics teaching, and the structure of medical education. *Academic Medicine* 69: 861–871.
91 Smith, D.G. (2012). Building institutional capacity for diversity and inclusion in academic medicine. *Academic Medicine* 87: 1511–1515.
92 Carter-Pokras, O., Bereknyei, S., Lie, D. et al. (2010). Surmounting the unique challenges in health disparities education: a multi-institution qualitative study. *Journal of General Internal Medicine* 25 (Suppl 2): S108–S114.
93 Lie, D.A., Boker, J., Crandall, S. et al. (2008). Revising the Tool for Assessing Cultural Competence Training (TACCT) for curriculum evaluation: findings derived from seven US schools and expert consensus. *Medical Education Online* 1 (13): 1–11.

94 Wilkinson, T.J. (2016). Stereotypes and the hidden curriculum of students. *Medical Education* 50 (8): 802–804. doi: 10.1111/medu.13008.

95 Meiboom, A., Diedrich, C., De Vries, H., Hertogh, C. and Scheele, F. (2015). The hidden curriculum of the medical care for elderly patients in medical education: a qualitative study. *Gerontology & Geriatrics Education* 36 (1): 30–44.

96 Phillips, S.P. and Clarke, M. (2012). More than an education: the hidden curriculum, professional attitudes and career choice. *Medical Education* 46: 887–893.

97 Birden, H., Glass, N., Wilson, I. et al. (2013). Teaching professionalism in medical education: a Best Evidence Medical Education (BEME) systematic review. BEME Guide no. 25. *Medical Teacher* 35 (7): e1252–e1266.

98 Aronson, L. (2011). Twelve tips for teaching reflection at all levels of medical education. *Medical Teacher* 33: 200–205.

99 Hickson, G., Pichert, J.W., Webb, L.E. and Gabbe S.G. (2007). A complementary approach to promoting professionalism: identifying, measuring, and addressing unprofessional behaviors. *Academic Medicine* 82 (11): 140–148.

100 Tariq, S.G., Thrush, C.R., Gathright, M. et al. (2016). Crystal clear or tin ear: how do medical students interpret derogatory comments about patients and other professionals? *Medical Education Online* 13: doi: 10.3402/meo.v21.31221.

101 Acosta, D. and Ackerman-Barger, K. (2017). Breaking the silence: time to talk about race and racism. *Academic Medicine* 92 (3): 285–288.

103 Sorensen, J., Norredam, M., Dogra, N. et al. (2017). Enhancing cultural competence in medical education. *International Journal of Medical Education* 8: 28–30.

103 Liaison Committee on Medical Education (2017). Functions and Structure of a Medical School: Standards for Accreditation of Medical Education Programs Leading to the M.D. degree. http://lcme.org/publications (accessed 30 October 2017).

104 Andrulis, D.P., Siddiqui, N.J., and Cooper, M.R. (2013). The Affordable Care Act and Racial and Ethnic Health Equity Series. Data, Research and Quality: Implementation Progress At-A-Glance. Texas Health Institute.

105 Equality Act (2010) Chapter 15. London: The Stationery Office.

106 Institute of Medicine (2004). *Health Literacy: A Prescription to End Confusion*. Washington DC: National Academy of Sciences.

107 Berkman, N.D., Sheridan, S.L., Donahue, K.E., et al. (2011). Health Literacy Interventions and Outcomes: An Updated Systematic Review. Evidence Report/Technology Assessment No. 199. (Prepared by RTI International—University of North Carolina Evidence-based Practice Center under Contract No. 290–2007-10056-I.) AHRQ Publication No. 11-E006. Rockville, MD. Agency for Healthcare Research and Quality. March 2011. http://www.ahrq.gov/clinic/tp/lituptp.htm.

Further Reading

Acosta, D. and Ackerman-Barger, K. (2017). Breaking the silence: time to talk about race and racism. *Academic Medicine* 92 (3): 285–288.

Dogra, N., Bhatti, F., Ertubey, C., et al. (2015). Teaching Diversity to Medical Undergraduates: Curriculum Development, Delivery and Assessment. AMEE guide No. 103.

Kirkcaldy, B. (2013). *Chimes of Time: Wounded Health Professionals. Essays on Recovery*. Leiden: Sidestone Press.

Lie, D., Carter-Pokras, O., Braun, B., and Coleman, C. (2012). What do health literacy and cultural competence have in common? Calling for a Collaborative Health Professional Pedagogy. *Journal of Health Communication: International Perspectives* 17 (Sup 3): 13–22.

McCann, M., Carter-Pokras, O., and Braun, B. (2013). Cultural Competency and Health Literacy Primer: A Guide for Teaching Health Professionals and Students. Maryland Department of Health and Mental Hygiene's Office of Minority Health and Health Disparities, University of Maryland School of Public Health, and Herschel S. Horowitz Center for Health Literacy.

36 Developing Medical Educators: A Journey, not a Destination

Yvonne Steinert
Centre for Medical Education, Faculty of Medicine, McGill University, Montreal, Quebec, Canada

KEY MESSAGES

- The term 'medical educator' describes a number of roles, including that of teacher and assessor, curriculum planner and evaluator, educational leader and manager, researcher, and scholar.
- There are many ways in which to develop as a medical educator; 'formal' staff development activities are only one way to achieve this goal. Other approaches include 'learning on the job', mentorship and role modelling, belonging to a community of educators, and organisational support and development.
- The content of staff development activities should move beyond the enhancement of teaching effectiveness to include leadership, the promotion of scholarship, and organisational change and development.
- Staff development activities should be guided by knowledge of core competencies for medical educators.
- Students and residents should be introduced to staff development activities early in their careers.
- Medical education is a social endeavour and the idea of community, and communities of practice, is fundamental to the development of medical educators.

Introduction

The Latin origin of the word 'doctor' (*doceo*) translates as 'I teach', but the majority of doctors, although expert in *what* they teach, have had little or no training in *how* to teach [1]. Doctors are also minimally prepared for the various roles that are subsumed under the term 'medical educator'. As Jason and Westberg have stated: 'The one task that is distinctively related to being a faculty member is teaching; all the other tasks can be pursued in other settings. Paradoxically, the central responsibility of faculty members is typically the one for which they are least prepared' [2].

In the past, it was assumed that intelligent people who have been students for many years have learnt – or can automatically learn – to be successful faculty members, and little or no support for staff development was provided [3]. This is no longer true. Increasing attention has been placed on the design and development of staff development programmes, in diverse contexts and settings [4–7]. Also, in recent years, a number of regulatory and international bodies have started to pay attention to the accreditation of teachers and teaching [8–10], and they have highlighted the importance of staff development in the certification of educators [11]. In the UK, for example, the role of teacher has for some time been recognised as a core professional activity for all doctors, and one that cannot be left to chance, aptitude, or inclination [12].

The goal of this chapter is to focus on the development of medical educators. To achieve this objective, we will try to define what is meant by a *medical educator*, describe the required core competencies, and examine different ways of developing medical educators. As staff development is one of the most common ways to achieve this objective, much of this chapter will examine what is known about formal staff development programmes. However, we will also address the role and importance of work-based learning, communities of practice, mentorship and role modelling, and organisational support and development.

What is a Medical Educator?

The medical education literature tends to use the terms 'teacher' and 'medical educator' interchangeably, with no clear definitions. To inform this chapter, we conducted a series of semi-structured interviews with 12 medical educators at McGill to ascertain their definitions and conceptions of being a medical educator [13]. Definitions included a broad range of conceptualisations, some of which are highlighted below.

- The medical educator is someone who critically reflects on the quality of the educational experience and tries to innovate and improve on what they have done.
- Medical educators have a passion, not just an interest, but a passion for bringing out the best, or finding ways

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

to bring out the best, in learners that they work with to help develop the best physicians that we can …

- A medical educator is someone who uses theories and principles of education in their activities. This includes teaching, scholarship, curricular design and evaluation, and research across the educational continuum …

What is striking in these definitions is the emphasis on reflection, passion, innovation, and informed practice across a continuum of tasks and activities.

In this chapter, the term 'medical educator' will encompass a broad spectrum of roles that include teaching and assessment, curriculum design and evaluation, educational leadership and innovation, and research and scholarship. Moreover, medical educators, whether in the clinical or classroom setting, will refer to individuals who actively reflect on what they do, using experience and available evidence to inform their educational practice and to enhance the teaching and learning of future health care professionals. As one individual reflected: [We] continually ask ourselves how we can do this better: 'How can we get [our students] to be the best that they can be' … 'If my program is working, can I still make it better? If it is not working, why not and how can I improve it?' [13].

What is Faculty Development?

Staff development, or faculty development, as it is often called, has become an increasingly important component of medical education. Faculty development activities have been designed to improve teacher effectiveness at all levels of the educational continuum (i.e. undergraduate, postgraduate, and continuing professional development), and diverse programmes have been offered to health care professionals in many settings [14]. In this chapter, staff development will refer to all activities health professionals pursue to improve their knowledge, skills, and behaviours as teachers and educators, leaders and managers, and researchers and scholars, in both individual and group settings [7]. Moreover, the term 'staff' or 'faculty' refers to *all* individuals who are involved in the teaching and education of learners, at all levels of the continuum, across all health professions [7]; it does not connote a particular employment or contractual relationship. Importantly, the goal of staff development is to teach faculty members the skills relevant to their institutional and faculty role and to sustain their vitality, both now and in the future [14].

Faculty development can provide medical educators with knowledge and skills about teaching and learning, curriculum design and delivery, learner assessment and programme evaluation, leadership and management, as well as research and scholarship. It can also reinforce or alter attitudes or beliefs about education and scholarly activity, provide a conceptual framework for what is often performed on an intuitive basis, and introduce clinicians and basic scientists to a community of medical educators interested in medical education and the enhancement of teaching and learning for students, patients, and peers.

Staff development can also serve as a useful instrument in the promotion of organisational change [15, 16]. That is, staff development can help build consensus, generate support and enthusiasm, and implement a change initiative; it can also help change the culture within the institution by altering the formal, informal, and hidden curriculum [17, 18] and by enhancing organisational capacities [19]. As Swanwick [20] has stated, staff development should be: 'An institution-wide pursuit with the intent of professionalizing the educational activities of teachers, enhancing educational infrastructure, and building educational capacity for the future …'

In many ways, staff development can play an important role at both the *individual* and the *organisational* level [21]. In addition, although staff development activities predominantly focus on teaching and instructional effectiveness, there is a critical need for these activities to address the other roles of medical educators, including that of curriculum designer, educational leader and manager, and scholar.

A Curriculum for Staff Development?

Interestingly, most staff development programmes have been developed independently of a curriculum for teachers and educators. Rather, they are often based on perceived or self-identified needs. However, as Purcell and Lloyd-Jones observed: 'Faculties have developed a plethora of teacher training programmes for medical teachers. But what is good medical teaching? Unless we know what it is, how can we develop it?' [12].

In recent years, a number of authors have proposed frameworks to synthesise and consolidate academic or pedagogical competencies (e.g. [3, 22–24]), and in many ways, it would be worthwhile to consider these frameworks in the design and delivery of staff development programmes. For example, the UK-based Academy of Medical Educators [25] has developed *Professional Standards* for medical educators that are divided into core values of medical educators and five domains, outlining 'detailed outcomes in terms of understanding, skills and behaviour required of medical educators'. The suggested core values include: professional integrity, educational scholarship, equality of opportunity and diversity, respect for the public, respect for patients, respect for learners, and respect for colleagues. Importantly, these core values underpin the professional practice and development of medical educators and serve as the foundation for the following five domains of educational practice:

- the design and planning of learning activities
- teaching and supporting learners
- assessment and feedback to learners
- educational research and evidence-based practice
- educational management and leadership.

Each of these domains describes a set of standards that are expected of medical educators at different levels of engagement and expertise and can serve as a useful tool for self-assessment as well as programme development.

In another context, Srinivasan et al. [24] described 'teaching as a competency' and detailed six competencies for medical educators:

- medical (or content) knowledge
- learner-centredness

- interpersonal and communication skills
- professionalism and role modelling
- practice-based reflection and improvement
- systems-based learning.

These competencies mirror those expected of residents (or junior doctors) in training and can similarly be used to define a curriculum for faculty development designed to enhance teaching effectiveness.

Other researchers have compared student and faculty perceptions of effective clinical teaching. For example, Buchel and Edwards [26] asked residents and faculty members to rate the attributes of effective teachers. Both residents and faculty agreed that clinical competence is one of the most important attributes of an effective clinical teacher. They also agreed that better educators were those who demonstrated enthusiasm for their educational responsibilities. At the same time, residents commented that it was important for a quality educator to respect their autonomy and independence as clinicians, whereas faculty members reported that this was one of the least important traits of an effective teacher. In addition, faculty members felt that serving as a role model worth emulating was essential, a factor stressed by previous authors [27]. Residents, however, did not believe that this was an important attribute and ranked it at the bottom of their list. Clearly, the perceptions of residents and faculty members are not always congruent, although an 'evidence-based' set of attitudes and behaviours should guide the development of staff development programmes.

It should also be noted that much less has been written about the roles of educational *leader* and *scholar*, roles that are often subsumed under the term 'medical educator'. In an interesting study, Bordage et al. [28] surveyed deans and associate deans to identify the educational and leadership skills required of 'programme directors with major educational and leadership responsibilities'. Their results indicated the importance of nine key skill areas: oral communication, interpersonal abilities, clinical competence, educational goal definition, educational design, problem solving and decision-making, team-building, written communication, and budgeting and financial management.

Spencer and Jordan [29] also highlighted the fact that educational change requires leadership and that we need to equip our colleagues to implement change. Clearly, the development of medical educators should address leadership competencies as well as those that promote scholarship in its broadest sense.

Boyer [30] identified the following four categories of scholarship:

- discovery
- integration
- application
- teaching.

Roughly considered, scholarship in education can involve the discovery of new knowledge (i.e. research), the integration or application of existing knowledge to new areas, and teaching [31]. Scholarship also provides a common ground for assessing the diverse roles and contributions of faculty members to the mission of the medical school and can take on many forms.

The scholarship of *discovery* has been synonymous with research in the traditional sense. In medical education, this may include original research (e.g. how expertise is developed) or forming new theory (e.g. how physicians learn in practice). Peer-reviewed grants and publications can also be products of the scholarship of discovery.

The scholarship of *integration* has been defined as making connections across the disciplines, illuminating data in a revealing way. Examples in medical education include systematic reviews or integrating concepts from other fields such as anthropology, sociology, or education into medicine.

The scholarship of *application* has been likened to 'service' in one's own field of knowledge, to the application of theory into practice. A common example is using research evidence to develop a new curriculum or assessment method based on available evidence [32]. In medical education, designing innovative instructional materials or developing a fellowship or faculty development programme are examples of the scholarship of integration and application.

The *scholarship of teaching* involves the capacity to effectively communicate one's own knowledge, skills, and beliefs. Moreover, teaching becomes scholarship when it is made public, is available for peer review and critique, and can be reproduced and built on by other scholars [33].

The promotion of scholarship, and helping educators foster scholarly activities among their colleagues, are important factors in the development of medical educators, and yet this area is often neglected.

As this discussion suggests, the development of a curriculum for medical educators merits attention. At the same time, it would also be worth focusing on a better understanding of the teacher's lived experience. In an interesting study, Higgs and McAllister [34] studied the 'experience of being a clinical educator' and discovered that this experience consisted of six interactive and dynamic dimensions:

- a sense of self (or self-identity)
- a sense of relationship with others
- a sense of being a clinical educator
- a sense of agency, or purposeful action
- seeking dynamic self-congruence
- the experience of growth and change.

Based on this research, and our experience at McGill, it would be worthwhile to take a more careful look at the 'lived experience' of being an educator and to use this experience as a framework for training. As one colleague stated: 'My pride as a medical educator comes from watching the light go on in my students' eyes and knowing why the light goes off …' [13] 'When I see junior colleagues work and demonstrate excellence in patient care, going the extra mile … I know that I have had an impact' [13]. Understanding the meaning of teaching for faculty members [35] and the important role of identity in teachers' sense of commitment and fulfilment would also enhance the design and delivery of faculty development programmes.

In summary, medical teachers and educators need to be prepared for complex [36] and demanding roles that include teaching and assessment, leadership and management, and scholarship in its broadest sense.

How Can We Develop Medical Educators?

'Formal' Approaches

The most common staff development formats include workshops, short courses and seminars, fellowships, and other longitudinal programmes [39, 40]. Other formats include degree programmes, peer coaching, augmented feedback, and online learning. A brief description of some of these formats, highlighted in Figure 36.1, follows.

Workshop, Seminars, and Short Courses

Workshops are popular because of their inherent flexibility and promotion of active learning. In particular, teachers value a variety of teaching methods within this format, including interactive lectures, small-group discussions and exercises, role-plays and simulations, and experiential learning [39]. Workshops are commonly used to promote skill acquisition (e.g. lecturing or small-group teaching skills) [42, 43], to prepare for new curricula (e.g. problem-based learning) [44, 45], or to help faculty adapt to new teaching environments (e.g. teaching in the ambulatory setting) [46, 47]. Workshops on leadership styles and skills [48] and/or curriculum design and innovation [49] can also help prepare educators for their leadership roles, whereas short courses on research methods [50] and writing for publication [51] can help prepare clinicians and basic scientists for their scholarly work.

To date, the majority of staff development programmes have focused on teaching improvement. That is, they aim to enhance teachers' performance in the classroom and the clinical setting, address conceptions of teaching and learning, promote the acquisition of specific teaching skills (e.g. small-group facilitation, giving feedback), focus on learner assessment, and review instructional design and curriculum development [39, 40]. A number of programmes also target specific core competencies (e.g. the teaching and assessment of communication skills, professionalism) and the use of technology in teaching and learning. Less attention has been paid to the personal development of health care professionals, educational leadership and scholarship, and organisational development and change. Although instructional effectiveness at the individual level is critically important, a more comprehensive approach is needed [14]. We clearly need to develop individuals who will be able to provide leadership to educational programmes, act as educational mentors, and design and deliver innovative educational programmes. As Cusimano and David [52] pointed out, there is an enormous need for more health care professionals trained in methods of educating others so that medical education will continue to be responsive to driving forces of change. As previously stated, staff development also has an important role to play in promoting teaching as a scholarly activity and in creating an educational climate that encourages and rewards educational leadership, innovation, and excellence.

Fellowships and other Longitudinal Programmes

Fellowships of varying length, format, and emphasis have been utilised in many disciplines [53–56]. More recently, integrated, longitudinal programmes have been developed as an alternative to fellowship programmes or sabbaticals. These programmes, in which faculty commit 10–20% of their time over one to two years, allow health care professionals to maintain most of their clinical, research, and

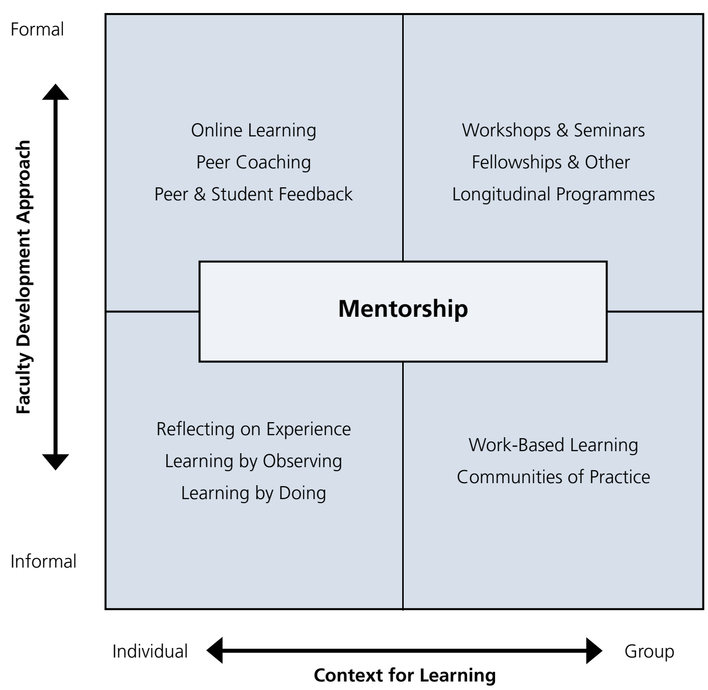


Figure 36.1 Approaches to faculty development*. *This figure was originally prepared for a chapter on 'Becoming a Better Teacher: From Intuition to Intent' [41]. (*Source*: Reprinted with permission by the American College of Physicians © 2010.)

administrative responsibilities, whilst furthering their own professional development. Programme components typically consist of a variety of methods, including university courses, monthly seminars, independent research projects, and participation in staff development activities. Integrated longitudinal programmes, such as a *Teaching Scholars Program* [57–60], have particular appeal because teachers can continue to practise and teach while improving their educational knowledge and skills. These programmes also allow for the development of educational leadership and scholarly activity in medical education [61].

In summary, although fellowships and other longitudinal programmes vary in structure, duration, and content, they all enable the acquisition of expertise and excellence in teaching, curricular design and evaluation, and educational leadership. Many of them also provide assistance in academic and career development [62, 63] and help create a community of teachers and educators. In addition, they encourage the dissemination of new knowledge and understanding to further the field of medical education.

Degree Programmes

Certificate or degree programmes are becoming increasingly popular in many settings. In part, this is due to what some authors have termed the 'professionalisation' of medical education [11, 12]. Several authors and organisations have argued for the need to certify medical educators and thereby ensure global standards; others do not agree and worry about disenfranchising keen and committed educators (see Box 36.1).

An advanced degree in medical education offers a grounding in educational theory and practice and can provide the foundation for educational research and scholarship. Cohen et al. [64] reported on 21 different Master's degree programmes in medical education in Holland [1], Canada [3], Australia [3], the USA [6], and the UK [8]. A more recent review of Master's degrees in medical education was conducted by Tekian and Harris [65], who highlighted an increased proliferation of Master's level programmes for health professions education. These authors described the commonalities (e.g. focus, content, educational requirements) and differences (e.g. structure, organisation) of 71 programmes and argued for the need to establish accreditation processes, based on common criteria and methods, for evaluating these programmes. In addition, they suggested that there is a need to address the geographic maldistribution of advanced degree programmes. Pugsley et al. [66] also commented on the variability in content and quality among Master's programmes in medical education and argued for increased standards and quality

BOX 36.1 FOCUS ON: The professionalisation of medical education

One of the Dutch terms for staff development is *Docentprofessionalisering*, which loosely translates as the 'professionalisation' of teaching. This is of particular interest as we witness the professionalisation of medical education in a number of venues. For example, in the UK, the Dearing Report made a number of recommendations about faculty in higher education, including the following which is pertinent in this context:

> We recommend that institutions of higher education begin immediately to develop or seek access to programmes for teacher training of their staff, if they do not have them, and that all institutions seek national accreditation of such programmes from the Institute for Learning and Teaching in Higher Education [37].

The UK General Medical Council's *Good Medical Practice* states that:

> Teaching, training, appraising and assessing doctors and students are important for the care of patients now and in the future. You should be willing to contribute to these activities … If you are involved in teaching you must develop the skills, attitudes, and practices of a competent teacher [8].

More recently, the General Medical Council has adopted a national plan to recognise and approve trainers, building on the work of the Academy of Medical Educators in the UK [25], who developed a series of standards for medical educators, organised into five domains.

On an international stage, the World Federation of Medical Educators has articulated a Staff Activity and Development Policy which states the following:

> The medical school must formulate and implement a staff activity and development policy which: allows a balance of capacity between teaching, research and service functions; ensures recognition of meritorious academic activities, with appropriate emphasis on teaching, research and service qualifications; ensures that clinical service functions and research are used in teaching and learning; ensures sufficient knowledge by individual staff members of the total curriculum; and includes teacher training, development, support and appraisal [10].

In a similar vein, the International Association of Medical Colleges has stated that 'opportunities for professional development must be provided to enhance faculty members' skills and leadership abilities in education and research' [9]. More specifically, their basic standard states that 'the medical school must have a staff policy which addresses a balance of capacity for teaching, research and service functions, and ensures recognition of meritorious academic activities, with appropriate emphasis on both research attainment and teaching qualifications'.

Although the emphasis on accreditation for teaching and standards for teaching has not received the same attention in Canada or the USA, there is clearly a movement towards increased accountability and accreditation of teaching, one of the roles of the medical educator [38]. Moreover, as Eitel et al. [11] suggest, staff development is one of the prerequisites for certification leading to the professionalisation of medical educators.

BOX 36.2 HOW TO: Become a medical educator

Colleagues at the Centre for Medical Education at McGill University identified the following ways of 'becoming a medical educator' [13].

- *By the nature of my responsibilities*

  One of the nice things about medical education is that you can often have an administrative position that allows you to have a lab. I was undergraduate program director for quite a while … by doing that it's given me a lab to try various innovations and evaluate them.
- *By participating in staff development and other training opportunities*

  Participating in faculty development workshops introduced me to a community of educators and got me 'hooked' … I haven't stopped learning since.
- *By pursuing an advanced degree*

  My advanced degree allowed me to look at things with education glasses on. It also gave me the opportunity to immerse myself in a group with similar interests and needs.
- *By wanting to*

  I had the interest and the desire. I have always wanted to be a teacher and I was good at teaching … I followed my passion.
- *By belonging to a community (of experts)*

  For me the most valuable part has been meeting regularly with a group of like-minded individuals committed to excellence and scholarship in medical education … I have become immersed in the culture.
- *By being mentored and through role modelling*

  I could not have done this alone …
- *By doing medical education*

  I have learned by doing – and using, either explicitly, or implicitly, what I have learned over the years to inform my teaching.

assurance. Nonetheless, despite these concerns, most universities in the UK now require staff to undertake a university certificate in teaching and learning, and many medical schools, in partnership with National Health Service (NHS) trusts, are providing opportunities for advanced educational training [67]. Degree programmes can be particularly helpful to individuals interested in educational leadership, administration, or research.

Peer Coaching

Peer coaching as a method of staff development has been described extensively in the education literature. Key elements of peer coaching include the identification of individual learning goals (e.g. improving specific teaching skills), focused observation of teaching by colleagues, and the provision of feedback, analysis, and support [68]. In addition, peer coaching is a highly personalised, learner-centred approach that requires a safe environment, mutual trust and collegiality, and reflection [69]. This underutilised approach, sometimes called *co-teaching* or *peer observation*, has particular appeal because it occurs in the teacher's own practice setting, enables individualised learning, and fosters collaboration [70]. It also allows health care professionals to learn about each other, as they teach together and, in this way, can nurture the development of medical educators.

Learner Feedback

Feedback and assessments from learners, at all levels of the educational continuum, can also be a helpful catalyst for faculty development [71, 72]. Unfortunately, teachers are often taken aback by learners' comments and observations, and this opportunity for self-improvement can easily be missed. However, an appreciative inquiry of student or resident assessments can provide useful information, especially if teachers pose the following questions:

- Is there a pattern that runs across diverse assessments?
- What am I doing well?
- What might I do differently?
- How can I use this as an opportunity for learning about myself?

'Informal' Approaches

Although staff development programmes are a popular way of developing medical educators, a number of alternative approaches should also be considered. Box 36.2 describes how medical educators at McGill University started the 'journey' of becoming a medical educator. For many, it started with their 'job responsibilities' and slowly evolved into a career path. The following section describes four of these important pathways.

Work-based Learning

Work-based learning, often defined as learning *for* work, learning *at* work, and learning *from* work [20], is fundamental to the development of medical educators, for whom 'learning on the job' is often the first entry into teaching and education. In fact, it is in the everyday workplace – where educators conduct their clinical, research, and teaching activities, and interact with faculty, colleagues, and students – that learning most often takes place. It would therefore be extremely worthwhile to help medical educators see their everyday experiences as 'learning experiences' and encourage them to reflect with colleagues and students on learning that has occurred in the clinical or classroom setting [73]. It would also be beneficial to bring staff development to the workplace.

It is interesting that staff development activities have traditionally been conducted away from the educator's workplace, requiring participants to take their 'lessons learnt' back to their own contexts. Perhaps it is time to reverse this trend and think about how we can enhance the learning that takes place in the work environment [74]. By working together and participating in a larger community, clinicians and basic scientists can build new knowledge and understanding and develop approaches to problems faced in teaching and learning.

Communities of Practice

Closely related to 'learning at work' is the concept of situated learning and communities of practice [75]. In my own setting, medical educators have commented on the role and value of a community of medical educators, brought together by a common interest in the enhancement of teaching and learning for students across the educational continuum and involvement in scholarly work in medical education, as a critical factor in their own development. As one colleague observed: 'For me it has been beneficial to be immersed in a group which is actually physically removed from where I do my clinical work … where I come to regularly and am forced to engage in the discussions and try out new ideas' [13].

Barab et al. [76] define a community of practice as a 'persistent, sustaining, social network of individuals who share and develop an overlapping knowledge base, set of beliefs, values, history, and experiences focused on a common practice and/or mutual enterprise'. Clearly, formal staff development programmes can play a pivotal role in developing communities of practice. At the same time, belonging to such a community can play a critical role in the development of medical educators.

Lave and Wenger [75] suggest that the success of a community of practice depends on the following five factors:

- the existence and sharing by the community of a common goal
- the existence and use of knowledge to achieve that goal
- the nature and importance of relationships formed among community members
- the relationships between the community and those outside it
- the relationship between the work of the community and the value of the activity.

In his later work, Wenger [77] also adds the notion that achieving the shared goals of the community requires a shared repertoire of common resources, including language, stories, and practices (see Box 36.3).

BOX 36.3 FOCUS ON: Cultivating communities of practice

Wenger et al. [90] describe seven principles for cultivating communities of practice that have direct relevance for the development of medical educators. They include the following:

Design for evolution Communities should build on pre-existing personal networks and allow for natural growth and development. Design elements should be catalysts for the community's natural evolution.

Open a dialogue between insider and outsider perspectives Good community design requires an insider's perspective to lead the discovery of what the community is about; however, an outside perspective is often needed to help members see untapped possibilities.

Invite different levels of participation People participate in communities for different reasons (e.g. because the community provides value, for personal connections, to enhance skill development). Different levels of engagement (e.g. core, active, peripheral) are to be expected and encouraged. Successful communities often 'build benches' for those on the side lines.

Develop both public and private community spaces Most communities have public events where members gather to exchange ideas, solve problems, or explore new concepts. However, communities are more than their 'calendar of events'. The heart of the community is the one-on-one networking that occurs informally. All communities require both public and private interactions.

Focus on value Communities thrive because they deliver value to the organisation, to the teams on which community members serve, and to the community members themselves. Articulation of its value helps a community to develop and grow.

Combine familiarity and excitement Successful communities offer the 'familiar comforts of a hometown' but also provide enough interest and variety to keep new ideas and new people cycling into the community.

Create a rhythm for the community There are many rhythms in a community (e.g. the syncopation of familiar and exciting events, the frequency of private interactions, the ebb and flow of people from the side lines into active participation, and the pace of the community's overall evolution). The rhythm of a community is the strongest indicator of its being alive.

In many ways, centres and/or departments of medical education, as well as the newer forms of academies [78–81], can offer a setting for communities of practice to develop. They can also help nurture and support both new and experienced teachers and educators. These centres and academies, which have been described in the literature and are increasing in number, can also help develop and sustain medical education as an academic discipline [38]. Critical to their success is a common purpose, open communication, and opportunities for dialogue, guidance, and institutional support.

Mentorship

Mentoring is a common strategy to promote the development, socialisation, and maturation of academic medical faculty [69, 82, 83]. However, although several formal mentorship programmes have been described [84, 85], this is an under-utilised approach for staff development, especially with regard to academic and career development. Daloz [86] describes a mentorship model that balances three key elements: support, challenge, and a vision of the individual's future career. This model can also serve as a helpful framework in staff development. Given the importance of mentoring for socialisation into the profession, the development of meaningful academic activities, career satisfaction, and the development of close collaborative relationships [85], we must work to promote and recognise mentorship as a way of developing medical educators.

Role Modelling

Role modelling is instrumental in the development of all medical educators' roles, although it is not usually recognised as such. In our setting, educators have commented on the value of mentors and the importance of role modelling in their formation: 'Medical education involves risks. Without the support of my mentors and role models, I would not have had the courage to accomplish what I did' [13].

Learning from role models occurs through observation and reflection and is a complex mix of conscious and unconscious activities [87]. Whilst we are all aware of the conscious observation of observed behaviours, understanding the power of the unconscious component is essential to effective role modelling. We should also remember that role models differ from mentors [88]. Role models inspire and teach by example – often whilst they are doing other things; mentors have an explicit relationship with a colleague over time [89].

Organisational Support and Development

A survey of faculty members' needs for faculty development [91] highlighted the necessity to look at staff development as 'development, orientation, and/or support'. Interestingly, most programmes focus on the 'development' part. Much less has been written about faculty orientation and/or support.

Support for medical educators can take different forms, including managerial and organisational support, provision of information, recognition of teaching excellence, and consideration of educational scholarship in promotion and tenure. As an example, support systems and materials are available in various areas of medical education [92] and range from textbooks tailored to the needs of those responsible for training doctors [93] to flexible and open learning materials and resources. Organisational support also includes the following:

- the development of institutional policies that support and reward excellence in teaching [11]
- a re-examination of the criteria for academic promotion and increased credit for educational initiatives [94]
- an increase in training and mentoring programmes
- enhanced resources for training teachers and junior faculty members.

In my own setting, the need for an orientation programme for 'new' faculty members has been expressed as follows: 'You have to know how to navigate the system and there are certain expectations and if somebody doesn't make that explicit to you, it's very hard to figure out …' [95]. As staff developers and medical educators, we often tend to focus on the individual teacher and overlook the importance of organisational support and development [15, 16].

Students and Residents as Teachers

Although the primary focus of this chapter addresses the development of faculty members as medical educators, many authors have expressed the view that staff development should start at entry to medical school [96]. As frequently observed, medical students teach in a variety of settings and participate regularly in peer-assisted learning [97]. Residents also play a critical role in the teaching of other residents and students [98], and, in fact, it has been estimated that residents spend as much as 25% of their time in teaching activities, including the supervision, instruction, and evaluation of medical students and more junior residents [99]. At the same time, residents have identified teaching as an important part of their responsibilities [100] and value learning about their educational roles [101].

In examining the role of undergraduate medical students, Dandavino et al. [96] outlined a number of reasons why medical students should learn about teaching. Medical students will become future residents and faculty members, and many of them will take on significant teaching roles. In addition, teaching is an essential component of the doctor–patient relationship and it is hypothesised that medical students will become more efficient communicators as a result of teacher training. It is also hoped that medical students will become better learners as a result of increased knowledge about teaching and learning. Studies have indicated that students are enthusiastic about their role in education [102], interested in learning about teaching [103], and often rated as effective teachers [104]. Similar observations have been made about residents. For example, studies have shown that residents contribute significantly to the education of medical students [98, 105], and that medical students perceive them as playing a critical role in their training [106, 107], especially as residents are closer to the students' experiences and able to draw upon their own teaching and learning practices. Some have also wondered whether improved teaching can increase residents' clinical competence [108, 109], but this question merits further investigation.

Within the context of promoting students and residents as teachers, a number of programmes have been described [110–114]. Some of these programmes include short workshops and seminars for students. For example, Bardach et al. [115] described four one-hour 'How to Teach' sessions for final-year students, whereas Nestel and Kidd [116] reported on the evaluation of a workshop designed to prepare students for their role as peer tutors. In both examples, students believed that formal instruction in teaching should become a required part of their experience; they also felt that they could use what they had learnt in their educational contexts. A number of schools also offer student-as-teacher programmes [111], whereas others have described elective activities in medical education, usually with a small number of students [117]. Elective experiences range in duration from one to four weeks and most students reported that the elective had imparted valuable knowledge and skills. Interestingly, however, very few medical schools seem to incorporate teaching skills into their undergraduate curricula in a systematic (or routine) fashion [102]. At the same time, many institutions offer opportunities for students and residents to serve as teachers through peer and near-peer teaching. For example, students can assume the role of tutors in student-led programmes [118, 119] or teaching assistants in specific courses [120]. As Peluso and Hafler [102] have said, peer and near-peer teaching opportunities are valuable in allowing students to develop educational sessions, organise tutoring programmes, or provide discrete teaching opportunities to more junior colleagues. However, they are limited by 'the confines of a preexisting curriculum and minimal control over the timing, content, and format of the teaching activities' [102].

A variety of resident-as-teacher programmes have also been described. The format of these activities mirrors many of those designed for students and faculty members, and include workshops, seminars, and teaching retreats [98]. The value of integrated curricula to improve teaching skills [121], and the role of chief resident as preparation for becoming a medical educator [122], has also been noted. The majority of available programmes are rated positively by residents, with noted changes in knowledge and attitudes towards teaching [98]. In addition, residents value the experiential nature of activities, the role of feedback and support for their roles as teachers, and the learning that occurs 'on the job'.

Looking forward, staff development programmes for students and residents should be considered an essential component of core curricula and include content about adult learning principles, as well as the broad array of teaching techniques that can be used during case presentations, formal lectures, and informal team discussions [96]. They should also be evaluated in a rigorous and systematic fashion. In addition, we should remember that medical students and residents can gain valuable experience as future educators from their lived experiences, including participation in peer-teaching, course design, relevant committees, and ongoing educational research activities [102]. Early and intentional exposure can help students and residents better appreciate the role of physicians as teachers and perhaps even encourage a career as a medical educator.

Although most of the relevant literature focuses on students and residents as teachers, the need to prepare learners for leadership and management roles has been highlighted by several authors [123, 124]. As Ackerly et al. stated: 'The active cultivation of future leaders is [urgently] required' [125]. Although the specific focus of described programmes does not specifically address leadership for medical educators, the learning formats resemble those of faculty members with an emphasis on the value of experiential learning, the role of mentorship, and the benefit of longitudinal projects. In addition, special pathways [126, 127] and joint MD-MBA programmes [128, 129] have been created to promote leadership development at an early stage and, irrespective of the programme, a skill-based approach that exposes learners to different career paths of physician leaders is encouraged [126]. It should also be noted that leadership (or management) is included as a core competency in different educational frameworks [130, 131] and, as a result, training in this area aligns well with both curricular expectations and health care needs. The same can be said of learners' roles as researchers and scholars. The importance of developing students and residents as researchers has also been highlighted in the literature, although integrating research skills into the role of becoming a future medical educator is more limited. Nonetheless, some valuable examples can be identified, including a research rotation for internal medicine residents [132], the completion of a scholarly project during residency [133], participation in research ethics boards for undergraduate students [134], and student electives in medical education [113, 117]. As is the case with faculty members, faculty development in this area must address the multiple roles that a medical educator can play.

Programme Effectiveness

Research on the impact of staff development activities has shown that overall satisfaction with programmes is high and that participants recommend these activities to their colleagues. Teachers also report a positive change in attitudes towards teaching, as well as self-reported changes in knowledge about educational principles and specific teaching behaviours [39, 40]. Other benefits include increased personal interest and enthusiasm, improved self-confidence, a greater sense of belonging to a community, and enhanced educational leadership and innovation. In our setting, participants in staff development activities have commented on the value of meeting like-minded colleagues and feeling a renewed sense of commitment and enthusiasm about teaching. As one participant reflected: 'I leave rejuvenated and ready to go out and teach a thousand students again!' [57]. Others have identified the value of conceptual frameworks for what they do intuitively, as illustrated in the following quote: 'I was given new tools to teach. Not only were they described to me in words, but they were also used in front of me and I was part and parcel of the demonstration' [57].

In 2006, as part of the Best Evidence in Medical Education (BEME) collaboration (Box 36.4), an international group of medical educators systematically reviewed the faculty development literature to ascertain the impact of faculty development initiatives on teaching effectiveness in medical education [39]. This systematic review, which was updated in 2016 [40], indicated that the majority of interventions targeted practising clinicians and included workshops, short courses, seminar series, longitudinal programmes, and 'other' interventions, such as peer coaching, augmented feedback, online modules, and site visits. These reviews also suggested that we need to broaden the focus beyond individual teaching effectiveness, develop programmes that extend over time, promote workplace learning, foster community development, and secure institutional support for faculty development.

At the same time, despite numerous descriptions of staff development programmes, there has been a paucity of research demonstrating the effectiveness of most faculty development activities [5, 39, 40]. Relatively few programmes have conducted comprehensive evaluations to ascertain what effect the programme is having on faculty members, and data to support the efficacy of these initiatives have often been lacking. Of the studies that have been conducted in this area, most have relied on the assessment of participant satisfaction; some have assessed the impact on cognitive learning or performance, and several others have examined the long-term impact of these interventions. However, moving forward it will be important to conduct more qualitative and mixed methods studies, assess behavioural and organisational change, evaluate transfer to practice, and explore the role of faculty development with the larger organisational context [40]. There is clearly a need for more rigorous research designs to capture the complexity of faculty development interventions. The use of newer methods of performance-based assessment, incorporating diverse data sources, is also indicated, as is the value of process-oriented studies comparing different faculty development strategies and the maintenance of change over time.

Theoretical Frameworks to Guide Faculty Development

MacDougall and Drummond [135] observed that there is no clear theoretical framework to describe how medical teachers and educators develop. In fact, despite an emphasis on educational 'know how' and practice, theory is noticeably absent from the staff development literature. In common with medical education itself, it would seem that a number of educational theories can be applied to faculty development and the development of medical educators. Many of these theoretical perspectives are covered in detail elsewhere in this book, notably in Chapter 2; however, three theoretical frameworks are particularly relevant to faculty development: situated learning [136], Knowles' principles of adult learning [137], and Kolb and Fry's experiential learning cycle [138].

BOX 36.4 WHERE'S THE EVIDENCE: Results of a systematic review

The BEME reviews of faculty development initiatives designed to improve teaching effectiveness in medical education [39, 40] reported the following outcomes:

- *High satisfaction with staff development programmes*

  Participants' overall satisfaction with staff development programmes was high. Moreover, they consistently found programmes to be acceptable, useful, and relevant to their objectives. They also valued the methods used, especially those with a practical and skills-based focus.
- *Changes in attitudes towards teaching and staff development*

  Participants reported positive changes in attitudes towards staff development and teaching as a result of their involvement in these activities. They cited a greater awareness of personal strengths and limitations, increased motivation and enthusiasm for teaching and learning, and a notable appreciation of the benefits of professional development.
- *Gains in knowledge and skills*

  Participants reported increased knowledge of educational principles and strategies as well as gains in specific teaching strategies and skills (e.g. instructional design, effective feedback). Where formal tests of knowledge were used, significant gains were shown.
- *Changes in behaviour*

  Self-perceived changes in teaching behaviours were consistently reported. While student evaluations did not always reflect participants' perceptions, observed changes in teaching performance were also detected. Observed changes that extended beyond teaching in the classroom or clinical setting included the design and delivery of new educational initiatives, new educational responsibilities or leadership positions, increased academic output and productivity, and career advancement.
- *Changes in organisational practice and student learning*

  Although changes in organisational practice and student learning were not frequently investigated, changes did include greater educational involvement and establishment of collegial networks, especially in longitudinal programmes.

Situated Learning

Situated learning is based on the notion that knowledge is contextually situated and fundamentally influenced by the activity, context, and culture in which it is used [136]. This view of knowledge as situated in authentic contexts has important implications for our understanding of staff development and the design and delivery of instructional activities for faculty members. Situated learning theory brings together the cognitive base and experiential learning that is needed to facilitate the acquisition of new behaviours. That is, it bridges the gap between the 'know what' and the 'know how' of teaching and learning by embedding learning in authentic activities. It also helps transform knowledge from the abstract and theoretical to the usable and useful [139]. The proponents of situated learning suggest that there should be a balance between the explicit teaching of a subject and the activities in which the knowledge learnt is used in an authentic context – both essential principles in staff development.

Some of the key components of situated learning include 'cognitive apprenticeship' [140], collaborative learning [136], reflection [141–144], deliberate practice, and articulation of learning skills [145], outlined in greater detail in Chapters 2 and 12. All of these elements can play a critical role in faculty development programming as well.

Closely tied to the notion of situated learning is the concept of 'legitimate peripheral participation' [75]. This social practice, which combines 'learning by doing' (also known as 'experiential learning') and apprenticeship into a single theoretical perspective, is the process by which a novice becomes an expert. That is, from a situated learning perspective, learners build new knowledge and understanding through gradual participation in the community of which they are becoming a part. In many ways, teachers go through this process as well. A key element of participation in the community is the opportunity to see and participate in the framing of problems and understand how knowledge is structured. According to Wenger [77], social participation within the community is the key to informal learning. It is also embedded in the practices and relationships of the workplace and helps create identity and meaning. In addition, it complements, and can substitute for, formal learning mechanisms. Informal learning is often not acknowledged as learning within organisations; rather, it is typically regarded as being 'part of the job'. However, learning at work is a key component of the development of medical educators, and there is value in rendering this learning as visible as possible so that it can be valued as an important component of staff development.

Principles of Adult Learning

Although some have argued that adult learning is not a theory [146] but merely a description of the adult learner, others believe that Knowles' principles of adult learning, also referred to as andragogy [137], form an important theoretical construct [147]. In either case, andragogy provides useful guidance for planning staff development programmes. Its key principles include the observation that adults come to learning situations with diverse motivations and expectations about learning goals and teaching methods, that much of adult learning is 'relearning' rather than new learning, and that most adults prefer to learn through experience. Clearly, the incorporation of these principles into the design of any educational programme will enhance receptivity, relevance, and engagement. An understanding of these principles can also influence pacing, meaning, and motivation.

Experiential Learning

Kolb and Fry [138] provide a description of the learning cycle that highlights the role of experience in the learning process. More specifically, they describe how experience is translated into concepts, which in turn guide the choice of new experiences [148]. In this model, which should be considered in the design of all instructional events, learning is viewed as a four-stage cycle and all learners need the ability to experience diverse situations (in both the classroom and the clinical setting), observe and reflect on what they have learnt (often in a large group session), develop their own theory and understanding of the world, and experiment new ways of being in order for learning to occur [138]. Importantly, attention to the experiential learning cycle will facilitate teaching *and* learning and ensure that different learning styles are respected and nurtured in any staff development programme.

Designing a Staff Development Programme

As formal staff development programmes are one of the most common ways of developing medical educators, this section will highlight some general guidelines that have been previously described and can be of use to staff developers [14].

Understand the Organisational Culture

Staff development programmes take place within the context of a specific institution or organisation. It is imperative to understand the culture of that institution and to be responsive to its needs. Staff development programmes should also capitalise on the organisation's strengths and work with the leadership to ensure success. In many ways, the cultural context can be used to promote or enhance staff development efforts. For example, as some authors have suggested, 'staff development during times of educational or curricular reform can take on added importance' [149]. It is also important to assess institutional support for staff development activities, ascertain available resources, and lobby effectively. Clearly, staff development cannot occur in a vacuum [14].

Determine Appropriate Goals and Priorities

As with the design of any programme, it is imperative to define goals and priorities carefully. What are we trying to achieve – and why is it important to do so? It is equally important to determine programme objectives as they will influence our target audience, choice of programme, overall content, and methodology. Determining priorities is not always easy, and it often involves consultations with key stakeholders. However, it is always essential to balance individual *and* organisational needs.

Conduct Needs Assessments to Ensure Relevant Programming

As stated earlier, staff development programmes should base themselves on the needs of the individual as well as the institution. Student needs, patient needs, and societal needs may also help direct relevant activities [14]. Assessing needs is necessary to refine goals, determine content, identify preferred learning formats, and assure relevance. It is also a way of promoting early 'buy-in'. Common methods include:

- written questionnaires or surveys
- interviews or focus groups with key informants (e.g. participants, students, educational leaders)
- observations of teachers 'in action'
- literature reviews
- environmental scans of available programmes and resources [150, 151].

Whenever possible, we should try to gain information from multiple sources and distinguish between 'needs' and 'wants'. Clearly, an individual teacher's perceived needs may differ from those expressed by their students or peers. Needs assessments can also help further translate goals into objectives, which will serve as the basis for programme planning and evaluation of outcome.

Develop Different Programmes to Accommodate Diverse Needs

Different staff development formats have been described in an earlier section. Clearly, we must design programmes that accommodate diverse goals and objectives, content areas, and the needs of the individual and the organisation. For example, if our goal is to improve our colleagues' lecturing skills, a half-day workshop on interactive lecturing might be the programme of choice. On the other hand, if we wish to promote educational leadership and scholarly activity among our peers, a teaching scholars programme or educational fellowship might be the preferred method [14]. In this context, it is also helpful to remember that staff *development* can include development, orientation, recognition, and support, and different programmes are required to accommodate diverse objectives.

Incorporate Principles of Adult Learning and Instructional Design

As previously mentioned, adults come to learning situations with a variety of motivations and expectations about teaching methods and goals. Incorporation of these principles into the design of a staff development programme is clearly needed, as physicians demonstrate a high degree of self-direction and possess numerous experiences that should serve as the basis for learning [14].

Principles of instructional design should also be followed. For example, it is important to develop clear learning goals and objectives, identify key content areas, design appropriate teaching and learning strategies, and create appropriate methods of evaluation of both the students and the curriculum. It is equally important to integrate theory with practice and to ensure that the learning is perceived as relevant to the work setting and to the profession. Learning should be interactive, participatory, and experientially based, using the participants' previous learning and experience as a starting point. Detailed planning and organisation involving all stakeholders is critical, as is the creation of a positive learning environment. However, although theory should inform practice, staff development initiatives must remain relevant and practical (see Box 36.5).

Offer a Range of Diverse Educational Methods

In line with principles of adult learning, staff development programmes should try to offer a variety of educational methods that promote experiential learning, reflection, feedback, and immediacy of application. Common learning methods include interactive lectures, case presentations, small-group discussions and individual exercises, role-plays and simulations, videotape reviews, and live demonstrations. Practice with feedback is also key, as is the opportunity to reflect on personal values and attitudes. Computer-aided instruction, debates and reaction panels, journal clubs, and self-directed readings are additional methods to consider. In line with our previous example, a workshop on interactive lecturing might include interactive plenary presentations, small group discussions and exercises, and opportunities for practice and feedback. A fellowship programme might include group seminars, independent projects, and structured readings. Whatever the method, the needs and learning preferences of the participants should be respected, and the method should match the objective. Health care professionals learn best 'by doing', and experiential learning should be promoted whenever possible [14].

Promote Buy-in and Market Effectively

The decision to participate in a staff development programme or activity is not as simple as it might at first appear. It involves the individual's reaction to a particular offering, motivation to develop or enhance a specific skill, being available at the time of the session, and overcoming the psychological barrier of admitting need [149]. As faculty developers, it is our challenge to overcome reluctance and to market our 'product' in such a way that resistance becomes a resource to learning. In our context, we have seen the value of targeted mailings, professionally designed promotional materials, and 'branding' of our product to promote interest. Continuing education credits, as well as free and flexible programming, can also help facilitate motivation and attendance. Buy-in involves agreement on importance, widespread support, and dedication of time and resources at both the individual and the system level, and must be considered in all programming initiatives [152].

Work to Overcome Commonly Encountered Challenges

Common challenges faced by faculty developers include the following:

- lack of institutional support and resources for programme planning

BOX 36.5 WHERE'S THE EVIDENCE: Key features of staff development

The BEME reviews of faculty development initiatives designed to improve teaching effectiveness in medical education [39, 40] found that the following 'key features' contribute to the effectiveness of staff development activities.

- *Evidence-informed educational design*

  This refers to the inclusion of a theoretical or conceptual framework to inform programme design, adherence to principles of instructional design and adult learning to promote skill development, and the use of multiple instructional methods to accommodate different learning styles and achieve intended outcomes
- *Relevant content*

  Perceived relevance (and applicability) of the programme's content to participants' educational and clinical responsibilities is key in encouraging faculty members to be involved in faculty development
- *Experiential learning*

  This includes the importance of applying what has been learned, practising skills, and receiving feedback on performance.
- *Opportunities for feedback and reflection*

  The role of feedback (from peers and colleagues) is critically important in promoting change, as is the opportunity to reflect on educational practices, values, and beliefs.
- *Educational projects*

  This feature, which is most common in programmes that extend over time, allows participants to focus on issues that are important to them and apply what they have learned in the workplace.
- *Intentional community building*

  This includes the provision of a safe and supportive learning environment, explicit encouragement of collaboration and networking, and the facilitation of effective peer and colleague relationships, both during and after the faculty development intervention.
- *Longitudinal programme design*

  Longitudinal programmes, which appear to be associated with other design features such as opportunities for practice and application, feedback and reflection, and relationship building and networking, tend to promote sustained change over time.
- *Institutional support*

  Human resources for programme design and delivery, as well as financial support of programmes and participants, are critical to programmatic success.

- conflicting goals and balancing individual and organisational priorities
- assessing needs
- motivating faculty to participate
- obtaining faculty buy-in
- promoting a 'culture change' that reflects the importance of faculty development and a renewed interest in teaching and learning.

Staff developers must inevitably work to overcome these problems through a variety of means that can include creative programming, skilled marketing, and the delivery of high-quality activities. Flexible scheduling and collaborative programming, which address clearly identified needs, can also help ensure success at the system level.

Prepare Staff Developers

The recruitment and preparation of staff developers is rarely reported. However, it is important to recruit carefully, train effectively, partner creatively, and build on previous experiences [152]. Medical educators can be involved in a number of ways: as co-facilitators, as programme planners, or as consultants. In our own setting, we try to involve new faculty members in each staff development activity and conduct a preparatory meeting (or 'dry run') to review content and process, solicit feedback, and promote 'ownership'. We also conclude each activity with a 'debriefing' session to discuss lessons learnt and plan for the future. Whenever possible, staff developers should be individuals who are well respected by their peers and have some educational expertise and experience in facilitating groups. It has been said that 'to teach is to learn twice'; this principle is clearly one of the main motivating factors for staff developers.

Evaluate and Demonstrate Effectiveness

The need to evaluate staff development programmes and activities is clear. In fact, we must remember that the evaluation of staff development is more than an academic exercise, and our findings must be used in the design, delivery, and marketing of our programmes. It has also been stated earlier that staff development must strive to promote education as a scholarly activity; we must role model this approach in all that we do.

In preparing to evaluate a staff development programme or activity, we should consider the goal of the evaluation (e.g. programme planning versus decision-making, policy formation versus academic inquiry), available data sources (e.g. participants, peers, students, or residents), common methods of evaluation (e.g. questionnaires, focus groups, objective tests, observations), resources to support assessment (e.g. institutional support, research grants), and models of programme evaluation (e.g. goal attainment, decision facilitation). Kirkpatrick and Kirkpatrick's levels of evaluation [153] are also helpful in conceptualising and framing the assessment of outcome. They include the following:

- *reaction* – participants' views on the learning experience
- *learning* – change in participants' attitudes, knowledge, or skills
- *behaviour* – change in participants' behaviour
- *results* – changes in the organisational system, the patient, or the learner.

At a minimum, a practical and feasible evaluation should include an assessment of utility and relevance, content, teaching and learning methods, and intent to change. Moreover, as evaluation is an integral part of programme planning, it should be conceptualised at the beginning of any programme. It should also include qualitative and quantitative assessments of learning and behaviour change using a variety of methods and data sources. For more on evaluation methods see Chapter 30 in this book.

Conclusion

As the demands for accountability in higher education gain momentum, pressures to change professional conduct in medical education will continue to grow [11]. Moreover, as the emphasis on global standards in medical education increases [10, 154], so will the need for the professional development of medical educators. As Glicken and Merenstein [18] aptly state, faculty members often come to medical teaching with the 'wisdom and experience that dictates what their students need to know', although they have not been trained for the job at hand. Clearly, it is our responsibility to enable this training, either through formal or informal approaches.

We also need to remember that medical education is a social endeavour. Faculty members at McGill University have identified core attributes (including reflection, passion, enthusiasm, and pride) and skills (such as the ability to maintain multiple perspectives, situate learning, work with others, and see the 'big picture') needed to become a medical educator [13]. They have also highlighted the benefit of a community of scholars in the formation of medical educators (as outlined in Box 36.2). In many ways, these suggestions can serve as a road map for developing medical educators, as each of us finds joy and satisfaction in this journey of discovery. As two colleagues described: 'To me a medical educator is someone who devotes part of their career and time and interests towards medical teaching and education. It is not the title that counts. It is more the effort and the amount of time and the process and the products that get produced that defines the level of medical educator we become' [13]. 'You become a medical educator once you add the intangible – the passion and the commitment, the dedication and the creativity …' [13].

Acknowledgements

I would like to thank Drs Peter McLeod, Linda Snell, and Peter Cantillon for their helpful comments on an earlier version of this chapter. I would also like to gratefully acknowledge my colleagues at the Centre for Medical Education and the Faculty of Medicine at McGill University for their continued participation, insights, and feedback on our faculty development programmes and activities. My own understanding of this complex process would not have been possible without this 'community of practice'. Some of the material used in this chapter has been published previously in a contribution by the author to *A Practical Guide for Medical Teachers* [14] and to *Teaching Medical Professionalism* [155] and is reproduced here with permission.

References

1 Miller, G. (1980). *Educating Medical Teachers*. Cambridge, MA: Harvard University Press.
2 Jason, H. and Westberg, J. (1982). *Teachers and Teaching in US Medical Schools*. New York: Appleton-Century-Crofts.
3 Harris, D., Krause, K., Parish, D., and Smith, M. (2007). Academic competencies for medical faculty. *Family Medicine* 39: 343–350.
4 Clark, J.M., Houston, T.K., Kolodner, K. et al. (2004). Teaching the teachers: national survey of faculty development in departments of medicine of US teaching hospitals. *Journal of General Internal Medicine* 19: 205–214.
5 Skeff, K.M., Stratos, G.A., and Mount, J.F.S. (2007). Faculty development in medicine: a field in evolution. *Teaching and Teacher Education* 23: 280–285.
6 McLean, M., Cilliers, F. and Van Wyk, J.M. (2008). Faculty development: yesterday, today and tomorrow. *Medical Teacher* 30: 555–84.
7 Steinert, Y. ed. (2014). *Faculty Development in the Health Professions: A Focus on Research and Practice*. Dordrecht: Springer.
8 General Medical Council (2012). Good Medical Practice. http://www.gmc-uk.org/education/10264.asp (accessed 20 March 2017).
9 International Association of Medical Colleges (2012). Composite Standards. http://www.iaomc.org/cs.htm (accessed 20 March 2017).
10 World Federation for Medical Education (2007). *Global Standards Programme*. www.wfme.org (accessed 20 March 2017).
11 Eitel, F., Kanz, K.G., and Tesche, A. (2000). Training and certification of teachers and trainers: the professionalization of medical education. *Medical Teacher* 22: 517–526.
12 Purcell, N. and Lloyd-Jones, G. (2003). Standards for medical educators. *Medical Education* 37: 149–154.
13 Steinert, Y. (2008). From teacher to medical educator: the spectrum of medical education. Unpublished report, Centre for Medical Education, Montreal.
14 Steinert, Y. (2009). Staff development. In: *A Practical Guide for Medical Teachers* (ed. J. Dent and R. Harden), 369–376. Edinburgh: Elsevier Churchill Livingstone.
15 Steinert, Y., Cruess, R.L., Cruess, S.R. et al. (2007). Faculty development as an instrument of change: a case study on teaching professionalism. *Academic Medicine* 82: 1057–1064.
16 Jolly, B. (2014). Faculty development for organizational change. In: *Faculty Development in the Health Professions: A Focus on Research and Practice* (ed. Y. Steinert), 119–137. Dordrecht: Springer.
17 Hafferty, F.W. (1998). Beyond curriculum reform: confronting medicine's hidden curriculum. *Academic Medicine* 73: 403–407.
18 Glicken, A.D. and Merenstein, G.B. (2007). Addressing the hidden curriculum: understanding educator professionalism. *Medical Teacher* 29: 54–57.
19 Bligh, J. (2005). Faculty development. *Medical Education* 39: 120–121.

20 Swanwick, T. (2008). See one, do one, then what? Faculty development in postgraduate medical education. *Postgraduate Medical Journal* 84: 339–343.
21 Wilkerson, L. and Irby, D.M. (1998). Strategies for improving teaching practices: a comprehensive approach to faculty development. *Academic Medicine* 73: 387–396.
22 Milner, R.J., Gusic, M.E., and Thorndyke, L.E. (2011). Perspective: toward a competency framework for faculty. *Academic Medicine* 86: 1204–1210.
23 Molenaar, W.M., Zanting, A., van Beukelen, P. et al. (2009). A framework for teaching competencies across the medical education continuum. *Medical Teacher* 31: 390–396.
24 Srinivasan, M., Li, S.T., Meyers, F.J. et al. (2011). "Teaching as a competency": competencies for medical educators. *Academic Medicine* 86: 1211–1220.
25 Academy of Medical Educators (2012) Professional standards. http://www.medicaleducators.org (accessed 22 June 2018).
26 Buchel, T.L. and Edwards, F.D. (2005). Characteristics of effective clinical teachers. *Family Medicine* 37: 30–35.
27 Wright, S.M. and Carrese, J.A. (2002). Excellence in role modeling: insight and perspectives from the pros. *Canadian Medical Association Journal* 167: 638–643.
28 Bordage, G., Foley, R., and Goldyn, S. (2000). Skills and attributes of directors of educational programmes. *Medical Education* 34: 206–210.
29 Spencer, J. and Jordan, R. (2001). Educational outcomes and leadership to meet the needs of modern health care. *Quality in Health Care* 10: ii38–ii45.
30 Boyer, E. (1990). *Scholarship Reconsidered: Priorities of the Professoriate*. Princeton, NJ: Princeton University Press.
31 Simpson, D.E. and Fincher, R.M. (1999). Making a case for the teaching scholar. *Academic Medicine* 74: 1296–1299.
32 Steinert, Y. and Snell, L. (2011). Educational innovation and scholarship: from curriculum design to implementation. In: *Educational Design: A CanMEDS Guide for the Health Professions* (ed. J. Sherbino and J.R. Frank). Ottawa: Royal College of Physicians and Surgeons of Canada.
33 Glassick, C.E. (2000). Boyer's expanded definitions of scholarship, the standards for assessing scholarship and the elusiveness of the scholarship of teaching. *Academic Medicine* 75: 877–880.
34 Higgs, J. and McAllister, L. (2007). Educating clinical educators: using a model of the experience of being a clinical educator. *Medical Teacher* 29: e51–e57.
35 Steinert, Y. and Macdonald, M.E. (2015). Why physicians teach: giving back by paying it forward. *Medical Education* 49: 773–782.
36 Fraser, S.W. and Greenhalgh, T. (2001). Coping with complexity: educating for capability. *British Medical Journal* 323: 799–803.
37 Dearing, R. (1997). *Higher Education in the Learning Society: report of the National Committee of Inquiry into Higher Education*. London: HMSO.
38 Walsh, K. (2007). Professionalism in medical education: could an academy play a role? *Medical Teacher* 29: 425–426.
39 Steinert, Y., Mann, K., Centeno, A. et al. (2006). A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME guide no. 8. *Medical Teacher* 28: 497–526.
40 Steinert, Y., Mann, K., Anderson, B. et al. (2016). A systematic review of faculty development initiatives designed to enhance teaching effectiveness: a 10-year update: BEME guide no. 40. *Medical Teacher* 38: 769–786.
41 Steinert, Y. (2010). Becoming a better teacher: from intuition to intent. In: *Theory and Practice of Teaching Medicine* (ed. J. Ende), 73–93. Philadelphia: American College of Physicians.
42 Nasmith, L. and Steinert, Y. (2001). The evaluation of a workshop to promote interactive lecturing. *Teaching and Learning in Medicine* 13: 43–48.
43 Nasmith, L., Steinert, Y., Saroyan, A. et al. (1997). Assessing the impact of a faculty development workshop: a methodological study. *Teaching and Learning in Medicine* 9: 209–214.
44 Nayer, M. (1995). Faculty development for problem-based learning programs. *Teaching and Learning in Medicine* 7: 138–148.
45 Olmesdahl, P.J. and Manning, D.M. (1999). Impact of training on PBL facilitators. *Medical Education* 33: 753–755.
46 Quirk, M.E., DeWitt, T., Lasser, D. et al. (1998). Evaluation of primary care futures: a faculty development program for community health center preceptors. *Academic Medicine* 73: 705–707.
47 Wilkerson, L. and Sarkin, R.T. (1998). Arrows in the quiver: evaluation of a workshop on ambulatory teaching. *Academic Medicine* 73: s67–s69.
48 Steinert, Y., Nasmith, L., and Daigle, N. (2003). Executive skills for medical faculty: a workshop description and evaluation. *Medical Teacher* 25: 666–668.
49 Steinert, Y., Cruess, S., Cruess, R., and Snell, L. (2005). Faculty development for teaching and evaluating professionalism: from programme design to curriculum change. *Medical Education* 39: 127–136.
50 Baldwin, C.D., Goldblum, R.M., Rassin, D.K., and Levine, H.G. (1994). Facilitating faculty development and research through critical review of grant proposals and articles. *Academic Medicine* 69: 62–64.
51 Steinert, Y., McLeod, P.J., Liben, S., and Snell, L. (2008). Writing for publication in medical education: the benefits of a faculty development workshop and peer writing group. *Medical Teacher* 30: e280–e285.
52 Cusimano, M.D. and David, M.A. (1998). A compendium of higher education opportunities in health professions education. *Academic Medicine* 73: 1255–1259.
53 Gruppen, L.D., Simpson, D., Searle, N.S. et al. (2006). Educational fellowship programs: common themes and overarching issues. *Academic Medicine* 81: 990–994.
54 Gruppen, L.D. (2014). Intensive longitudinal faculty development programs. In: *Faculty Development in the Health Professions: A Focus on Research and Practice* (ed. Y. Steinert), 197–216. Dordrecht: Springer.
55 Frohna, A.Z., Hamstra, S.J., Mullan, P.B., and Gruppen, L. (2006). Teaching medical education principles and methods to faculty using an active learning approach: the University of Michigan Medical Education Scholars Program. *Academic Medicine* 81: 975–978.
56 Burdick, W., Amaral, E., Campos, H., and Norcini, J. (2011). A model for linkage between health professions education and health: FAIMER international faculty development initiatives. *Medical Teacher* 33: 632–637.
57 Steinert, Y., Nasmith, L., McLeod, P.J., and Conochie, L. (2003). A teaching scholars program to develop leaders in medical education. *Academic Medicine* 78: 142–149.
58 Steinert, Y. and McLeod, P.J. (2006). From novice to informed educator: the teaching scholars program for educators in the health sciences. *Academic Medicine* 81: 969–974.
59 Rosenbaum, M.E., Lenoch, S., and Ferguson, K.J. (2005). Outcomes of a teaching scholars program to promote leadership in faculty development. *Teaching and Learning in Medicine* 17: 247–252.
60 Muller, J.H. and Irby, D.M. (2006). Developing educational leaders: the teaching scholars program at the University of California, San Francisco, School of Medicine. *Academic Medicine* 81: 959–964.
61 Steinert, Y., Naismith, L., and Mann, K. (2012). Faculty development initiatives designed to promote leadership in medical education. A BEME systematic review: BEME guide no. 19. *Medical Teacher* 34: 483–503.
62 Wilkerson, L., Uijtdehaage, S., and Relan, A. (2006). Increasing the pool of educational leaders for UCLA. *Academic Medicine* 81: 954–958.

63 Hatem, C.J., Lown, B.A., and Newman, L.R. (2006). The academic health centre coming of age: helping faculty become better teachers and agents of educational change. *Academic Medicine* 81: 941–944.

64 Cohen, R., Murnaghan, L., Collins, J., and Pratt, D. (2005). An update on master's degrees in medical education. *Medical Teacher* 27: 686–692.

65 Tekian, A. and Harris, I. (2012). Preparing health professions education leaders worldwide: a description of masters-level programs. *Medical Teacher* 34: 52–58.

66 Pugsley, L., Brigley, S., Allery, L., and Macdonald, J. (2008). Counting quality because quality counts: differing standards in master's in medical education programmes. *Medical Teacher* 30: 80–85.

67 Hesketh, E.A., Bagnall, G., Buckley, E.G. et al. (2001). A framework for developing excellence as a clinical educator. *Medical Education* 35: 555–564.

68 Flynn, S.P., Bedinghaus, J., Snyder, C., and Hekelman, F. (1994). Peer coaching in clinical teaching: a case report. *Family Medicine* 26: 569–570.

69 Boillat, M. and Elizov, M. (2014). Peer coaching and mentorship. In: *Faculty Development in the Health Professions: A Focus on Research and Practice* (ed. Y. Steinert), 159–179. Dordrecht: Springer.

70 Orlander, J.D., Gupta, M., Fincke, B.G. et al. (2000). Co-teaching: a faculty development strategy. *Medical Education* 34: 257–265.

71 Archer, J., Swanwick, T., Smith, D. et al. (2013). Developing a multisource feedback tool for postgraduate medical educational supervisors. *Medical Teacher* 35: 145–154.

72 Bukhari, A., Mahboob, U., Atiq, S., and Ahmed, J. (2016). Significance of students' feedback on improving the teaching practices of the faculty. *Khyber Medical University Journal* 8: 142–146.

73 Steinert, Y. (2014). Learning from experience: from workplace learning to communities of practice. In: *Faculty Development in the Health Professions: A Focus on Research and Practice* (ed. Y. Steinert), 141–158. Dordrecht: Springer.

74 Boud, D. and Middleton, H. (2003). Learning from others at work: communities of practice and informal learning. *Journal of Workplace Learning* 15: 194–202.

75 Lave, J. and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. New York: Cambridge University Press.

76 Barab, S.A., Barnett, M., and Squire, K. (2002). Developing an empirical account of a community of practice: characterizing the essential tensions. *The Journal of the Learning Sciences* 11: 489–542.

77 Wenger, E. (1998). *Communities of Practice: Learning, Meaning and Identity*. New York: Cambridge University Press.

78 Cooke, M., Irby, D.M., and Debas, H.T. (2003). The UCSF academy of medical educators. *Academic Medicine* 78: 666–672.

79 Irby, D.M., Cooke, M., Lowenstein, D., and Richards, B. (2004). The academy movement: a structural approach to reinvigorating the educational mission. *Academic Medicine* 79: 729–736.

80 Bligh, J. and Brice, J. (2007). The Academy of Medical Educators: a professional home for medical educators in the UK. *Medical Education* 41: 625–627.

81 Searle, N.S., Thompson, B.M., Friedland, J.A. et al. (2010). The prevalence and practice of academies of medical educators: a survey of US medical schools. *Academic Medicine* 85: 48–56.

82 Bligh, J. (1999). Mentoring: an invisible support network. *Medical Education* 33: 2–3.

83 Bland, C., Schmitz, C., Stritter, F. et al. (1990). *Successful Faculty in Academic Medicine: Essential Skills and How to Acquire Them*. New York: Springer.

84 Morzinski, J., Diehr, S., Bower, D.J., and Simpson, D.E. (1996). A descriptive, cross-sectional study of formal mentoring for faculty. *Family Medicine* 28: 434–438.

85 Pololi, L.H., Knight, S.M., Dennis, K., and Frankel, R.M. (2002). Helping medical school faculty realize their dreams: an innovative, collaborative mentoring program. *Academic Medicine* 77: 377–384.

86 Daloz, L. (1986). *Effective Teaching and Mentoring*. San Francisco: Jossey-Bass.

87 Cruess, S.R., Cruess, R.L., and Steinert, Y. (2008). Role modelling: making the most of a powerful teaching strategy. *British Medical Journal* 336: 718–721.

88 Ricer, R.E. (1998). Defining preceptor, mentor, and role model. *Family Medicine* 30: 328.

89 Paice, E., Heard, S., and Moss, F. (2002). How important are role models in making good doctors? *British Medical Journal* 325: 707–710.

90 Wenger, E., McDermott, R., and Snyder, W. (2002). *Cultivating Communities of Practice*. Boston: Harvard Business School Press.

91 Steinert, Y. and Walsh, A. (2006). *A Faculty Development Program for Teachers of International Medical Graduates*. Ottawa: Association of Faculties of Medicine of Canada.

92 Grant, J. (1998). Offering training and support to trainers. In: *Delivering the New Doctor* (ed. E. Paice). Edinburgh: Association for the Study of Medical Education.

93 Jolly, B. and Grant, J. (1997). *The Good Assessment Guide: A Practical Guide to Assessment and Appraisal for Higher Specialist Training*. London: Joint Centre for Education in Medicine.

94 Azer, S.A. (2005). The qualities of a good teacher: how can they be acquired and sustained? *Journal of the Royal Society of Medicine* 98: 67–69.

95 Steinert, Y., McLeod, P., Boillat, M. et al. (2008). Faculty development: a field of dreams? *Medical Education* 43: 42–49.

96 Dandavino, M., Snell, L., and Wiseman, J. (2007). Why medical students should learn how to teach. *Medical Teacher* 29: 558–565.

97 Topping, K.J. (1996). The effectiveness of peer tutoring in further and higher education: a typology and review of the literature. *Higher Education* 32: 321–345.

98 Hill, A.G., Yu, T.C., Barrow, M., and Hattie, J. (2009). A systematic review of resident-as-teacher programmes. *Medical Education* 43: 1129–1140.

99 Seely, A.J.E. (1999). The teaching contributions of residents. *Canadian Medical Association Journal* 161: 1239–1241.

100 Busari, J.O., Prince, K.J., Scherpbier, A.J., van der Vleuten, C.P. and Essed, G.G. (2002). How residents perceive their teaching role in the clinical setting: a qualitative study. *Medical Teacher* 24: 57–61.

101 Thomas, P.S., Harris, P., Rendina, N., and Keogh, G. (2002). Residents as teachers: outcomes of a brief training programme. *Education for Health* 15: 71–78.

102 Peluso, M.J. and Hafler, J.P. (2011). Medical students as medical educators: opportunities for skill development in the absence of formal training programs. *The Yale Journal of Biology and Medicine* 84: 203–209.

103 Bing-You, R.G. and Sproul, M.S. (1992). Medical students' perceptions of themselves and residents as teachers. *Medical Teacher* 14: 133–138.

104 Haist, S.A., Wilson, J.F., Brigham, N.L. et al. (1998). Comparing fourth-year medical students with faculty in the teaching of physical examination skills to first-year students. *Academic Medicine* 73: 198–200.

105 Edwards, J.C., Friedland, J.A., and Bing-You, R. (2002). *Residents' Teaching Skills*. New York: Springer.

106 Sternszus, R., Cruess, S.R., Cruess, R.L. et al. (2012). Residents as role models: impact on undergraduate trainees. *Academic Medicine* 87: 1282–1287.

107 Walton, J.M. and Patel, H. (2008). Residents as teachers in Canadian paediatric training programs: a survey of program director and resident perspectives. *Paediatrics and Child Health* 13: 675–679.

108 Busari, J.O. and Scherpbier, A.J. (2004). Why residents should teach: a literature review. *Journal of Postgraduate Medicine* 50: 205–210.

109 Seely, A.J., Pelletier, M.P., Snell, L.S., and Trudel, J.L. (1999). Do surgical residents rated as better teachers perform better on in-training examinations? *American Journal of Surgery* 177: 33–37.

110 Busari, J.O., Scherpbier, A.J., van der Vleuten, C.P., and Essed, G.G. (2006). A two-day teacher-training programme for

medical residents: investigating the impact on teaching ability. *Advances in Health Sciences Education: Theory and Practice* 11: 133–144.

111 Pasquinelli, L.M. and Greenberg, L.W. (2008). A review of medical school programs that train medical students as teachers (MED-SATS). *Teaching and Learning in Medicine* 20: 73–81.

112 Pasquale, S.J. and Pugnaire, M.P. (2002). Preparing medical students to teach. *Academic Medicine* 77: 1175–1176.

113 Soriano, R.P., Blatt, B., Coplit, L. et al. (2010). Teaching medical students how to teach: a national survey of students-as-teachers programs in U.S. medical schools. *Academic Medicine* 85: 1725–1731.

114 Wamsley, M.A., Julian, K.A., and Wipf, J.E. (2004). A literature review of 'resident-as-teacher' curricula: do teaching courses make a difference? *Journal of General Internal Medicine* 19: 574–581.

115 Bardach, N.S., Vedanthan, R., and Haber, R.J. (2003). 'Teaching to teach': enhancing fourth year medical students' teaching skills. *Medical Education* 37: 1031–1032.

116 Nestel, D. and Kidd, J. (2002). Evaluating a teaching skills workshop for medical students. *Medical Education* 36: 1094–1095.

117 Craig, J.L. and Page, G. (1987). Teaching in medicine: an elective course for third-year students. *Medical Education* 21: 386–390.

118 Schaffer, J.L., Wile, M.Z., and Griggs, R.C. (1990). Students teaching students: a medical school peer tutorial programme. *Medical Education* 24: 336–343.

119 Walker-Bartnick, L.A., Berger, J.H., and Kappelman, M.M. (1984). A model for peer tutoring in the medical school setting. *Journal of Medical Education* 59: 309–315.

120 Evans, D.J. and Cuffe, T. (2009). Near-peer teaching in anatomy: an approach for deeper learning. *Anatomical Sciences Education* 2: 227–233.

121 Morrison, E.H., Rucker, L., Boker, J.R. et al. (2004). The effect of a 13-hour curriculum to improve residents' teaching skills: a randomized trial. *Annals of Internal Medicine* 141: 257–263.

122 Moser, E.M., Kothari, N., and Stagnaro-Green, A. (2008). Chief residents as educators: an effective method of resident development. *Teaching and Learning in Medicine* 20: 323–328.

123 Berkenbosch, L., Bax, M., Scherpbier, A. et al. (2013). How Dutch medical specialists perceive the competencies and training needs of medical residents in healthcare management. *Medical Teacher* 35: e1090–e1102.

124 Blumenthal, D.M., Bernard, K., Bohnen, J., and Bohmer, R. (2012). Addressing the leadership gap in medicine: residents' need for systematic leadership development training. *Academic Medicine* 87: 513–522.

125 Ackerly, D.C., Sangvai, D.G., Udayakumar, K. et al. (2011). Training the next generation of physician-executives: an innovative residency pathway in management and leadership. *Academic Medicine* 86: 575–579.

126 Kuo, A.K., Thyne, S.M., Chen, H.C. et al. (2010). An innovative residency program designed to develop leaders to improve the health of children. *Academic Medicine* 85: 1603–1608.

127 Paller, M.S., Becker, T., Cantor, B., and Freeman, S.L. (2000). Introducing residents to a career in management: the Physician Management Pathway. *Academic Medicine* 75: 761–764.

128 Larson, D.B., Chandler, M., and Forman, H.P. (2003). MD/MBA programs in the United States: evidence of a change in health care leadership. *Academic Medicine* 78: 335–341.

129 Sherrill, W.W. (2000). Dual-degree MD-MBA students: a look at the future of medical leadership. *Academic Medicine* 75: S37–S39.

130 Accreditation Council for Graduate Medical Education (2012). Competency Evaluation System. http://www.acgme.org/acgmeweb (accessed 20 March 2017).

131 Frank, J.R., Snell, L., and Sherbino, J. (2015). CanMEDS 2015 physician competency framework. http://www.royalcollege.ca/rcsite/canmeds/canmeds-framework-e (accessed 20 March 2017).

132 Kanna, B., Deng, C., Erickson, S.N. et al. (2006). The research rotation: competency-based structured and novel approach to research training of internal medicine residents. *BMC Medical Education* 6: 52.

133 Rivera, J.A., Levine, R.B., and Wright, S.M. (2005). Completing a scholarly project during residency training: perspectives of residents who have been successful. *Journal of General Internal Medicine* 20: 366–369.

134 Walton, N.A., Karabanow, A.G., and Saleh, J. (2008). Students as members of university-based Academic Research Ethics Boards: a natural evolution. *Journal of Academic Ethics* 6: 117–127.

135 MacDougall, J. and Drummond, M.J. (2005). The development of medical teachers: an enquiry into the learning histories of 10 experienced medical teachers. *Medical Education* 39: 1213–1220.

136 Brown, J.S., Collins, A., and Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher* 18: 32–42.

137 Knowles, M.S. (1984). *Andragogy in Action*. San Francisco: Jossey-Bass.

138 Kolb, D. and Fry, R. (1975). Towards an applied theory of experiential learning. In: *Theories of Group Processes* (ed. C. Cooper). London: Wiley.

139 Cruess, R.L. and Cruess, S.R. (2006). Teaching professionalism: general principles. *Medical Teacher* 28: 205–208.

140 Mann, K. (2006). Learning and teaching in professional character development. In: *Lost Virtue: Professional Character Development in Medical Education* (ed. N. Kenney). New York: Elsevier.

141 Schön, D. (1983). *The Reflective Practitioner: How Professionals Think in Action*. New York: Basic Books.

142 Hewson, M.G. (1991). Reflection in clinical teaching: an analysis of reflection-on-action and its implications for staffing residents. *Medical Teacher* 13: 227–231.

143 Lachman, N. and Pawlina, W. (2006). Integrating professionalism in early medical education: the theory and application of reflective practice in the anatomy curriculum. *Clinical Anatomy* 19: 456–460.

144 Robertson, K. (2005). Reflection in professional practice and education. *Australian Family Physician* 34: 781–783.

145 McLellan, H. (1996). *Situated Learning Perspectives*. Englewood Cliffs, NJ: Educational Technology Publications.

146 Norman, G.R. (1999). The adult learner: a mythical species. *Academic Medicine* 74: 886–889.

147 Merriam, S.B. (1996). Updating our knowledge of adult learning. *Journal of Continuing Education in the Health Professions* 16: 136–143.

148 Boud, D., Keogh, R., and Walker, D. (1985). *Reflection: Turning Experience into Learning*. London: Kogan Page.

149 Rubeck, R.F. and Witzke, D.B. (1998). Faculty development: a field of dreams. *Academic Medicine* 73: S32–S37.

150 Grant, J. (2002). Learning needs assessment: assessing the need. *British Medical Journal* 324: 156–159.

151 Lockyer, J. (1998). Needs assessment: lessons learned. *Journal of Continuing Education in the Health Professions* 18: 190–192.

152 Steinert, Y. (2005). Staff development for clinical teachers. *The Clinical Teacher* 2: 104–110.

153 Kirkpatrick, D.L. and Kirkpatrick, J.D. (2006). *Evaluating Training Programs: The Four Levels*. San Francisco: Berrett-Koehler Publishers.

154 Karle, H. (2007). European specifications for global standards in medical education. *Medical Education* 41: 924–925.

155 Steinert, Y. (2009). Educational theory and strategies for teaching and learning professionalism. In: *Teaching Medical Professionalism* (ed. R. Cruess, S. Cruess and Y. Steinert), 31–52. New York: Cambridge University Press.

Further Reading and Resources

Steinert, Y. ed. (2014). *Faculty Development in the Health Professions: A Focus on Research and Practice*. Dordrecht: Springer.

www.merlot.org The MERLOT system provides access to curated online learning and support materials and content creation tools, led by an international community of educators, learners and researchers.

www.mededportal.org MedEd PORTAL is a peer-reviewed, open-access journal, that promotes educational scholarship and the dissemination of teaching and assessment resources in the health professions.

37 Educational Leadership and Management

Judy McKimm[1] *and Tim Swanwick*[2,3]
[1]College of Medicine, Swansea University Medical School, Swansea, UK
[2]Health Education England, London, UK
[3]NHS Leadership Academy, Leeds, UK

KEY MESSAGES

- Leadership is a social construct that reflects the preoccupations of the time.
- Leadership and management are inextricably intertwined, and both are essential for organisational success.
- Leadership is about change and movement, management about order and consistency.
- Consideration of followership is essential to an understanding of leadership practice.
- Educational leadership in health operates in complex environments across multiple boundaries.
- Effective leadership requires a deep contextual awareness and the ability to scan the horizon.
- Leadership is about both being and doing; leaders need to 'walk the walk'.

Introduction

Education leaders in medical and other health professions carry a 'double burden'. They have to manage and lead educational endeavours within universities and other educational institutions in a rapidly changing global environment, while working in close collaboration with a range of health care delivery partners with their own financial and service-driven agendas. The complexity and pace of change in health care means that the need for effective leaders and leadership development has never been so important; however, in a sector that is continually focused on the achievement of targets and balancing budgets, training and development are not always given the highest priority.

At the national level, health care education is driven by key governmental health and educational policy initiatives that set the context for educational leaders. These include aspirational policies, such as 'widening access' and expanding higher education, as well as an inexorable drive towards greater public accountability – a problematic area for leaders who work in universities traditionally run on collegial models of participative but autonomous decision-making and consensus.

Bush [1] notes that it is vital for education leaders to maintain 'education' as their central concern. This is difficult to achieve in the health care sector, where there are inherent tensions between the purposes of *education*, the demands of *vocational training* (which has as its goal the production of a competent, safe future workforce), and the demands of delivering a high-quality service to patients and communities. Funding for the different stages and parts of health professions education is often misaligned and, depending on the health and education structure in different countries, derived from different places: the public sector (e.g. government departments), private and voluntary sectors, charities, endowments, and non-governmental organisations.

In many countries, leadership development in the health services has been intrinsically linked to service redesign, improvement, and delivery. For example, in England, leadership in the National Health Service (NHS) is firmly sited in service improvement, and the NHS Leadership Academy (http://www.leadershipacademy.nhs.uk) lays out an underpinning philosophy about leadership development that assumes that leadership skills can be learned, aligns leadership qualities with management capabilities, locates leadership at all levels within the organisation, and aims to create an empowerment culture based on supporting people, the transformation of followers into leaders, and the development and creation of a shared vision [2] (see Figure 37.1).

In higher (or tertiary) education, which educates and trains the medical and health care workforce of the future at pre-registration level, public sector service improvement does not have the same priority, and policy agendas differ. Most higher education strategies focus on delivering a high-quality student experience, ensuring equity of access and outcome, enhancing research outputs, and ensuring graduates are fit for employment or further study. Performance indicators use various metrics to identify and reward universities and colleges for achieving these aims. The inherent tensions and contradictions in leading change in today's public sector organisations are exacerbated in health professions

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.

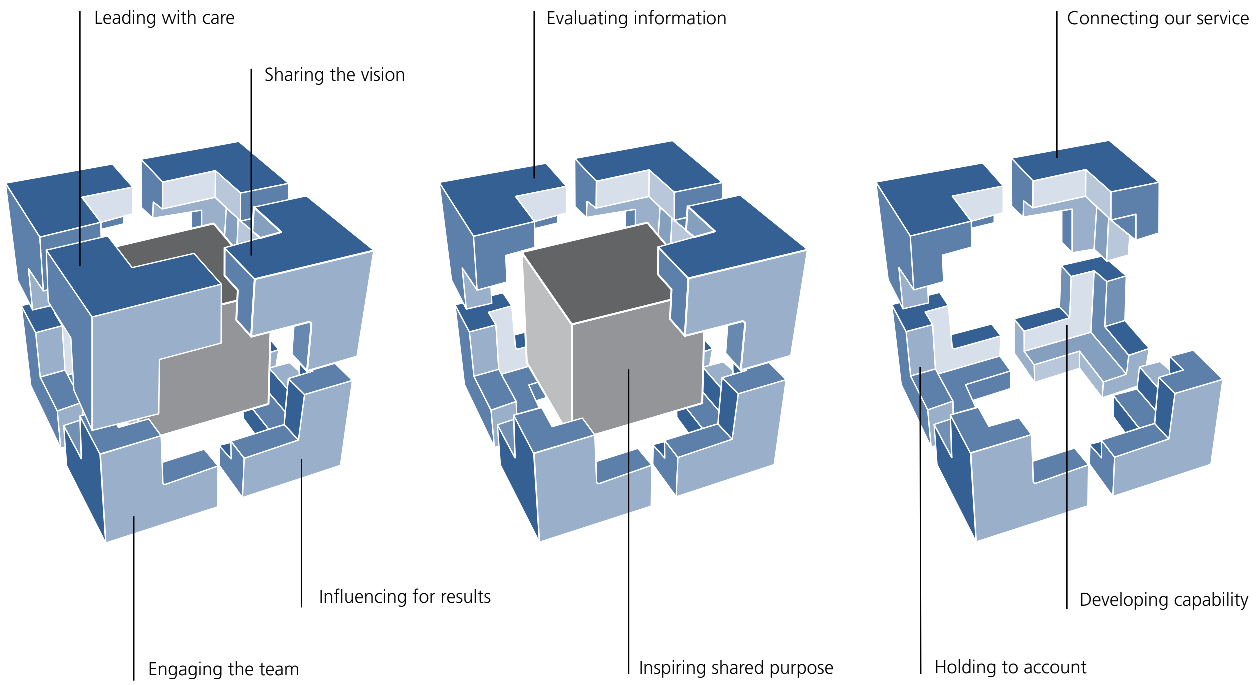


Figure 37.1 Health care Leadership Model. *Source*: The Health care Leadership Model is ©NHS Leadership Academy, 2013 [2]. All rights reserved.

education by differences between health service and education policies, values, systems, and practices.

Another unique feature that faces health care educators is the challenge of leading education across professional boundaries. Leaders in health care education have to be aware of the needs and demands of multiple stakeholders in the current drive to produce competent health care practitioners, and so be open to working in a flexible way. Working with academics, clinicians, and health care managers and professionals with differing values and education systems means that leaders need to quickly establish credibility with different stakeholders, use and develop sound interpersonal skills, and demonstrate effective management capabilities.

Leadership training and development is often patchy and not easy to access for more junior staff. Higher education has traditionally been focused on developing the management skills of heads of departments and more senior management rather than implementing a planned, system-wide approach to succession planning and developing leadership capacity at all levels. Changes in skills mix and professional roles, changes in funding and commissioning arrangements, and emerging organisational changes such as mergers, collaborations, and partnerships will mean that tomorrow's health care education leaders will require a greater range of leadership and management skills than ever before.

Leadership and Management

There are many definitions of leadership. Leadership theory looks not only at individuals but also at the relationship between leaders and followers and the context or situation in which they operate. In recent years, many theorists have found it useful to distinguish between *management* and *leadership*. Although there is considerable overlap and interdependency, differentiating between the two concepts will help us begin to explore what leadership is and some of the assumptions made about, and expectations of, people working in educational settings.

For Bennis and Nanus, 'managers are people who do things right and leaders are people who do the right thing' [3]. Managers are often described as performing *functions* in organisations, usually holding a particular formal title or fulfilling a particular role. Management is concerned with planning, organising, coordinating, commanding, or controlling the activities of staff [4]. Leaders, by contrast, aim to influence and guide others into pursuing particular objectives or visions of the future and to motivate them into wanting to follow. Yukl suggests that 'leadership is the process of influencing others to understand and agree about what needs to be done and how it can be done effectively and the process of facilitating individual and collective efforts to accomplish the shared objectives' [5].

Kotter [6] contrasts leadership and management and suggests that whereas leadership sets a direction and develops a vision for the future, that is, producing change and movement, management is more concerned with planning and with providing order and consistency in organisations rather than the process of producing change. Covey et al. put a slightly different slant on this, suggesting that managers work within an existing 'paradigm', solve problems, and manage people, whereas leaders create new paradigms, challenge systems, seek new opportunities, and lead people [7].

BOX 37.1 Management versus leadership (adapted from Northouse [8])

Management	Leadership
Produces order and consistency through:	*Produces change and movement through:*
• planning and budgeting	• setting direction
• problem solving	• problem defining
• organising and staffing	• building commitment
• controlling and monitoring	• motivating and sustaining

The differences between leadership and management identified by many authors over the years are succinctly summarised by Northouse [8] (see Box 37.1). However, contemporary leadership theory, for example, Blagg and Young [9], Yukl [5], and Storey [10], suggests that in practice it is often the same person who operates in both capacities and that the most effective people in organisations can utilise both management and leadership understanding and skills. Furthermore, in universities, administrators are often distinguished from academics, reflecting the way in which university systems have developed. In this context, the terms 'leaders' and 'managers' are not so widely used, although individuals clearly carry out management and leadership functions. In practice, the barriers and constructs that often worked to place academics and administrators seemingly in opposition to one another are breaking down, as educational institutions are becoming more aware of the need to operate efficiently and effectively in an increasingly competitive global market. Educational leaders can no longer rely wholly on formal positional authority but must seek to develop an ability to lead 'from the middle' of their organisations and influence effectively across increasingly large and complex systems.

Followership

As we have discussed, educational leadership operates within dynamic, rapidly changing complex environments, and recent research and thinking on relational and adaptive leadership highlights the shift from leader-centric theories to those which are inter-relational, focus on followership, and acknowledge systems thinking and networks. As Uhl-Bien et al. note: 'our understanding of leadership is incomplete without an understanding of followership' [11]. Traditional leader-centric thinking tends to link a leader's activities directly to group or organisational outcomes, rather than seeing them as a consequence of a dynamic process between leader and follower or team member. From a leader-centric perspective, followers (like managers) are often described in terms of subordination or derogation and little credence is given to the importance of followership behaviours in facilitating and forming leadership. Followership theory enhances our understanding of leadership by describing the influence that followers have on their leaders, and the process of 'co-creation' that occurs as a result of the impact that followers have on leaders and leadership styles [11]. Followership is an active process, distinct from 'not-leading' or 'not-leading right now', and varies as much as leadership practice, ranging from 'star-followers' who represent an engaged and dynamic positive influence, to those who obstruct team progress and challenge leadership authority [12]. Thus followers can have a very positive or negative powerful impact on the direction of a team and the efficacy of the leadership. Being and becoming a 'leader' and a 'follower' is therefore a fluid process, all followers lead and all leaders follow, depending on the context.

A Brief History of Leadership

The next two sections consider leadership models and theories, charting their development and unpicking the contribution that they make in enhancing our understanding of what leaders do and what leadership is. It is important to remember that leadership is a social construct; theories and models of leadership tend, therefore, to be of their time and reflect the prevailing public 'mood' and the preoccupations of (largely North American) enterprise. Leadership theories and models can be broadly divided into three overlapping categories, as follows:

- those which focus on the personal qualities or personality of the leader as an individual
- those which relate to the interaction of the leader with other people
- those which seek to explain leadership behaviours in relation to the environment or system [13].

In the first half of the twentieth century, leadership theory revolved around personal qualities, an approach based on the assumption that leaders are born rather than made. This gave rise to the pervasive idea of the 'great man' or 'heroic' leader who will be a leader under any circumstances. The 'great man' theories emphasised characteristics such as intelligence, energy, and dominance, but several major reviews of the literature failed to consistently identify personality traits that differentiated leaders from non-leaders. Interestingly, trait theory has made a comeback in recent years with the emotional intelligence theories of Goleman [14] and research on the influence of personality on leadership behaviours [15]. 'Who you are' undoubtedly affects 'how you lead' but the relationship is by no means a simple one (see Box 37.2).

From the 1950s onwards, attention shifted from personal characteristics of leaders to describing how their *style* and behaviours impacted on groups. Blake and Mouton [16], for instance, consider the differences between 'task-focused' and 'people-focused' leaders, and argue that high levels of concern for both is what is required, while authors such as Tannenbaum and Schmidt [17] described a range of leadership styles involving varying degrees of delegated autonomy, from the 'autocratic' to the 'abdicatory'.

While these models introduce the idea of leadership as a group of behaviours, they give little indication as to which

BOX 37.2 FOCUS ON: Personal qualities (adapted from McKimm et al. [32])

The literature identifies a wide range of personal qualities required for leadership, such as integrity, humility, charisma, and the ability to communicate. The majority of the qualities linked with successful leadership can be summarised under three headings: resilience, emotional intelligence, and grit.

- *Resilience* is the ability to 'bounce back' from adversity or challenge [33].
- *Emotional intelligence* (EI) comprises a combination of self-awareness, empathy, social skills, self-motivation, and self-regulation [34].
- *Grit* is a combination of resilience, passion, hard work, perseverance, determination, and direction [35].

To be a 'gritty' leader you have to have a deep interest in what you're doing, take opportunities to practise skills and show self-discipline, cultivate a conviction and purpose about your work (it has to matter), and have hope and confidence that you can *do* leadership as well as *be* a leader.

Alongside these personal qualities, leaders need *credibility* which can be developed as follows:

- Understand your *industry* (e.g. medical or health professions education) in the wider socio-cultural, political, and economic context. This will help you keep abreast of trends and policies that might affect the education you provide and help you identify and seize opportunities.
- Educational leaders need a sound understanding of *education systems* – structures, funding, and programmes – and possess a theoretical educational knowledge.
- Become an *expert* in a specific area, project, or initiative. This helps to build credibility particularly when your power and influence is relatively low because of position in the organisation and professional hierarchies.
- Understand your *strengths and weaknesses,* this helps you to build effective teams by compensating for your deficits, and structure your leadership development.

behaviours worked best in what circumstances. This was addressed through the work of, among others, Fiedler [18] and Hersey and Blanchard [19] who describe 'contingency theories'. The latter achieved widespread popularity through the *One Minute Manager* series, with their explication of *situational* leadership. Hersey and Blanchard's argument is that leaders need to adapt their style to variations in the competence and commitment of followers. The four styles identified are: directing, coaching, supporting, and delegating. A situational approach to leadership is also adopted by Adair in his 'three circles' model [20]. Adair recommends that, depending on the circumstances, the focus of a leader's attention should be distributed flexibly between the *task*, the *individual*, and the *team*.

It became apparent in the 1980s that none of the existing leadership theories offered advice on how to cope in environments of continuous change. Various authors highlighted that the models described so far were effectively managerial or transactional and so helped people plan, order, and organise at times of stability but were inadequate at describing how people or organisations might be led through periods of significant change. A new paradigm emerged – that of *transformational* leadership [21]: leadership through the transformation of the willingness of others to work towards the goal of some future desired state.

A fusion of the transformational notion of social influence and leadership trait theory led, in the 1980s and 1990s, to the emergence of *charismatic* leadership as a solution to organisational problems. Effective leaders were viewed as dominant personalities brimming with self-confidence and offering a clearly articulated ideology. They offered a strong role model and had high expectations of followers. However, it quickly became apparent that this culture of organisational 'heroes' was both unhealthy and unsustainable, and more recently there has been a shift in thinking that emphasises the thoughtful, inclusive, person-centred, and values-led leader, highlighting notions of followership and relational leadership [22, 23] and bringing to the fore ideas of *engaging* [24], *distributed* [25], *collective* [26], and *collaborative* [27] leadership. More recent considerations of leadership stress the need for leaders to be aware of, and responsive to, the complexity of the systems and organisations in which they work [28, 29]. From these perspectives, models such as *adaptive* or *eco-leadership* have emerged [30].

However, as many authors, including Bryman [31] and Storey [10], have noted, these stages signal a change in emphasis – not a shift away from previous models. Like the abominable snowman, a definitive model of leadership remains an elusive thing to track down.

Educational Management and Leadership

Writers and researchers who have focused on educational leadership and management, as opposed to general theories of leadership, tend to locate phenomena associated with educational leadership within various paradigms [36], perspectives [37], models [1], or metaphors [38]. All these terms tend to be applied in similar ways rather than reflect significant differences in meaning, and we will use them interchangeably here.

National education policies emphasise the centrality of educational leadership and management. There is no shortage of leadership advice from government, consultants, academics, and education authorities to leaders in schools, colleges, and universities, but as Bush notes, 'many of these prescriptions are atheoretical in the sense that they are not

BOX 37.3 Typology of management models (adapted from Bush and Glover [41])

- Formal
- Collegial
- Political
- Subjective
- Ambiguity
- Cultural

underpinned by explicit values or concepts' [39]. Although leadership and management are practical activities, theories can help inform and influence practice by suggesting new ways to do things and frameworks or lenses through which events, interrelationships, and situations can be seen. Pragmatic decisions might appear to be common sense, but 'common sense' is grounded in assumptions, limitations, a view about an organisation, and a particular frame of reference. 'Theory-for-practice' [40] encourages us to look at organisations and relationships in a different light, helping us interpret and make sense of complexity and ambiguity, and make relevant and informed management and leadership decisions.

Most of the theory and research on educational management and leadership relates to the context of schools and colleges, with very little in the context of the education of health professionals. However, the broader research is useful in helping us look at the organisations and relationships involved in delivering medical education.

Bush and Glover [41] describe a typology of management models, building on previous work by Leithwood et al. [42] This framework is shown in Box 37.3. As discussed above, we believe that effective leadership, management, and followership are inextricably intertwined. For the remainder of this section we will consider general leadership and organisational theories relevant to the six management paradigms listed in Box 37.3. On the way, we will also consider the applicability of each model to the leadership and management of health professions education.

Formal Management

Formal management models view educational organisations as hierarchical, bounded bureaucratic systems with an official structure – often represented by an organisational chart – and relationships characterised by virtue of the legitimate authority vested in them. Managerial decisions are seen as objective, rational, and goal oriented. Leaders have positional power over those beneath them but are accountable to internal and external bodies or committees, such as the senate, councils, or boards.

Theories drawn from structural or system sociology underpin these managerial models, most notably the concept of *bureaucracy*. Weber [43] describes a bureaucratic organisation as existing to achieve established purposes, with work best achieved through specialisation, division of labour, coordination, and control – a description that sits well with contemporary education organisations.

In *managerial leadership*, leaders are responsible for:

- setting goals and ensuring the organisation reaches the goals
- maintaining the integrity of the educational structure and working within it using power and authority to the organisation's advantage
- 'bounding' the system in alignment with how the organisation sees the relationship with the external environment.

In these (leader-centric) models, leaders are often seen as holding much of the organisation's power and are expected to make decisions based on their own knowledge and expertise. In practice, most organisations invest power and authority across a senior management team rather than in a single leader, thus ensuring that the hierarchy stays in place and that the formal systems, structures, and procedures are maintained. These 'Apollonian' [44] structures can provide a platform for the 'hero' leader; indeed, in education, such a figure may actively be sought out. Universities, for instance, often invest a huge amount of time, energy, and resources in seeking out the best new vice-chancellor. Storey notes that in a managerial organisation: 'The need for leadership is often addressed in terms of the "reputational capital" that a celebrated leader can bring … this is very interesting and revealing because it highlights the importance of stakeholder perception' [10].

In reality, modern educational organisations are complex and 'messy'; professional staff expect to be consulted and involved in decision-making; goals are difficult to state and measure, as many of them are long term; and individuals do not necessarily behave in a prescribed way. In medical and health professions education, additional areas of conflict arise because many individuals have authority and power in one arena – for instance, in their role as a consultant clinician – but may have little organisational power or authority in an academic setting. This, coupled with traditional power discrepancies between the professions in terms of authority, accountability, status, and reward, lead to additional tensions in health care education organisations that do not exist in schools, colleges, or other university departments.

So, while it is recognised that bureaucracy may be an appropriate organisational model for educational management [45], inherent weaknesses exist – predominantly a tendency for the organisation to become preoccupied with the maintenance of its own structure and processes (rules, regulations, and procedures), thus forgetting its original educational purposes or relegating them to a position of subordinance [46]. Another general criticism is that bureaucracies are deprofessionalising, stripping away the autonomy of practitioners [47]. If the leader focuses too much on the system and its underpinning processes, and forgets to balance this with paying attention to the teams and individuals that work in the organisation, then there is risk of over-management and lack of creativity and responsiveness to change.

Minztberg [48] dissects organisations into five generic components and in doing so identifies a simple organisational taxonomy, namely organisations as:

- a machine bureaucracy
- a professional organisation

- an entrepreneurial start-up
- an adhocracy.

A *machine bureaucracy* can be conceptualised as a factory or production line, pumping out a standardised product under instructions from management. In the *entrepreneurial start-up*, the owner-manager takes the strategic decisions while remaining hands-on in the day-to-day business of the organisation. The *adhocracy* is an ambiguous and organic form in which structural elements combine, disassociate, and recombine to address complex problems. Health care education tends to be structured in *professional bureaucracies*, and this has profound implications for management and leadership.

In a professional organisation, Mintzberg suggests that a small *strategic apex*, assisted by a supportive *technostructure* and *staff*, manages through a small *middle line* a large *operating core* of professionals. For the coordination of its work, a professional bureaucracy relies on the standardisation of skills through training and indoctrination, and although the operating professionals have considerable control, the standards to which they perform are set outside the structure of the organisation, usually by a self-governing association of fellow professional organisations. The professional tends to resist rationalisation of their skills as this drives the structure towards that of a machine bureaucracy, thus destroying professional autonomy. Organisational strategy in the professional organisation is represented by the cumulative effect over time of multiple projects or initiatives undertaken by its members. As a consequence of these characteristics, the professional organisation tends to be rather inflexible, good at producing a standard unchanging product, but ill-suited to adaptation. Response to change is slow – evolution as opposed to revolution – and 'top-down' calls for reform are usually resisted.

In a professional organisation operating in a *collegial* management model, leaders and managers are seen to sit 'alongside' colleagues rather than manage from 'above' [49] even though the organisational charts may portray otherwise. Collegiality assumes a common set of values, an authority of expertise, and decisions reached through a process of discussion leading to consensus. Collegiality has been shown in mainstream education to improve the quality of decision-making, to bring about more effective implementation, and, perhaps most importantly, to be the type of management teachers want [50].

Unfortunately, as will be apparent, collegial management in the public sector is in conflict with the prevailing bureaucratic culture, a conflict heightened by governmental pressures for increased output and greater public accountability.

Collegial Management

The collegial management model is normative, predicated upon members of the organisation sharing beliefs, values, and norms and that decision-making is made democratically. However, this belief is often illusory rather than grounded in reality, and leaders need to spend time ensuring that consultation and participation in decision-making is real. They need to seek out and clearly define the beliefs and values that individuals and groups from different professional backgrounds share. Collegial decisions are founded on expert authority.

The collegial model is highly appropriate for small professional organisations. Larger units, such as universities, also retain many elements of collegiality in that they have extensive committee structures to facilitate representation and participation, and academic autonomy and freedom are valued at the 'bottom-up' approach. However, there has always been a separation between academic and non-academic staff and the right to participation in decision-making that these roles enjoy. More recently, universities have been described as moving towards a more 'top-down' managerial model, partly in response to increased competition in a global marketplace. This shift also enables clear organisational goals that 'fit' all disciplines and professional groups within the educational organisation to be agreed. It has also led Hargreaves to describe the emergence of a phenomenon widespread in the public sector, that of 'contrived collegiality' [50], consultation without the possibility of control.

The management model of collegiality leads us to consider transformational, participative, and relational leadership models, all of which rest on the notion of the leader as 'first among equals' [51], utilising the skills of influencing, negotiation, listening, facilitation, and consensus building rather than acting in the role of a commanding and controlling hero leader. Such leaders are aware of the complexity and interactivity of decision-making processes and are able to recognise and harness the expertise of the professionals whom they lead, as well as demonstrate sensitivity towards the 'informal codes of professional practice which govern expectations' between the leader and his or her staff [52].

A number of what have come to be known as 'new paradigm' models of leadership [31] emerged in the 1980s. New paradigm models depict leaders as managers of meaning rather than in terms of the way they influence processes. Such leaders define organisational reality through the articulation of a clear vision and its supporting values, a major re-interpretation of leadership from an authority relationship (sometime referred to as *transactional leadership*, which may or may not involve some form of pushing or coercion) to a process of influencing followers by inspiring them, or pulling them to a vision of some future state. This new form of leadership is called *transformational leadership* because such individuals are said to 'transform' followers [53].

Transformational leadership is characterised by the '4 I's' of Bass and Avolio: [21]

- idealised influence
- inspirational motivation
- individualised consideration
- intellectual stimulation.

Transformational leadership and its near relatives, *interpersonal* and *relational leadership*, are rooted in the philosophical belief that meaning is socially constructed. Grint describes the 'constitutive approach' as one in which leaders not only mould followers' interpretation of the environment but also present it as being the correct interpretation [54]. This is one of the critiques of transformational leadership, in that leaders may be tempted to manipulate meaning to suit their own ends. Writers such as Maccoby [55] highlight that there is a dark side to such charismatic leadership, that of the narcissistic leader, who can challenge change and disrupt the status quo but also can be arrogant and grandiose, thriving on risk and seeking power, glory, and admiration. A wave of corporate scandals over the last 20 years has served to highlight the dangers of a narcissistic leader. Further on in this chapter, we shall see how the concepts of *values-led* and *servant leadership* help to compensate for some of the possibly negative aspects of the transformational approach.

Another post-charismatic model of leadership is described by Fullan [56], a model based around embedded learning, devolved leadership in teams, and learning as a product of confrontation, mistakes, and experimentation. Practice is made public in a collaborative culture.

Distributed leadership or 'leadership at all levels' is currently receiving considerable attention in education and the health care sector [57] though not without critical challenge [58, 59]. The case for distributed leadership is made on the basis of the following:

- a belief in leadership teams: a team being a more effective unit than an individual
- the fact that as organisations become more complex places to manage and lead, there is a need for leadership at all levels, with 'coherence making' [29] as everyone's responsibility
- the development of pools of talent, from which tomorrow's leaders may be drawn.

A related model is that of *collaborative leadership*, focusing on a commitment to partnership working for the good of service users, which draws on many of the approaches outlined above. However, it also demands a specific set of qualities and behaviours. Collaborative leadership shares the burden but also shares the power and resources. This requires that: '… collaborative leaders are personally mature. They have a solid enough sense of self that they do not fear loss of control' [60]. Above all, collaborative leaders must listen, and be seen to do so. As van Zwanenberg notes: 'The person in a position of leadership who is not open to actively listening, questioning, and reflecting in a very conscious way will be judged as a hypocrite if they continue to talk the language of partnership …' [61].

All of the above models value human capital and the development of sustainable organisations, and for leaders and managers this has a clear message. An effective organisation will align the motivational goals of the individual employee with its own needs, not only to motivate employees but to avoid risk. A strong organisation is resilient and (through the people involved in its activities) can respond appropriately to internal and external change. The consequences of poor role/person fit are profound, and Argyris [62] found that, over time, frustration intensified, leading to absenteeism, psychological withdrawal, restriction of output, feather bedding, sabotage, and formation of power groups to redress the balance. The alternative strategy was to climb the hierarchy to what was perceived to be a better job and perpetuate the whole process.

To summarise a wealth of literature on the human side of organisations, successfully led organisations *empower* employees and *invest* in people (see Box 37.4).

BOX 37.4 Leading through people

Empower employees	Invest in people
• Redesign work	• Hire the right people
• Provide autonomy	• Reward them well
• Encourage participation	• Promote from within
• Focus on job enrichment	• Train and educate
• Emphasise teamwork	• Share the wealth
• Ensure upward influence	

Political Management

Power flows in educational organisations through three main conduits, as follows:

- bureaucratic processes (boards, committees, management hierarchies)
- collegiality (consensual decision-making informed by expert knowledge)
- micro-politics [63] (negotiation, bargaining, conflict, subterfuge).

The potential sources of power in organisations have been analysed by a number of theorists [63, 64]. Hoyle [40], for instance, has distinguished four main types, as follows:

- *structural* – power as the property of office
- *personality* – charisma and 'the aura of authority'
- *expertise* – which includes access to information as well as specialised knowledge
- *opportunity* – power through the occupancy of key administrative roles.

Politics then is ultimately about the deployment of such resources through the wielding and yielding of power and the day-to-day interactions of people and their ideologies. Sometimes, considerable power resides in the lowliest of roles, such as the hospital laundry worker or switchboard operator. Authority [65] is a related concept, differentiated from power in Box 37.5. Baldridge's work on universities in the USA [51] highlights the role that interest and pressure groups have on influencing decision-making and forcing change. His model of how policies emerge is iterative and dynamic: competition for scarce resources leading to formal and informal groupings jostling and positioning themselves for control, with power accruing to dominant coalitions. Personal and professional interests feed into the melee, the former in relation to status, working conditions, and reward, and the latter to gain commitment to particular ways of working, curricular models, teaching and learning methods, or student groupings.

Hoyle describes how educational leaders can adopt political strategies to achieve their aims [40]. These include *dividing and ruling*, *co-optation*, *displacement*, *controlling information*, and *controlling meetings*, with the last of these strategies achieved through the dubious strategies of rigging agendas, losing recommendations, lobbying members of the group, invoking outside bodies, or massaging minutes.

The exercise of power is also a major preoccupation of *transactional leadership*, a leadership model based on the following four assumptions:

- people are motivated by reward and punishment
- social systems work best with a clear chain of command

BOX 37.5 FOCUS ON: Power and authority

The relationship between power and authority is not straightforward. It is perfectly possible to have one without the other. Power, as has been described, is about the control and manipulation of scarce resources – whether they be physical, economic, knowledge-based, or to do with personal qualities [3]. Authority relates to the right, or perceived right, to exercise power.

According to the economist and sociologist Max Weber [67], all leaders are surrounded in a myth of superiority that derives from three forms of legitimate authority: charismatic, traditional, and rational–legal domination.

Charismatic domination derives from the personal qualities of the leader. Many religious leaders, and some political ones, derive their authority from this form of domination.

Traditional domination relates to positions that preserve long-established values and social relationships. The UK Royal Family have authority, not necessarily because of charisma but because an otherwise meritocratic society allows them to continue to adopt that social position derived purely from inheritance.

Rational or legal domination is the most advanced, and relates to authority as a function of delegated power within a bureaucracy. Those in authority give orders (and expect to be obeyed) because their office gives them the right to give orders. However, there is a context-specific element to this, and orders are only to be obeyed if they are relevant to the situation in which they are given. The head of your department or school has rational–legal authority, and who knows, may also have charisma …

Power in educational institutions may go hand in hand with authority, such as those budgetary decisions made by the dean, head, principal, or executive board. However, power in organisations is often wielded without authority through a process of 'influence' and micro-politics.

- authority resides within line management
- subordinates are there to do what their manager tells them.

Leadership of the transactional variety starts then to look very much like management, a similarity emphasised by Kotter [66], who asserts that transformational leadership is about strategy and people, whereas transactional leadership revolves around systems and resources. Kotter, though, is also clear that both these leadership paradigms are needed in an effective organisation.

> Leadership is different from management, but not for the reason people think. Leadership isn't mystical and mysterious. It has nothing to do with having charisma or other exotic personality traits. It's not the province of the chosen few. Nor is leadership necessarily better than management or a replacement for it: rather leadership and management are two distinctive and complementary activities. Both are necessary for success in an increasingly complex and volatile business environment [8, p. 103].

Subjective Management

Bush [1] describes subjective management models as those which 'assume that organisations are the creations of people within them'. Participants are thought to interpret situations in different ways, and these individuals' perceptions are derived from their background and values. In this socially constructed view, organisations' structures, people, and activities have different meanings for each of their members and exist only in the experience of those members. Different writers vary in the emphasis they put on individuals rather than the organisation; some see the tension between the individual perspective and organisational collective as a 'chasm', whereas others see it more like a dialectic, with interdependence between constantly changing individual and collective meanings.

Post-modern leadership is a relatively undefined concept; it has similarities to the subjective model in that different actors are seen as bringing multiple meanings and realities to any organisation or situation. 'Situations must be understood at local level with particular attention to diversity' [68]; this emphasises the fluid and 'chaotic' nature of contemporary society in which power is *enacted* (ibid) through all members, thus leading to empowerment. Leaders need to rely on their skills as interpreters of meaning articulated in different ways by multiple stakeholders. The hero visionary does not fit in the post-modern world; leaders need to listen to, and focus on, people as individuals with their own world views.

Ambiguity Management

Ambiguity models emphasise organisational change, uncertainty, and unpredictability. Organisational goals, systems, and processes are seen as unclear and not well understood, and participation in decision-making is fluid as individuals opt in and out. Organisations are characterised by fragmentation and 'loose coupling' [69], particularly educational organisations in which members have a high degree of autonomy and discretion. The most famous of the models is the 'garbage can' model described by Cohen and March [70] in their research into colleges and universities, identifying that decisions are the result of fluid processes in which problems, solutions, participants, and choice opportunities all interact to generate outcomes or decisions.

Gilbert [71] defines *situational* or *contingent leadership* as 'an approach based on the commonsensical idea that there will be interactions in most situations between the leader's attitude and attitudes, the tasks to be undertaken, the strengths and weaknesses of the team and the environment in which the leader and team have to operate'. Contingency leadership models see leaders as having to adapt their stance and style to the particular situation rather than seeing one leadership approach as being the 'right' one.

BOX 37.6 Leadership that gets results: Leadership styles (adapted from Goleman [14])

	Coercive	Authoritative	Affiliative	Democratic	Pace-setting	Coaching
What the leader does	Demands immediate compliance	Mobilises people towards a vision	Creates harmony and builds bonds	Forges consensus through participation	Sets high standards for performance	Develops people for the future
When it works best	In a crisis	When a clear direction is needed	To heal rifts and motivate people in stressful circumstances	To get buy-in from valuable employees	To obtain quick results from a well-motivated team	To help an employee improve their performance
Overall effect on climate	−	++	+	+	−	+

Peck et al. [72] note, for example, that 'transformational characteristics may be more appropriate in the earlier stages of a transition process, with the transactional characteristics required to stabilise transformational change'. It is not that one is superior or inferior to the other but rather that effective leaders appear to be able to utilise many styles and select those appropriate to the specific situation.

What are those styles? A Hay-McBer study identified six drivers of organisational climate or working atmosphere [14] (see Box 37.6). Effective leaders used all styles flexibly dependent on the situation, with the *authoritative* leadership style found to have the most positive effect on business performance.

In the complex world of medical and health care education, effective leaders will be those who can respond to different situations in a measured way, selecting the appropriate leadership style to suit the situation. This point is particularly relevant for those who are coming from one leadership role (e.g. in clinical practice) into another (e.g. academia), as Gilbert [71] highlights: 'Success in one environment does not necessarily translate to another'. However, being able to adopt different styles of leadership is not the same as changing one's personality. Followers value authenticity and consistency in leaders very highly, these behaviours generate trust which is a key element of high-performing teams.

Cultural Management

Cultural models of management are very influential in mainstream education, and they emphasise the importance of informal modes of influence and of the centrality of values, beliefs, and ideologies. Individuals are seen as actors who bring their ideas and values to their relationships with others. In turn, this influences the way in which an organisation develops norms and traditions – 'the way we do things around here' [73] – which are reinforced by symbols, ceremonies, and rituals. Symbols are central to the construction of meaning, and in medical and health care education these abound as part of the undergraduate (preregistration) socialisation process and continue throughout professional life. The importance of understanding the meaning ascribed to linguistic, behavioural, and visual or physical symbols cannot be underestimated by leaders, particularly those who want to work interprofessionally or across organisations.

A strong organisational culture enables people to identify themselves and their aspirations with the purpose of the organisation and can heighten faith and confidence in an organisation in the midst of environmental turbulence and adversity. Organisational culture can be influenced in a variety of ways. The academic trappings (gowns, maces, medals, crests, etc.) adopted by many professional organisations in medicine are good examples of how cultural expression serves to define, or brand, an organisation. Small changes in culture make big differences, and it is the leader's responsibility to pay attention to the detail as well as the big picture. Quite often, the solution to an underperforming organisation is not to undertake a massive restructuring exercise but to change the organisational culture. Paraphrasing the theories of Edgar Schein [74], Figure 37.2 illustrates the levels at which this must be undertaken.

Leaders also need to be aware of how changing demographics and widening participation impact on organisational and professional cultures; issues concerning gender, race, ethnicity, and class can lead to a need to reframe cultural norms and values. Leaders may well have to deal with uncomfortable moments as once-dominant ideologies and positions are challenged and come under scrutiny.

Leaders with a strong ethical and values-led approach, which acknowledges cultural shifts and responds appropriately, will be more effective than those who operate merely on exchange transactions. Such person-centred, values-driven, or moral leaders need to ensure that their central values accord with those in the educational organisation, or there will be dissonance and discomfort for all those involved. In times of financial hardship and competing agendas, maintaining integrity and an ethical stance can sometimes be very difficult, particularly so if we do not fully know ourselves. Such self-knowledge features large in the writings of Daniel Goleman on emotional intelligence. Goleman [34] argues that in order to be successful in dealing with others, it is important to have effective self-awareness, control, and management of your own emotions, and a deep understanding of your own goals, responses, and behaviour patterns.

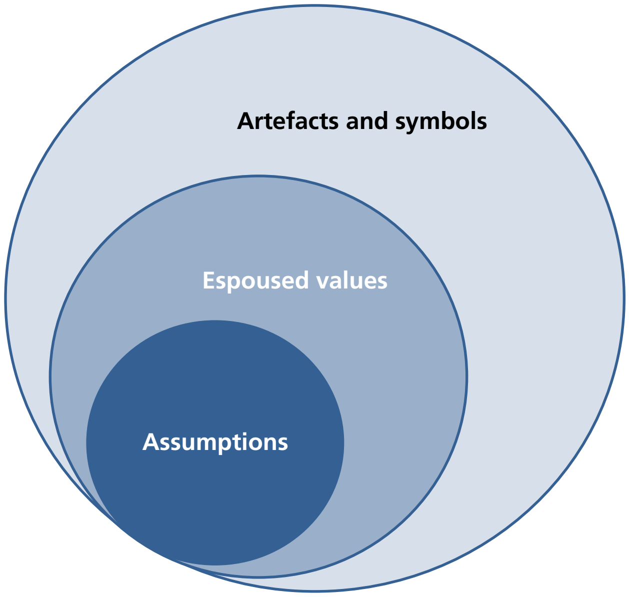


Figure 37.2 Layers of organisational culture [74].

Personal values and concern for others are also at the heart of Greenleaf's *servant leadership* [75], where, as Collins puts it, 'professional will is seen alongside personal self-effacement' [76]. The difference between the servant leader and other forms of leadership is that the servant leader 'is servant first … It begins with the natural feeling that one wants to serve, to serve first. Then conscious choice brings one to aspire to lead' [75]. The primacy of serving a higher purpose has made Greenleaf's approach popular in the church and not-for-profit sectors, and the central concept of 'stewardship' sits easily alongside the drive to develop sustainable organisations and embed succession planning through the diffusion of leadership throughout the organisation. Both private and public sector organisations are increasingly recognising the importance of a diverse workforce. Taking an inclusive leadership approach makes sound business sense and helps generate fresh perspectives and the sense of belonging that most followers need [77]. Central to these leadership approaches are a strong sense of social accountability and responsibility, a systems perspective, and an awareness of unconscious bias in interacting with, recruiting, and rewarding others. The Linkage Inclusive Leadership Model [78] for example, identifies three key areas of focus of inclusive leaders:

- Leads self – they minimise unconscious biases through candid conversations and being personally vulnerable and open to learning.
- Leads relationships – they build relationships with people by networking, adapting their style to others and encouraging others' development, competence, and confidence.
- Leads culture – they build a safe, respectful, and trustful culture and work environment that allows people to be their true selves.

One particular values-led approach that is particularly relevant to medical education is that of *instructional leadership*, which Bush describes as 'focussing on the direction rather than the process of leadership … with a firm emphasis on the purpose of education and the need to focus on teaching and learning as the prime purpose of educational institutions' [1]. In professional education there is a renewed focus on teaching and learning and in supporting teachers' professional development, e.g. in the establishment of bodies such as the UK Academy of Medical Educators [79] and in the appointment of senior managers in health care organisations and universities with named responsibilities for 'teaching and learning' or education.

Leading Teams

Recent thinking has shifted the leadership focus back from whole organisations to distributed leadership involving 'followers' working within small groups and teams [57, 80]. *Team building* has long been a part of organisational and professional development activities. Manz and Sims suggest that a new management style will be needed for team-based organisations – *super-leadership* – with the locus of control shifting from the leader to the team [81].

Teachers use team-leading and team-building concepts all the time when they are managing groups of students and trainees, and the majority of teacher training programmes include small-group teaching, looking at the role of the teacher as facilitator in helping a group achieve its outcomes in terms of both task and process. In educational leadership, aspiring or existing leaders can draw on their educational expertise and use their skills in a transferable way with the teams they are leading.

Research by the UK Government's Performance and Innovation Unit [82] suggests that the climate within a given team can account for up to 30% of its performance and that the leader of the team has a primary role in creating the appropriate climate. The climate of a team is difficult to break down into its component parts, but includes elements such as levels of personal autonomy, adequate reward and recognition, and clarity of roles and boundaries. The converse has also been identified in what West and Lyubovnikova [83] refer to as 'pseudo-teams', groups that lack reflectivity, interdependence, and shared objectives. An effective leader then will need an understanding of the nature of teams, the necessary conditions for them to function well and be able to achieve a balance between individual members' needs, the task, and team processes [20]. Team leaders need to have certain behaviours that facilitate teamworking. Kozlowski et al. [84] identify the following helpful behaviours:

- developing shared knowledge among team members
- acting as a mentor
- instructing others
- facilitating group processes
- providing information
- monitoring performance
- promoting open communication
- providing goals
- allocating resources efficiently.

Other skills include the ability to lead participative meetings, listening skills, the ability to handle conflict, group-centred decision-making skills, team-building skills, and coaching and developing skills.

Teams are dynamic units that are constantly evolving, and this evolution involves a number of identifiable stages, famously identified by Tuckman [85]. The stage at which a team finds itself – forming, storming, norming, performing, or mourning – is directly related to the quality and quantity of its output. A team leader needs to know which stage his or her team is at and support the team through its various transitions.

For a team to achieve its potential, it is also essential that the operational roles of its members are matched as closely as possible to their individual working and social style. People function within teams in different ways, but these can be identified using a range of tools and roles can be allocated accordingly to ensure that the team reaches its full potential. Belbin [86] famously classifies individual behaviours in teams into nine distinct 'roles' – observable tendencies of individuals to relate with their colleagues and contribute to the team (see Box 37.7). Such team roles are tendencies, not fixed patterns, and most team members will have at least one secondary tendency. Not every team will have all roles at its disposal, but Belbin's research found that teams that did, tended to perform more effectively. From this perspective, no single team role is more important than another and the potential weaknesses of each role should be seen as allowable. Successful teams thrive on their diversity and allow members to play on their strengths. The task of the leader is to exploit the strengths, allow for the weaknesses, and exploit the natural motivation that an understanding of the needs and capacities of everyone in the team engenders.

BOX 37.7 Belbin team roles [86]

Plant
An 'ideas person'. Thoughtful, creative, brilliant, radical. Better at thinking than communicating. Not interested in details.

Co-ordinator
The 'chairperson'. A good team captain. Involves all team members and plays a mediator role in discussions. Co-ordinates work rather than doing it.

Monitor evaluator
The 'critic'. Evaluates everything carefully, seeing both sides of every argument. Cold and objective. Can be perceived as negative and/or unenthusiastic.

Implementer
A 'doer'. A reliable worker who puts ideas into action and gets on with it. May not be very imaginative or flexible.

Completer finisher
The conscientious details person. Pays attention to detail and completes the job. Can be a worrier.

Resource investigator
The 'networker'. The group's ambassador and detective, making friends and tracking down information and resources. Initial enthusiasm can fade before the project is completed.

Shaper
The 'driver'. Pushes through ideas and keeps projects moving, enjoying the cut and thrust of the action. Can upset others as they do this.

Teamworker
The 'peacemaker'. Lubricates the team with diplomacy and helps keep the team working effectively. Everybody's friend. Can be indecisive.

Specialist
The 'expert'. Provides technical or other specialist knowledge to the team. Input is usually restricted to their own specialism.

Leading Change

Like leadership, change is an elusive concept, and the literature on change and change management is vast [87–89]. As a working definition, change can be considered as the transformation of an individual or a system from one state to another, a process that may be initiated by internal factors or external forces or both. An understanding of the nature of change and change management is essential for any educational leader, none more so than those who work in the health sector. Behind every effort to lead change is a set of assumptions or theories, whether implicit or explicit, about how to bring that change about. In considering recent developments in health care education, we wish to emphasise a shift in the conceptualisation of change and change management that has occurred in recent years.

Mintzberg [90] summarises this shift in thinking in arguing that change can be thought of as either a pre-planned and predictable event or alternatively as an open-ended, ongoing, and unpredictable process aimed at aligning and realigning in response to a changing environment. Early attempts at delineating change, such as that of Lewin [91], focused on defining phases or stages of the process that might be managed before moving on to the next. Lewin's model describes unfreezing the present situation, a transition stage, then moving to a new situation and refreezing. This approach would apply to the development and implementation of a new curriculum. Bullock and Batten [92] similarly define a four-stage model of exploration–planning–action–integration. Implicit in both, and indeed in many other models, is the constancy of both the existing and the new state and the assumption that, given the appropriate skills and resources, the movement to the new state can be brought about in a controlled manner. Yet more models of change emphasise the problem-solving approach of problem identification, solution generation, choice of solution, implementation, and evaluation [93, 94], again stressing a rational, linear progression through the change process.

More recently, there has been acceptance that a rational approach to change will only get you so far, and that it is change, not stability, that is the steady state.

House [37] notably outlines a trilogy of perspectives on educational change: *technological*, *cultural*, and *political*. Not only does this provide a useful model for the analysis of change events but it acknowledges the importance of issues other than the innovation itself, particularly how innovations are interpreted and integrated into their destination social fabric (the cultural perspective) and how power, authority, and competing interests are played against each other (the political perspective). That is not to say that a techno-rational approach to change is wrong, but, as we have discussed, having multiple perspectives or frames [95] through which to examine problems enables a richer repertoire of management strategies to be deployed.

The second emerging theme is the acceptance that change is here to stay. Furthermore, as Fullan points out, 'it is a mug's game to hope for the speed of change to slow down' [96]. Change and reform in education and health services have become a way of life. In recent years, seemingly disconnected and often conflicting policies have rained down on health and education professionals with increasing rapidity as politicians attempt to make their mark in a rapidly revolving political cycle. However, this is not a new phenomenon, and Baker has concluded after a lifetime of study of educational reform that: 'Planned change for (educators) is not the cumulative development of a comprehensive strategy. Rather, it is "one dammed thing after another"' [97].

Leading Systems

Multiple perspectives on change and the 'acceleration of just about everything' [98] are then acknowledged features of our social era, a period that has been labelled variously as post-modernity [99], late capitalism [100], and the informational society [101]. These key features have led naturally to an interest by social scientists in complex systems. Indeed, Hargreaves et al. [102] adds a fourth 'post-modern' perspective to House's original typology, acknowledging *complexity* – alongside two other characteristic elements of late capitalism, consumerism and diversity – as a key feature of twenty-first century educational change. New paradigms of change management are developing borne out of the language of chaos and complexity theory, offering insights into the nature of emergence, innovation, learning, and adaptation [103].

Human systems have a tendency to avoid the onset of chaos if at all possible. Sometimes, this occurs to such a great extent in organisations that stability and equilibrium set in, leading to stereotypical patterns of behaviour and poor change responsiveness. However, the environment is sufficiently chaotic and elements within it sufficiently interrelated not to allow the state to persist for long, and an oscillation sets in between the twin pulls of the urge to maintain integrity (autopoiesis) and the impulse to explore and adapt to the changing environment.

In social systems, stasis occurs where there is a high level of agreement and an equally high level of certainty. Chaos ensues when there is neither agreement nor certainty. Between these two extremes lies an area of variable agreement and variable certainty, and this is the zone of complexity where anything is possible, the 'swampy lowlands' [104] wherein lie the really important problems worth grappling with.

Businesses [105], schools [50], and health care organisations [106, 107] have all been viewed as 'complex adaptive systems' where 'order is not totally predetermined and fixed, but the universe (however defined) is creative, emergent (through iteration, learning and recursion), evolutionary and changing, transformative and turbulent' [87]. Despite this constant disequilibrium, the complex adaptive system strives to create order out of chaos, constantly reorganising in response to the changing environment. Complex adaptive systems display 'perpetual novelty' [108] and a propensity for problem solving. So what does complexity mean for educational leadership?

Broadly, the literature identifies a spectrum of possible ways that leaders might interact with a complex system, ranging from the descriptive to the manipulative [107] (see Figure 37.3). At the interactionist end of this spectrum are the systems thinkers of the 1980s who position themselves outside the system, recommending influencing it through judicious and thought-through interventions [109]. More nihilistic are academics such as Stacey [110], who see the leader's role as one of articulating emerging themes and making tentative suggestions about what might be going on. Axelrod and Cohen [111] occupy some of the middle ground in their advice for those that seek to influence, to pay attention to internal processes, particularly those that foster variation, encourage interaction of system agents, and promote selection pressures. Alongside them, Johnson and Scholes [112] are two of a number of authors who advocate concentration on developing a culture in which more-or-less the right thing is then likely to happen, more of the time.

Across the spectrum, the message is the same, essentially a non-directive approach to change management, as Mintzberg and colleagues advise: 'The best way to manage change is to allow for it to happen, to be pulled by the concerns out there rather than being pushed by the concepts in here' [113, p. 324]. Pascale et al. [114] similarly advocate perturbing the system and standing back to observe the results. Fullan [29], too, countenances 'never a check list – always complexity'.

Fraser and Greenhalgh [115] summarise a number of complexity concepts applicable to education and training, as follows.

- Neither the system nor its external environment is, or ever will be, constant.
- Individuals within a system are independent and creative decision makers.
- Uncertainty and paradox are inherent within the system.
- Problems that cannot be solved can nevertheless be 'moved forward'.
- Effective solutions can emerge from minimum specification.
- Small changes can have big effects.
- Behaviour exhibits patterns (that can be termed 'attractors').
- Change is more easily adopted when it taps into attractor patterns.

Both Fraser and Greenhalgh and Fullan acknowledge the complex nature of educational change and that self-organisation of an educational system can occur given minimal direction. Both also acknowledge the

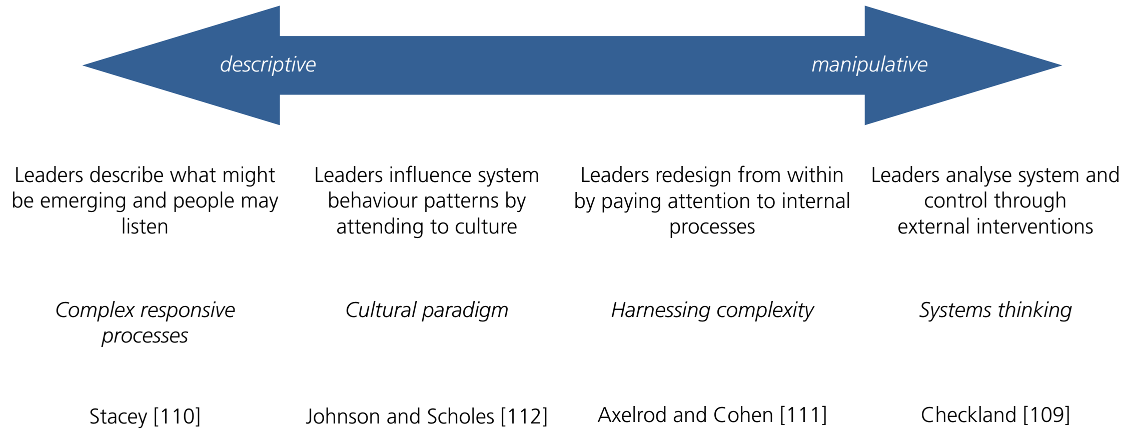


Figure 37.3 Leading in complex systems.

importance of *attractors* or traits within systems that set the direction for this self-organisation. Values, principles, goals, theories, and leadership can all act as attractors, pulling together ideas, information, people, and energy in generating new and emergent patterns.

Setting Direction

As we identified earlier, one of the key features of leadership, as opposed to management, is the setting of direction through the articulation of a vision and the development of a strategy.

Strategy, though, is a word that has multiple meanings concisely summarised by Mintzberg et al. [113], namely:

- *a plan* – a guide to take one along a path from a current state to some future state
- *a pattern* – a consistent pattern of behaviour
- *a position* – the location of a particular product in a market
- *a perspective* – a way of interacting
- *a ploy* – a tactic designed to outwit a competitor.

Also, one, some, or all of these constructions may be employed in the process of strategic planning.

Strategic planning has a number of functions. It sets direction, focuses effort, defines the organisation, and provides a consistency of approach and behaviour. However, there are downsides, as a strategic plan, beautifully written and long in gestation, may encourage 'groupthink', blinding the leadership of an organisation to new opportunities. Consistency, too, is not always desirable, as it may serve to overlook the rich diversity within any organisation. A strategic plan needs therefore to be flexible enough to take into account changes in the environment or new ideas and ways of doing things that emerge during the plan's lifetime.

The classic model of strategic planning traces a journey from a vision (often expressed as 'purpose' or a 'mission'), through aims, objectives, and action plans, to monitoring arrangements (see Figure 37.4). It is a pathway underpinned by the expressed and (one hopes) lived values of the organisation.

To be successful, strategic planning needs to take account of the views of all stakeholders. For a medical school department, these might include the university, students, teachers and tutors, the hospital, the local postgraduate organisation, other medical schools, patient groups, professional bodies, and regulators. Taking people with you is key, and to settle on a strategic direction too early in the process is fraught with danger. In Claxton's words, 'slow knowing' is the key: 'One needs to be able to soak up experience of complex domains – such as human relationships – through one's pores, and to extract subtle, contingent patterns that are latent within it. And to do that one needs to be able to attend to a whole range of situations patiently without comprehension; to resist the temptation to foreclose on what that experience may have to teach' [116, p. 192].

Strategic planning should therefore have at its heart a reiterative process of evaluative analysis, a gauging and re-gauging of the operating environment of the organisation, canvassing the views of stakeholders, understanding trends, and developing an awareness of the shaping forces at work. Through this process, leaders may also act to 'seize the future' [2] as they work outside their own organisation to shape the key uncertainties that lie ahead. A number of approaches may be used to gather information to inform a strategic plan, and some of these are illustrated in Box 37.8. Current drivers in health professions education are listed in Box 37.9.

Having articulated a vision and, through an iterative process, agreed on a number of strategic aims or broad statements of intent, organisational objectives can be defined. To be effective, objectives should be SMART: specific, measurable, achievable, realistic, time bounded. Using project management techniques [117], tactics and action plans can then be developed to meet the stated objectives and plans can be put in place for monitoring and evaluation. Throughout this iterative process, the leader needs to ensure that at all levels, the strategic plan is concordant with the organisation's agreed values. Without a strong backbone of values, a strategic plan becomes meaningless, and commitment will be lacking. In order to keep organisational values fresh and meaningful, they themselves should be reviewed and rearticulated on a regular basis.

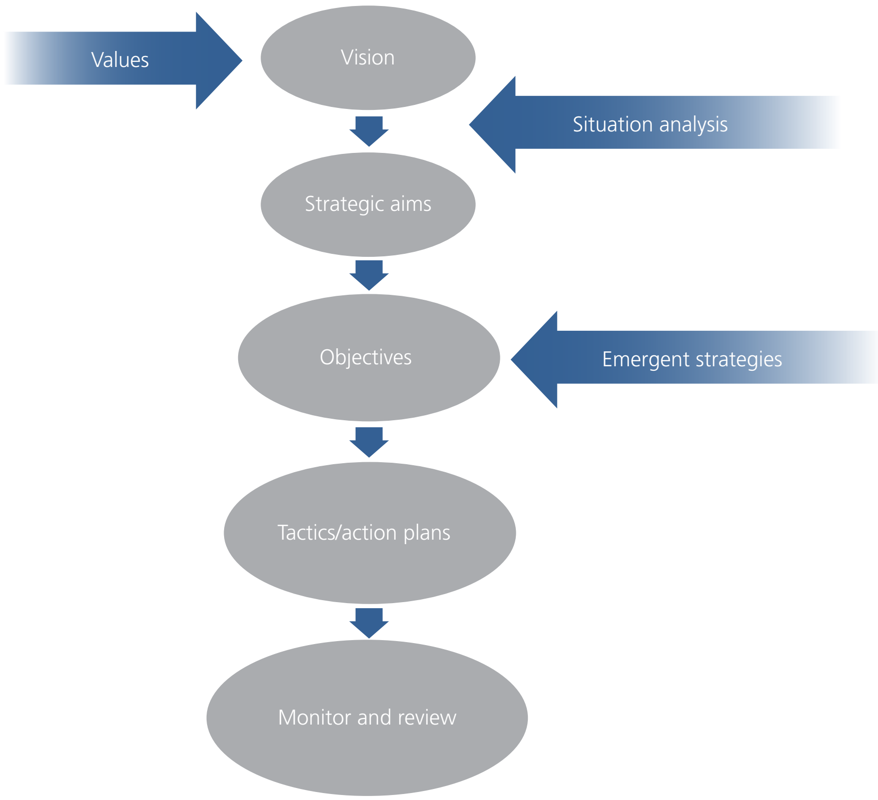


Figure 37.4 Strategic planning.

Leading for Improvement

Donald Berwick, one of the great contemporary thinkers on health care improvement, defined what he called the 'central law of improvement' [120], namely, that *every system is perfectly designed to achieve the results it achieves*. From this, he developed the idea that real improvement can only occur from changing systems, not changing within systems. Effective leaders of improvement then, rather than exhorting or cajoling people to work harder within a system, challenge the very system itself and provide a vision of a superior alternative.

A simple model for achieving change that results in improvement – and change and improvement are not necessarily the same thing – asks three basic questions [121]:

- What are we trying to achieve?
- How will we know if a change leads to an improvement?
- What changes could we make that we think will result in improvement?

There are several leadership elements to these questions. Clarity about what is to be achieved is an absolute obligation on the leader, to set the aims and mission of the organisation or curriculum and to communicate these clearly. The more clearly aims are articulated, the greater the chance of them being realised.

The second of the questions relates to the measurement of outcomes, not as a way of instituting punishment or reward (e.g. star ratings and league tables) but as a means to learn. If we introduce (for example) this new assessment system, how will we know that it makes a difference? How will we know whether a problem-based learning curriculum will lead to better doctors?

Also, in an educational field full of self-determining professionals, a third key leadership task is to provide an environment that encourages innovation and experimentation. Constant small improvements that involve everyone in the organisation expand the available 'gene pool' of ideas and potential solutions. The 'never-ending proposition' of 'coherence making' [96] then becomes everyone's responsibility.

Change-adept organisations empower their members and build capacity through personal and professional development. In a change-responsive organisation, blame is eliminated, innovation is encouraged, and there is an open environment of continuous improvement. In schools, promoting the professional development of colleagues has been found to be central in helping teachers cope with and instigate change [122]. And developing educational expertise and mastery helps emerging leaders to establish credibility. In addition, Lomax [123] suggests that a research perspective is needed, encouraging all staff to take a scholarly and constructively critical approach to their work. In this way, a professional learning community can be developed.

In developing a climate of continuous learning and improvement, we are in effect building a *learning organisation*. Learning organisations concern themselves with learning at all levels and place a high value on both individual and organisational development. As such, they are said to be quicker to adapt to change [124].

BOX 37.8 HOW TO: Develop a strategic plan

A number of tools may assist the process of strategic planning.

SWOT analysis
SWOT enables the organisation to identify its specific competencies, and alongside consideration of its values ensures that there is good alignment between the external environment and internal situation. The mnemonic stands for: the organisation's internal Strengths and Weaknesses, and the Opportunities and Threats that exist in the external environment.

Gap analysis
Gap analysis is done to map the gap that exists between implied and specified future requirements and the existing situation. Gap analysis highlights services and/or functions that have been accidentally left out, deliberately eliminated, or are yet to be developed or procured. The underlying method is simple and revolves around two simple questions:

- Where are we now, i.e. where will we be if we don't change anything?
- Where do we want to be, i.e. what does the environment demand of us?

PESTLE analysis
PESTLE provides a means of identifying possible futures facing an organisation and is useful for broadening perspectives on future strategy by scanning the external environment. The PESTLE headings are: Political, Economic, Social, Technological, Legal, and Environmental.

Porter's 'Five Forces'
Another model, developed by Michael Porter of the Harvard Business School [118], is designed to aid an analysis of the organisation's environment and industry, and the attractiveness of various products or services. The five forces are the:

- risk of new competitors
- threat of substitute products
- bargaining power of buyers
- bargaining power of providers
- degree of competitive rivalry between existing institutions.

Through choosing an appropriate strategy, the organisation can alter the impact of these forces to its advantage, e.g. by collaborating on products (such as a new programme), rather than competing with rivals.

Stakeholder analysis
Stakeholder analysis provides a tool to explore how the views of others might affect the implementation of your strategy. All individuals or groups that might have an interest in your plan are listed and defined as 'opposed', 'neutral', or 'supportive'. How far would you have to move an individual or group for the plan to succeed? It may be enough just to shift some stakeholders to neutral, but essential to obtain the active support of others. How might the plan be presented differently to meet the needs of each group (think politically here) and/or what changes might you have to make in order to make it acceptable or desirable?

Scenario-planning
Traditional forecasting methods often fail to predict significant changes, particularly when the environment is turbulent and unpredictable. Scenario-planning enables us to devise a number of possible futures for the organisation concerned. A strategy can then be developed that would best equip the organisation to survive in those possible worlds. Scenario-planning is best carried out by a group that includes representation from all parts of the organisation and outside, as the idea is to devise scenarios that take into account as many different perspectives as possible.

Improvement as described becomes predominantly a developmental issue and one of getting the organisational culture right. What if things go wrong? As discussed in the section above, rewarding people well, both materially and through positive feedback, is important in terms of motivation. However, there is a flip side to this. The leader who does not intervene when performance is slipping or praises contributions that are substandard, as well as those that are excellent, cheapens their positive regard and will lose the respect of their followers. Without clear and consistent standards to which followers are held accountable, praise and reward will have no value.

Challenges for Health Care Education Leaders

Leading in medical and education, especially at senior level, can lead to particular issues and challenges not encountered by those working in other educational

BOX 37.9 Current drivers in health professions education [119]

- Widening access and inclusivity
- Values-based selection
- Expanding student numbers
- Internationalisation
- Flexibility of programmes
- Technology-enhanced learning
- Increasing accountability for education quality
- Patient safety and quality improvement
- Workforce planning and education for a transformed service
- Reformulation of clinical roles
- Professionalisation of clinical education
- Global workforce challenges
- Financial constraint

disciplines or contexts. Developing and delivering education across professional, subject discipline, and organisational boundaries requires leaders to understand and be willing to work within different organisational structures, systems, funding mechanisms, cultures, norms, and values, and it is often difficult to straddle the clinical–academic 'divide'.

McKimm [125] carried out interviews with a small number of established leaders in medical and health care education in the UK and identified a set of issues and challenges consistently noted by the interviewees. Similar issues were also noted by Lieff and Albert [126] in their research on Canadian medical education leaders.

Personal Issues

- Maintaining an appropriate work–life balance is difficult, especially for those with domestic commitments.
- The culture of senior management practice in both clinical and academic life has impact on career progression for those with domestic responsibilities.
- Some women noted the existence of the 'glass ceiling' and impact of taking career breaks.
- Leaders who had trained as clinicians had to make decisions on how to manage both clinical and academic careers and when/if to leave clinical practice.
- Leaving clinical practice is tied in with maintaining leadership credibility.
- Aspiring leaders with non-traditional career backgrounds feel that they will be overlooked in consideration for senior management positions, especially at medical schools.

Organisational and Cultural Issues

- Many organisational barriers to leadership exist; leaders need to understand the culture and anthropology of their own organisation to succeed.
- Some leaders felt that their profession, discipline, or clinical specialty is perceived in a stereotyped or 'less serious' way by others, and this has impeded career development.
- Clinicians from some specialties may be better able to cope with the dual demands of clinical and academic life than those from others.

Balancing Competing Agendas

- An overwhelming issue identified is working within a rapidly changing and complex health care system.
- Dual demands of working in higher education (which is highly accountable) and the health services (subject to rapid change) put a strain on health care education leaders not found in other sectors of higher education.
- Tensions between management styles, cultures, values, and demands of higher education and health care organisations – a 'crowded stage' with multiple task masters.

The Wider Agenda

- Health care education leaders need to be aware of their influential role in changing and improving health care systems.
- Senior medical education leaders need an accessible forum for discussion of high-level educational change.
- Leaders need to be aware of the wider educational and health care agendas and help drive issues such as interprofessional learning and collaboration, promoting diversity and innovation in educational management and leadership.

The issues and challenges raised by the leaders in the studies highlight that most leaders in medical and health care education will at some point have to make decisions about their leadership direction and whether that lies predominantly in higher education, a clinical setting, undergraduate education, or postgraduate education. Many opportunities exist for leading at various levels and in different contexts in medical education. Becoming aware of the opportunities available and matching this to experience and capabilities are a part of the leadership journey, often revisited many times during a career.

Leadership Development

As Storey points out: 'the public sector has … addressed the leadership development agenda in a particularly high profile and emphatic way' [10]. In parallel with the shifting perspectives on leadership itself from trait and style, through task versus people, and transformational and other 'new paradigm' approaches, to the current position that explores what McDermott [127] calls 'leadership from within' (emphasising reflection, emotions, values, and openness to experience), traditional approaches to learning leadership from 'experts' have given way to experiential and learner-centred approaches. Learning capability is now seen as a key strategic resource, and leadership is seen as central to the development of 'learning organisations' [128].

In addition to the management and leadership programmes offered by universities and the health care

BOX 37.10 HOW TO: Design a leadership development programme

The aim of leadership development can be viewed as 'increasing the capacity of organisations and the people within them to better achieve their purpose' [129]. An underpinning assumption of all such development is that the knowledge, skills, and behaviours required to lead can be learned. Commonly used interventions include courses, seminars and workshops, action-learning, multi-source feedback, coaching and mentoring, simulation, psychometric tools, elearning, and structured work-based experiences. What is important though, is not so much *which* interventions are deployed but *how* they are assembled programmatically. Swanwick and McKimm [13] propose a number of design principles for effective leadership development programmes, namely that they should be practical, work-orientated, supportive of individual development, link theory to practice, and build networks. Broadly the aim is to create a development offer rooted in real-world experience that supports a learner-centred journey of personal learning and transformation. Petrie [130] highlights two types of processes required to achieve this: *horizontal* development – the acquisition of skills, knowledge and new perspectives – and *vertical* development – in which a learner's evolving sense of the world becomes ever more complex. He suggests that vertical development is achieved through three interlinked sets of activities; 'heat experiences' (taking people out of their comfort zone), 'colliding perspectives' (exposing individuals to differing views and opinions), and 'elevated sense-making' (making time for reflection and reframing). The most effective leadership development programmes are embedded within a culture of organisational learning and are linked to human resource practices that value the individual and recognise both performance and potential.

sector, different professional groups have taken different approaches to leadership development, and a range of leadership development programmes are on offer to health care educators. Some of these emphasise the value of leadership development in an interprofessional context (learning from and with other professionals), but most are focused either on clinical leadership or strategic management and leadership development in higher education. Very few programmes exist that are discipline-based or that specifically aim to cross the health care/higher education divide.

Aspiring or current educational leaders can access a range of programmes to support their ongoing professional development. The most effective of these provide a mixture of strategic management and organisational and leadership theory coupled with practical exercises geared to the context in which the leader is or will be working. Many programmes use tools and exercises designed to deepen self-awareness and reflection and encourage support for individuals in the form of action learning sets, coaching, or mentoring. However, effective programmes need also to be longitudinal and oriented around individuals and the organisations in which they work [13]. For this to happen, management and leadership development must be embedded in the culture of every educational organisation. See Box 37.10.

Conclusion

Leadership is one of those concepts that seems very easy to define, until you try to do so. This chapter offers an overview of the way in which leadership theory has developed over the past 50 or so years, illuminating some of the most influential concepts and indicating how these might apply to the health care education leaders of today and tomorrow. We have not had room to do justice to all the ideas surrounding leadership (or management or followership) nor have we explored some of the pressing issues facing leaders of diverse organisations, such as those concerned with gender or race.

We hope that this brief guide to leadership highlights what health care education might require of its leaders of the future: what kind of leader and what attributes, skills, and qualities those leaders might need to demonstrate.

If we are to develop a health workforce capable of delivering high-quality services in all areas, not just in pockets of excellent care, then health care education will require educational leaders at all levels who can manage as well as lead, and who can work effectively and collaboratively across boundaries to ensure equitable and sustainable services, even when resources and morale are low.

References

1 Bush, T. (2003). *Theories of Educational Management and Leadership*. London: Sage.
2 NHS Leadership Academy (2013). Healthcare Leadership Model. http://www.leadershipacademy.nhs.uk/resources/healthcare-leadership-model (accessed 15 March 2017).
3 Bennis, W. and Nanus, N. (1985). *Leaders: The Strategies for Taking Charge*. New York: Harper & Row.
4 Fayol, H. (1949). *General and Industrial Management*. London: Pitman Publishing.
5 Yukl, G. (2002). *Leadership in Organisations*, 5e. Upper Saddle River, NJ: Prentice Hall.
6 Kotter, J.P. (1990). What leaders really do. *Harvard Business Review* 68 (3): 103–111.
7 Covey, S., Merrill, A.R., and Merrill, R.R. (1994). *First Things First*. New York: Simon & Schuster.
8 Northouse, P. (2012). *Leadership: Theory and Practice*, 6e. London: Sage.
9 Blagg, D. and Young, S. (2001). What makes a good leader? Harvard Business School Bulletin http://www.alumni.hbs.edu/bulletin/2001/february/leader.html (accessed 9 June 2013).
10 Storey, J. ed. (2004). *Leadership in Organisations: Current Issues and Key Trends*. London: Routledge.

11 Uhl-Bien, M., Riggio, R.E., Lowe, K.B., and Carsten, M.K. (2014). Followership theory: a review and research agenda. *The Leadership Quarterly* 25: 83–104.
12 Kelley, R.E. (2008). Rethinking followership. In: *The Art of Followership: How Great Followers Create Great Leaders and Organizations* (ed. R.E. Riggio, I. Chaleff and J. Lipman-Blumen). San Francisco: Jossey-Bass.
13 Swanwick, T. and McKimm, J. (2013). Faculty development for leadership and management. In: *Faculty Development in the Health Professions* (ed. Y. Steinert). New York: Springer Publications.
14 Goleman, D. (2000). Leadership that gets results. *Harvard Business Review* 78 (2): 78–90.
15 Judge, T.A., Bono, J.E., Ilies, R., and Gerhardt, M.W. (2002). Personality and leadership: a qualitative and quantitative review. *Journal of Applied Psychology* 87 (4): 765–780.
16 Blake, R. and Mouton, J. (1964). *The Managerial Grid*. Houston, TX: Gulf.
17 Tannenbaum, R. and Schmidt, W. (1958). How to choose a leadership pattern: should a leader be democratic or autocratic or something in between? *Harvard Business Review* 36: 95–101.
18 Fiedler, F. (1964). A contingency model of leadership effectiveness. In: *Advances in Experimental Social Psychology* (ed. L. Berkowitz), 149–190. New York: Academic Press.
19 Hersey, P. and Blanchard, K. (1988). *Management of Organizational Behaviour*. Englewood Cliffs, NJ: Prentice Hall.
20 Adair, J. (1973). *Action-Centred Leadership*. New York: McGraw-Hill.
21 Bass, B. and Avolio, B. (1994). *Improving Organizational Effectiveness through Transformational Leadership*. Thousand Oaks, CA: Sage.
22 Kellerman, B. (2007). What every leader needs to know about followers. *Harvard Business Review* 85 (12): 84–91.
23 Vanderslice, V. (1988). Separating leadership from leaders: an assessment of the effect of leader and follower roles. *Human Relations* 41: 677–696.
24 Alimo-Metcalfe, B. and Alban-Metcalfe, J. (2006). More (good) leaders for the public sector. *International Journal of Public Sector Management* 19 (4): 73–95.
25 Bolden, R. (2010). Leadership, management and organisational development. In: *Gower Handbook of Leadership and Management Development*, 5e (ed. R. Thorpe, J. Gold and A. Mumford), 117–132. Farnham: Gower.
26 McKimm, J. (2017). Collaborative leadership and partnership working. In: *ABC of Clinical Leadership*, 2e (ed. T. Swanwick and J. McKimm). Oxford: Wiley-Blackwell.
27 King's Fund and Center for Creative Leadership (2014). Developing collective leadership for health care. London: King's Fund. www.kingsfund.org.uk/publications/developing-collective-leadership-health-care (accessed 16 March 2017).
28 Mennin, S. (2010). Self-organisation, integration and curriculum in the complex world of medical education. *Medical Education* 44: 20–30.
29. Fullan, M. (2001). *Leading in a Culture of Change*. San Francisco, CA: Jossey-Bass.
30 Western, S. (2011). An overview of leadership discourses. In: *Educational Leadership: Context, Strategy and Collaboration* (ed. M. Preedy, N. Bennett and C. Wise). Milton Keynes: The Open University.
31 Bryman, A. (1996). Leadership in organisations. In: *Handbook of Organisational Studies* (ed. S.R. Clegg, C. Harvey and W.R. Nord), 276–292. London: Sage.
32 McKimm, J., Forrest, K., and Thistlethwaite, J. ed. (2017). Educational leadership. In: *Medical Education at a Glance*. Chichester: Wiley Blackwell.
33 Souba, W.W. (2016). Resilience – back to the future. *JAMA Surgery* 10: 896–897.
34 Goleman, D. (1996). *Emotional Intelligence*. London: Bloomsbury.
35 Duckworth, A. (2016). *Grit: The Power of Passion and Perseverance*. New York: Scribner.
36 Boyd, W. (1992). The power of paradigms: reconceptualising educational policy and management. *Educational Administration Quarterly* 28 (4): 504–528.
37 House, E. (1981). Three perspectives on innovation: technological, political and cultural. In: *Improving Schools: Using What We Know* (ed. R. Lehming and M. Kane), 42–114. Beverley Hills, CA: Sage.
38 Morgan, G. (1998). *Images of Organization: The Executive Edition*. San Francisco, CA/Thousand Oaks, CA: Berrett-Kochler Publications/Sage.
39 Bush, T. (1999). Crisis or crossroads? The discipline of educational management in the late 1990s. *Educational Management and Administration* 27 (3): 239–252.
40 Hoyle, E. (1986). *The Politics of School Management*. London: Hodder & Stoughton.
41 Bush, T. and Glover, D. (2002). *School Leadership: Concepts and Evidence*. Nottingham: National College for School Leadership.
42 Leithwood, K., Jantzi, D., and Steinbach, R. (1999). *Changing Leadership for Changing Times*. Buckingham: Open University Press.
43 Weber, M. (1947). *The Theory of Social and Economic Organization*. New York: Free Press.
44 Handy, C. (1995). *Gods of Management: The Changing Work of Organizations*. New York: Doubleday.
45 Lungu, G. (1985). In defense of bureaucratic organization in education. *Educational Management and Administration* 13: 172–178.
46 Osborne, A. (1990). The nature of educational management. In: *Education Management for the 1990s* (ed. J. West-Burnham), 9–10. Harlow: Longman.
47 Bush, T. (1995). *Theories of Educational Management*, 2e. London: Paul Chapman.
48 Mintzberg, H. (1992). *Structure in Fives: Designing Effective Organisations*. Harlow: Prentice Hall.
49 Bush, T. (2002). Educational management: theory and practice. In: *The Principles and Practice of Education Management* (ed. T. Bush and L. Bell), 15–31. London: Paul Chapman.
50 Hargreaves, A. (1994). *Changing Teachers, Changing Times: Teachers' Work and Culture in a Postmodern Age*. London: Cassell.
51 Baldridge, J.V., Curtis, D.V., Ecker, G., and Riley, G.L. (1978). *Policy Making and Effective Leadership*. San Francisco, CA: Jossey-Bass.
52 Coulson, A. (1986). *The Managerial Work of Headteachers*. Sheffeld: Sheffield City Polytechnic.
53 Alimo-Metcalfe, B. (1998). *Effective Leadership*. London: Local Government Management Board.
54. Grint, K. (1995). *Management: A Sociological Introduction*. Oxford: Polity Press.
55 Maccoby, M. (2000). Narcissistic leaders: the incredible pros, the inevitable cons. *Harvard Business Review* 78 (1): 69–77.
56 Fullan, M. (2001). *The New Meaning of Educational Change*, 3e. London: Routledge Falmer.
57 Bennett, N., Wise, C., Woods, P.A., and Harvey, J.A. (2003). Distributed leadership: a review of literature. *National College for School Leadership* Spring: 1–57. http://oro.open.ac.uk/8534/1/bennett-distributed-leadership-full.pdf (accessed 26 June 2018).
58 Gronn, P. (2008). The future of distributed leadership. *Journal of Educational Administration* 46 (2): 141–158.
59 Bolden, R., Petrov, G., and Gosling, J. (2009). Distributed leadership in higher education: rhetoric and reality. *Educational Management Administration and Leadership* 37: 2.
60 Turning Point Leadership Development National Excellence Collaborative (2002). Academics and Practitioners on Collaborative Leadership. http://www.turningpointprogram.org/Pages/pdfs/lead_dev/LDC_panels_lowres.pdf (accessed 9 June 2013).
61 Van Zwanenberg, Z. (2003). *Modern Leadership for Modern Services*. Scottish Leadership Foundation, Alloa.
62 Argyris, C. (1964). *Integrating the Individual and the Organization*. New York: Wiley.
63 Ball, S. (1987). *The Micropolitics of the School: Towards a Theory of School Organisation*. London: Methuen.

64 Hales, C. (1997). Power, authority and influence. In: *Organizational Effectiveness and Improvement in Education* (ed. A. Harris, M. Preedy and N. Bennett), 22–30. Buckingham: Open University Press.
65 Handy, C. (1993). *Understanding Organisations*, 4e. London: Penguin Books.
66 Kotter, J. (1996). *Leading Change*. Boston: Harvard Business School Press.
67 Weber, M. (1998). The types of legitimate domination. In: *Economy and Society* (ed. G. Roth and C. Wittich), 215–254. Berkeley, CA: University of California Press.
68 Keough, T. and Tobin, B. (2001). Postmodern leadership and the policy lexicon: from theory, proxy to practice. Paper presented at the Pan-Canadian Education Research Agenda Symposium, Quebec.
69 Weick, K. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly* 21 (1): 1–19.
70 Cohen, M.D. and March, J.G. (1976). *Leadership and Ambiguity: The American College President*. Boston: Harvard Business School Press.
71 Gilbert, P. (2005). *Leadership: Being Effective and Remaining Human*. Lyme Regis: Russell House Publishing.
72 Peck, E., Dickinson, H., and Smith, J. (2006). Transforming or transacting? The role of leaders in organisational transition. *British Journal of Leadership in Public Services* 2 (3): 4–14.
73 Deal, T.E. (1985). The symbolism of effective schools. *Elementary School Journal* 85 (5): 601–620.
74 Schein, E.H. (1985). *Organizational Culture and Leadership*. San Francisco: Jossey-Bass Publishers.
75 Greenleaf, R.K. (1977). *Servant Leadership: A Journey into the Nature of Legitimate Power and Greatness*. Mahwah, NJ: Paulist Press.
76 Collins, J. (2001). *Good to Great*. London: Random House.
77 Hollander, E.P., Park, B.B., and Elman, B. (2008). Inclusive Leadership and Leader-Follower Relations: Concepts, Research, and Applications. The Member Connector, International Leadership Association.
78 Morrow, C. (2014). The Linkage Inclusive Leadership Model. http://www.diversityjournal.com/13313-moving-dial-measuring-inclusive-leadership (accessed 7 February 2017).
79 Bligh, J. and Brice, J. (2007). The academy of medical educators: a professional home for medical educators in the UK. *Medical Education* 41 (7): 625.
80 Alimo-Metcalfe, B. and Alban-Metcalfe, J. (2004). Leadership in public sector organisations. In: *Leadership in Organisations: Current Issues and Key Trends* (ed. J. Storey), 173–202. London: Routledge.
81 Manz, C.C. and Sims, H.P. (1991). Super-leadership: beyond the myth of heroic leadership. *Organisational Dynamics* 19: 18–35.
82 Performance and Innovation Unit (2000). Strengthening Leadership in the Public Sector. http://www.cabinet-office.gov.uk/innovation (accessed 11 September 2006).
83 West, M.A. and Lyubovnikova (2013). Illusions of team working in healthcare. *Journal of Health Organization and Management* 27 (1): 134–142.
84 Kozlowski, S.W.J., Gully, S.M., Salas, E., and Cannon-Bowers, J.A. (1995). Team leadership and development: theory, principles, and guidelines for training leaders and teams. Third Symposium on Work Teams. Dallas, TX: University of North Texas.
85 Tuckman, B. (1965). Developmental sequence in small groups. *Psychological Bulletin* 63: 384–389.
86 Belbin, R.M. (1991). *Management Teams: Why They Succeed or Fail*. London: Heinemann.
87 Morrison, K. (1998). *Management Theories for Educational Change*. London: Paul Chapman.
88 Handy, C. (1989). *The Age of Unreason*. London: Business Books.
89 Iles, V. and Sutherland, K. (2001). *Organisational Change: A Review for Healthcare Managers, Researchers and Professionals*. London: National Coordinating Centre for NHS Service Delivery and Organisation.
90 Mintzberg, H. (1987). Crafting strategy. *Harvard Business Review* 65 (4): 66–75.
91 Lewin, K. (1958). Group decisions and social change. In: *Readings in Social Psychology* (ed. G. Swanson, T. Newcomb and L. Hartley), 459–473. New York: Rinehart & Winston.
92 Bullock, R. and Batten, D. (1985). It's just a phase we're going through: a review and synthesis of OD phase analysis. *Group & Organization Studies* 10: 383–412.
93 Buchanan, D. and Boddy, D. (1992). *The Expertise of the Change Agent*. Harlow: Prentice Hall.
94 Bolman, L.G. and Deal, T.E. (2013). *Reframing Organisations*, 5e. San Francisco: Jossey-Bass.
95 Bank, J. (1992). *The Essence of Total Quality Management*. Harlow: Prentice Hall.
96 Fullan, M. (2003). *Change Forces with a Vengeance*. London: Routledge Falmer.
97 Baker, P., Curtis, D., and Beneson, W. (1991). Structures and processes of planned change. *The School Community Journal* Fall/Winter (1): 11–19.
98 Gleick, J. (1999). *Faster: The Acceleration of Just about Everything*. New York: Pantheon.
99 Harvey, D. (1989). *The Condition of Postmodernity*. Cambridge, MA: Polity Press.
100 Jameson, F. (1991). *Postmodernism, or, the Cultural Logic of Late Capitalism*. London: Verso.
101 Castells, M. (1996). *The Rise of the Network Society*. Oxford: Blackwell.
102 Hargreaves, A., Earl, L., and Schmidt, M. (2002). Perspectives on alternative assessment reform. *American Educational Research Journal* 39 (1): 69–95.
103 Battram, A. (1998). *Navigating Complexity*. London: The Industrial Society.
104 Schön, D. (1983). *The Reflective Practitioner: How Professionals Think in Action*. New York: Basic Books.
105 Lewin, R. and Regine, B. (2000). *The Soul at Work*. New York: Simon & Schuster.
106 Plesk, P. and Greenhalgh, T. (2001). The challenge of complexity in health care. *British Medical Journal* 323: 625–628.
107 Kernick, D. (2002). Complexity and healthcare organisation. In: *Complexity and Healthcare: An Introduction* (ed. K. Sweeney and F. Griffiths), 93–121. Oxford: Radcliffe.
108 Waldrop, M. (1992). *Complexity: The Emerging Science at the Edge of Order and Chaos*. Harmondsworth: Penguin.
109 Checkland, P. (1981). *Systems Thinking, Systems Practice*. Oxford: Wiley.
110 Stacey, R. (2001). *Complex Responsive Processes in Organisations*. London: Routledge.
111 Axelrod, R. and Cohen, M.D. (2000). *Harnessing Complexity*. New York: Basic Books.
112 Johnson, G. and Scholes, K. (1997). *Exploring Corporate Strategy*. Harlow: Prentice-Hall.
113 Mintzberg, H., Ahlstrand, B., and Lampel, J. (1998). *Strategy Safari: A Guided Tour through the Wilds of Strategic Management*. New York: Free Press.
114 Pascale, R., Milleman, M., and Gioja, L. (2000). *Surfing the Edge of Chaos*. New York: Crown Business Publishing.
115 Fraser, S. and Greenhalgh, T. (2001). Coping with complexity: educating for capability. *British Medical Journal* 323: 799–803.
116 Claxton, G. (1997). *Hare Brained and Tortoise Mind*. London: Fourth Estate.
117 Nokes, S., Greenwood, A., and Goodman, M. (2003). *The Definitive Guide to Project Management: The Fast Track to Getting the Job Done on Time and on Budget (Financial Times Series)*. Edinburgh: Pearson Education.
118 Porter, M.E. (1980). *Competitive Strategy: Techniques for Analyzing Industries and Competitors*, 1e. New York: The Free Press.
119 McKimm, J. and Swanwick, T. (2017). Educational leadership. In: *ABC of Clinical Leadership*, 2e. Oxford: Wiley-Blackwell.

120 Berwick, D. (1996). A primer on the improvement of systems. *British Medical Journal* 312: 619–622.

121 Langley, G.J., Nolan, K.M., and Nolan, T.W. (1992). *The Foundation of Improvement*. Silver Spring, MD: API Publishing.

122 Moyles, J., Suchuitsky, W., and Chapman, L. (1998). *Teaching Fledglings to Fly: Report on Mentoring in Primary Schools*. London: Association of Teachers and Lecturers.

123 Lomax, P. (1990). *Managing Staff Development in Schools: An Action Research Approach, BERA Dialogues 3*. Clevedon: Multilingual Matters.

124 Senge, P. (1990). *The Fifth Discipline: The Art and Practice of the Learning Organization*. New York: Doubleday.

125 McKimm, J. (2004). *Case Studies in Leadership in Medical and Health Care Education: special report 5*. Newcastle upon Tyne: Higher Education Academy Subject Centre for Medicine, Dentistry and Veterinary Medicine.

126 Lieff, S. and Albert, M. (2010). The mindsets of medical education leaders; how do they conceive of their work? *Academic Medicine* 85 (1): 57–62.

127 McDermott, G.R. (1994). Partnering with god: ignatian spirituality and leadership in groups. In: *Spirituality at Work: Discovering the Spirituality in Leadership* (ed. J.A. Conger), 132–161. San Francisco, CA: Jossey-Bass.

128 Drath, W.H. and Palus, C.J. (1994). *Making Common Sense: Leadership as Meaning Making in a Community of Practice*. Greensboro, NC: Center for Creative Leadership.

129 Bolden, R. (2005). What is Leadership Development? Purpose and practice. Leadership South West Research Report 2. http://business-school.exeter.ac.uk/documents/discussion_papers/cls/LSWreport2.pdf (accessed 24 December 2015).

130 Petrie, N. (2014). *Future Trends in Leadership Development*. Centre for Creative Leadership http://insights.ccl.org/articles/white-papers/future-trends-in-leadership-development-2 (accessed 2 April 2016).

Further Reading

Fullan, M. (2005). *Leadership and Sustainability: Systems Thinkers in Action*. Thousand Oaks, CA: Corwin Press (Sage).

Heifetz, R.A., Grashow, A., and Linsky, M. (2009). *The Practice of Adaptive Leadership: Tools and Tactics for Changing your Organisation and the World*. Harvard Business Press.

Hoyle, E. and Wallace, M. (2005). *Educational Leadership: Ambiguity, Professionals and Managerialism*. London: Sage.

Kegan, R. and Lahey, L.L. (2016). *An Everyone Culture: Becoming a Deliberately Developmental Organisation*. Boston, MA: Harvard Business School Press.

Kydd, L., Anderson, A., and Newton, W. ed. (2003). *Leading People and Teams in Education*. London: Sage.

Northouse, P. (2012). *Leadership: Theory and Practice*, 6e. London: Sage.

Preedy, M., Bennet, N., and Wise, C. (2012). *Educational Leadership: Context, Strategy and Collaboration*. Milton Keynes: Sage Publications/Open University.

Storey, J. ed. (2016). *Leadership in Organisations: Current Issues and Key Trends*, 3e. London: Routledge.

Index

Understanding Medical Education: Evidence, Theory, and Practice, Third Edition. Edited by Tim Swanwick, Kirsty Forrest and Bridget C. O'Brien.