

Practical Guide to the Evaluation of Clinical Competence

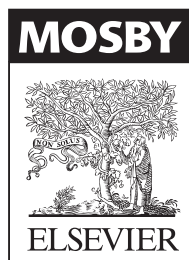
Practical Guide to the Evaluation of Clinical Competence

Eric S. Holmboe, MD, FACP

Senior Vice President for Quality Research and Academic Affairs
American Board of Internal Medicine
Philadelphia, Pennsylvania
Adjunct Professor of Medicine
Yale University School of Medicine
New Haven, Connecticut

Richard E. Hawkins, MD, FACP

Vice President, Assessment Programs
National Board of Medical Examiners
Philadelphia, Pennsylvania



MOSBY
ELSEVIER

1600 John F. Kennedy Blvd.
Suite 1800
Philadelphia, PA 19103-2899

PRACTICAL GUIDE TO THE EVALUATION OF
CLINICAL COMPETENCE

ISBN: 978-0-323-04709-8

Copyright © 2008 by Eric S. Holmboe and National Board of Medical Examiners.
Published by Mosby, Inc. All rights reserved.

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Permissions may be sought directly from Elsevier's Rights Department: phone: (+1) 215 239 3804 (US) or (+44) 1865 843830 (UK), fax: (+44) 1865 853333, e-mail: healthpermissions@elsevier.com. You may also complete your request on-line via the Elsevier homepage (<http://www.elsevier.com>), by selecting 'Customer Support' and then 'Obtaining permissions.'

Notice

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our knowledge, changes in practice, treatment and drug therapy may become necessary or appropriate. Readers are advised to check the most current information provided (i) on procedures featured or (ii) by the manufacturer of each product to be administered, to verify the recommended dose or formula, the method and duration of administration, and contraindications. It is the responsibility of the practitioner, relying on their own experience and knowledge of the patient, to make diagnoses, to determine dosages and the best treatment for each individual patient, and to take all appropriate safety precautions. To the fullest extent of the law, neither the Publisher nor the Editors assume any liability for any injury and/or damage to persons or property arising out of or related to any use of the material contained in this book.

Library of Congress Cataloging-in-Publication Data

Practical guide to the evaluation of clinical competence / [edited by]
Eric S. Holmboe, Richard E. Hawkins.—1st ed.
p. ; cm.

Includes bibliographical references and index.

ISBN 978-0-323-04709-8

1. Clinical competence—Evaluation. 2. Clinical medicine—Ability testing. I. Holmboe, Eric S. II. Hawkins, Richard E.

[DNLM: 1. Clinical Competence. 2. Internship and Residency. 3. Education, Medical, Graduate—standards. 4. Educational Measurement—methods. 5. Fellowships and Scholarships. W 20 P895 2008]

R837.A2P73 2008

616—dc22

2007018058

Acquisitions Editor: James Merritt
Developmental Editor: Nicole DiCicco
Electronic Developmental Editor: Carol Emery
Project Manager: Bryan Hayward
Design Direction: Gene Harris

Printed in the United States of America.

Last digit is the print number: 9 8 7 6 5 4 3 2 1

Working together to grow
libraries in developing countries
www.elsevier.com | www.bookaid.org | www.sabre.org

ELSEVIER

BOOK AID
International

Sabre Foundation

Preface

Evaluation of medical trainees is hard but invaluable work. Evaluation of clinical competence is a core element of professionalism and underlies effective self-regulation; it is essential to fulfilling our professional obligation to assure the public that the graduates of medical training programs are competent to enter the next stage of education and/or practice. As medical educators, it is also important that we develop and use high-quality evaluation methods and systems in order to fulfill a primary obligation to our students, residents, and fellows. Here, effective evaluation provides the feedback and guidance to support their professional growth and development.

We have spent much of the last 15 years of our professional lives thinking, learning, and then teaching, about evaluation. We have had the good fortune to participate in developing or refining evaluation methods and systems across the spectrum of education and practice. Like many of you, much of our initial learning was largely through trial and error, occurring as a result of being assigned positions of responsibility in determining the competence of students and residents. Subsequently, we adopted a more deliberative approach to developing and implementing evaluation methods, and the slope of our learning curves began to increase. Our efforts to appraise the impact and effectiveness of new or enhanced evaluation methods helped us gain additional knowledge and understanding of evaluation, more so than did review of the relatively limited amount of medical literature on resident evaluation available at the time.

The impression that our knowledge and understanding about evaluation was advanced more significantly through “on-the-ground” development and measurement of evaluation approaches than through reading the literature led us to believe that sharing our experiences and knowledge might be valuable to others in the community. We started by giving faculty development workshops at various professional conferences and began to write about our experience. The organization of this book, including most of the chapters, reflects the evolving structure and content of the materials developed for our faculty development workshops.

Although originating from, and built on, previous faculty development work, this book also reflects our own professional growth and enhanced understanding of evaluation. Our discussions and collaboration with experts in the medical education and evaluation community helped us become more comfortable in applying knowledge and the “science of evaluation” principles to the day-to-day challenges confronting clerkship and program directors.

The primary purpose of this book is to provide a *practical* guide to developing evaluation systems. However, we hope that it also serves as a useful resource through inclusion of the underlying evidence base to support recommendations and facilitate understanding. The book has been organized around the various evaluation tools and how individuals with responsibilities for evaluation can apply these tools in their own setting. Each chapter provides information on the strengths and weaknesses of the evaluation method, along with information about specific tools. However, no single tool can do the whole job; effective evaluation requires a multifaceted approach.

Effective evaluation also depends upon collaboration among a *team* of faculty and other educators; thus any change to an evaluation system must include not only buy-in from others, but also the investment to train educators to use evaluation tools effectively. Evaluation tools are only as good as the individual using them. If done well, evaluation can have a profoundly positive effect on patients, trainees, and faculty. Nothing can be more satisfying than knowing each and every one of your graduates is truly ready to move to the next career level. The public expects no less, and we should expect no less from ourselves. In that spirit, we welcome comments from you the reader on how we can improve upon this book.

*Eric S. Holmboe
Richard E. Hawkins*

Contributors

Robert McL. Boote, BS, JD

Ballard, Spahr, Andrews & Ingersoll, LLP
Philadelphia, Pennsylvania

John R. Boulet, PhD

Assistant Vice President
Research and Evaluation
Educational Commission for Foreign
Medical Graduates
Philadelphia, Pennsylvania

Carol Carraccio, MD, MA

Associate Chair for Education
Department of Pediatrics
University of Maryland School of Medicine
Baltimore, Maryland

Brian E. Clauser, EdD

Associate Vice President
Measurement Consulting Services
National Board of Medical Examiners
Philadelphia, Pennsylvania

Stephen G. Clyman, MD

Executive Director
Center for Innovation
National Board of Medical Examiners
Philadelphia, Pennsylvania

Margery H. Davis, MD

Professor of Medical Education
Director of the Centre for Medical
Education
University of Dundee
Dundee, Scotland

F. Daniel Duffy, MD, MACP

Professor of Medicine
University of Oklahoma
Norman, Oklahoma
Former Executive Vice President
American Board of Internal Medicine
Philadelphia, Pennsylvania

Michael L. Green, MD, MSc

Associate Professor of Medicine
Yale University School of Medicine
Associate Program Director for Ambulatory
Education
Yale Primary Care Internal Medicine
Program
New Haven, Connecticut

Richard E. Hawkins, MD, FACP

Vice President
Assessment Programs
National Board of Medical Examiners
Philadelphia, Pennsylvania

Eric S. Holmboe, MD, FACP

Senior Vice President for Quality Research
and Academic Affairs
American Board of Internal Medicine
Philadelphia, Pennsylvania
Adjunct Professor of Medicine
Yale University School of Medicine
New Haven, Connecticut

S. Barry Issenberg, MD, FACP

Associate Professor of Medicine
University of Miami Miller School of
Medicine
Gordon Center for Research in Medical
Education
Miami, Florida

Jocelyn M. Lockyer, MHA, PhD

Professor
Department of Community Health Sciences
Associate Dean
Continuing Medical Education and
Professional Development
University of Calgary
Calgary, Canada

Catherine R. Lucey, MD, FACP

Professor of Medicine
Vice Dean for Education
College of Medicine
Ohio State University
Columbus, Ohio

Melissa J. Margolis, MS

Measurement Scientist
Measurement Associates
National Board of Medical
Examiners
Philadelphia, Pennsylvania

John J. Norcini, PhD

President and CEO
Foundation for Advancement of
International Medical Education and
Research
Philadelphia, Pennsylvania

Louis Pangaro, MD

Professor and Vice Chair
Department of Medicine
Uniformed Services University of the
Health Sciences
Bethesda, Maryland

Ross J. Scalese, MD

Assistant Professor of Medicine
University of Miami Miller School of
Medicine
Gordon Center for Research in Medical
Education
Miami, Florida

David B. Swanson

Deputy Vice President
Professional Services
National Board of Medical Examiners
Philadelphia, Pennsylvania

Acknowledgments

Much love to my wife of 25 years, Eileen Holmboe, and for supporting me through many career turns and moves.

Gratitude and love to my children, Ken and Lauren, who have to put up with their all too often “part-time” dad.

Thanks and love for Mom, Dad, Kristi, and Kevin.

Eric Holmboe

Much love and gratitude to my wife, Marguerite Hawkins, and my mother, Jacqueline Hawkins, for their support over many years.

Richard Hawkins

We both would like to sincerely thank the many trainees with whom we worked over the years. More than any other group, we learned the most working with you. Your dedication and cooperation were much appreciated.

*Eric Holmboe
Richard Hawkins*

Faculty Guidelines to Training DVD

This DVD provides a set of video recorded trainee-patient encounters intended to facilitate faculty development programs focusing on improving direct observation and feedback skills. These tapes can be used in conjunction with the training methods discussed in Chapter 9.

This DVD provides brief descriptions of training scenarios. The scenarios are scripted to depict varying levels of trainee proficiency. Trainee deficiencies are demonstrated in each level—none are scripted to be “perfect.” There are three scenarios for each clinical skill of medical interviewing, physical examination, and counseling; each successive scenario displays progressively better performance. Following a brief introduction regarding the context of the trainee-patient encounter, the deficiencies, or “errors,” for each clinical encounter scenario are outlined. It is important to note that the deficiencies range from minor to more severe; the judgment of the faculty is an important component of the training exercises. You can use this guide to help you discuss and calibrate your own faculty in workshops. We recommend you review the scenarios carefully on your own with the guidelines to become comfortable with the clinical encounters before using them in workshops with your own faculty.

How Faculty Should Conduct an Effective Observation

A separate video is provided to illustrate important concepts in direct observation. The focus of this scenario is on the faculty physician performing the observation of a resident or student. This scenario provides multiple examples of what *not* to do as a faculty member in performing an observation. A number of points should be noted about this direct observation by the faculty attending:

1. The faculty member is late, disrupting the flow of the trainee-patient encounter.
2. He does a poor job of explaining to the patient what his role will be during the observation, and was not explicit with the trainee about what he will do during the physical examination.
3. His positioning to observe the physical examination is poor; he is seated behind the patient and in front of the trainee. This makes it very difficult for him to see if the physical examination maneuvers were

done correctly, and having the observer right in front of her may be distracting to the trainee.

4. He disrupts the blood pressure measurement by trying to wash his hands during the measurement.
5. He disrupts the eye examination by moving around and inserting himself into the examination process. One reason this may have distracted the trainee was the lack of explanation by the faculty physician on what he would do during the examination (see item 2 above).
6. He further disrupts the examination by asking the patient questions during the trainee’s examination of the lungs. This could have been avoided had the faculty physician reviewed the patient’s history and presenting complaint with the trainee before observing the physical examination. The medical history presentation could have been done at the bedside. This then cues the faculty physician on what he should be looking for during the physical examination.
7. The faculty physician is distracted by a knock at the door. This distraction causes him to miss a critical component of the examination; the cardiac examination in a patient with a presenting complaint of syncope.

These are the major take-home points from this direct observation encounter. Some basic principles for effective direct observation are as follows:

- Prepare for the observation.
 - Faculty: Know what you are looking for.
 - Resident: Let him/her know what to expect.
 - Patient: Let him/her know why you are there.
- Minimize intrusiveness—correct positioning using “triangulation,” when possible.
- Minimize interference with the trainee-patient interaction.
- Avoid distractions.

Medical Interviewing Tapes (Scenarios 1–3)

Context

The patient presents with chest discomfort that started early in the morning. The setting is an emergency department and the resident is performing the history

to decide whether the patient requires admission. This patient has several “key features” in his history: exertional nature of the pain and relief with rest; prior symptoms consistent with angina; and positive risk factors. These findings, along with past history, strongly suggest coronary artery disease.

Facilitator Notes

The videos show progressively better information collection (1=poor, 3=best of the three, but still with deficiencies), what may be referred to as the “doctor-centered” portion of the medical history. Suboptimal in all of the videos are the “patient-centered” aspects of the medical interview such as responding to concerns, body language, and so on. Video 3 is a little better in that the trainee starts with an open-ended question, asks for patient’s questions, and so forth.

Physical Examinations (Scenarios 4–6)

Context

A new patient presents to an urgent care clinic with a 2-day history of productive cough and mild shortness of breath. He states he has felt warm at home. He has difficulty performing activities of daily living without dyspnea on exertion. He has an 80 pack-year smoking history and still smokes 1 pack of cigarettes per day. He was told he had a “heart murmur” some years ago, but that no further evaluation was needed. He denies a history of heart attack or angina. He also denies chest pain or paroxysmal nocturnal dyspnea. This patient has not received regular medical care since moving to a retirement apartment complex 4 years ago. His only surgical history was an appendectomy as a child. He denies any change in his stools or bowel habits, but has noted a 5-pound weight loss in the last 6 months. He notes a little swelling in his ankles that usually goes down overnight. His only medication is a multivitamin and a daily aspirin when he “remembers to take them.” BP taken by nurse is 110/72 mm Hg and temperature is 101.5°F.

Physical Examination

1. Components of examination
 - a. Blood pressure measurement
 - b. HEENT examination
 - Should examine the throat and ears
 - Should examine the mucosa for pallor (eyes, mouth at minimum)—assessment for signs of anemia
 - c. Neck examination—should assess for adenopathy (assessment of nodes should cover all important lymphatic areas to include supraclavicular area)
 - d. Pulmonary exam—should assess for pneumonia and pleural effusion; should at minimum percuss and auscultate

- e. Cardiac examination including assessment of jugular venous pressure
- f. Assessment of extremities

Facilitator Notes

This patient has a right middle lobe pneumonia. Look particularly for proper examination by the trainee of the right middle lobe.

Counseling (Scenarios 7–9)

Context

This female patient has hyperlipidemia and has failed a trial of weight loss and dietary modification with no change in cholesterol level. Recent laboratory test results follow: total cholesterol is 285 mg/dL; High-density lipoprotein level is 45 mg/dL; triglyceride level is 170 mg/dL; and low-density lipoprotein level is 206 mg/dL. She returns to office today to start statin therapy. Other cardiovascular risk factors are age (>45 years); hypertension; father had heart attack at age 53; currently smoking.

Important issues for counseling are as follows:

- a. Discussion of patient’s role in decision making
- b. Discussion of the clinical issue or nature of the decision
- c. Discussion of the alternatives
- d. Discussion of pros and cons of the choices
- e. Discussion of the uncertainties with the decision
- f. Assessment of patient’s understanding
- g. Exploration of patient preference

Facilitator Notes

These videos can be used to highlight important components of informed decision making for a very common scenario. They could be useful as a way to incorporate the new Adult Treatment Panel (ATP) cholesterol guidelines to discuss how the feedback to the resident on these tapes could incorporate evidenced-based medicine principles. The video also demonstrates emotion from the patient—How should one respond? The video could be used to facilitate discussion of the NURS model (Name the emotion/reflect Understanding/show Respect/give Support) of responding to emotions from patients.

The Core Elements of Informed Decision Making

Level of Decision Making Required	Component
Basic (e.g., ordering a lab test)	1, 2, 7
Intermediate (e.g., prescribing new medication)	1, 2, 3, 4, 6, 7
Complex (e.g., invasive procedure)	All 7

Braddock Core Informed Decision-Making Components

Rate each required component performed by the trainee.

1. Discussion of patient's role in decision making
Rationale: Many patients are not aware that they can and should participate in decision making.
Example: "I'd like us to make to make this decision together."
2. Discussion of the clinical issue or nature of the decision
Rationale: A clear statement of what is at issue helps to clarify what is being decided and allows the physician to share some of her/his thinking about it.
Example: "This is a medication that would help with"
3. Discussion of alternatives
Rationale: A decision is always a choice among certain options, including doing nothing at all.
Example: "You could try the new medication or continue the one you are taking now."
4. Discussion of the pros (benefits) and cons (risks) of the options
Rationale: MDs frequently discuss the pros of one option and the cons of another option without fully exploring the pros and cons of each.
Example: "The new medication is more expensive, but you need to take it only once a day."
5. Discussion of the uncertainties associated with the decision
Rationale: While often difficult, a discussion of the uncertainties is crucial for a patient's comprehensive understanding of the options.
Example: "Most patients with your condition will respond to this medication, but not all."
6. Assessment of the patient's understanding
Rationale: Once the core disclosures are made, the physician must check with the patient to know if what the doctor said so far makes sense: This is a central goal of informed decision making.
Example: "Does that make sense to you?" "Do you have any questions so far?"
7. Exploration of patient preference
Rationale: Physicians may assume that patients will speak up if they disagree with a decision, but patients often need to be asked for their opinion.
Example: "Does that sound reasonable?"

Facilitator Key to the Clinical Scenarios

As noted previously, videos were scripted to illustrate varying levels of trainee proficiency; each contains some deficiencies. The scenarios were developed iteratively for use in a research study on direct observation. Provided on the following pages are the deficiencies for each encounter. You may not agree with all of them; this could be a rich source for dialogue with your workshop group. Some deficiencies are more blatant than

others; the key simply denotes presence or absence of the deficiency, not degree. All of these videos can be used for direct observation of competence training described in Chapter 9.

The facilitator key is provided in the format of an abstraction instrument. The key can be used to assess faculty performance in direct observation, and to provide feedback to the faculty. The criteria are based on an absolute (criterion-based) scale, not a relative (normative-based) scale for optimal patient care. Your faculty may have different points of view depending on the level of trainee you choose for your direct observation exercises.

Scenario 1: Level 1 History Taking

Clearly Unsatisfactory. On the 9-point mini-CEX instrument, this trainee should receive a rating of 3 or lower.

Description of Deficiency

Primary

1. Failed to introduce himself
2. No open-ended questions; all questions closed-ended
3. Question about quality of pain leading ("pressure or squeezing")
4. Failed to ask if patient is having chest pain now
5. Did not ask what made pain better (alleviating)
6. Did not ask duration of discomfort
7. Failed to ask about any prior episodes of chest discomfort
8. Failed to ask about diaphoresis
9. Did not ask how severe the pain was
10. Did not ask about past medical history (except cholesterol)

Secondary

11. Lacked patient centeredness
12. Failed to ask if the patient had ever smoked
13. Failed to ask about occupation
14. Did not ask age of father at time of his heart attack
15. Family history closed-ended
16. Failed to ask patient if patient had any questions
17. Failed to recognize patient concern at end of interview/offer empathetic or reassuring comment

Scenario 2: Level 2 History Taking

Marginal to low satisfactory. On a 9-point scale, this performance received an average rating of 4 among a group of communication experts. Research has shown that a 4 on a 9-point scale equates to a marginal, or barely satisfactory, performance.

Description of Deficiency

Primary

1. Failed to ask if patient is having chest pain now
2. Did not ask about duration of chest discomfort
3. Did not ask what made pain better

4. Did not ask about any prior episodes of chest discomfort

Secondary

5. Did not ask age of father at time of heart attack
6. Failed to ask about occupation
7. Lacked patient centeredness
8. Did not ask explicitly about heart disease in other family members
9. Failed to ask patient if patient had any questions

Scenario 3: Level 3 History Taking

High satisfactory (from a data-gathering perspective). Some experts viewed this scenario as poor in patient-centeredness.

Description of Deficiency

Secondary

1. Could be more patient centered (subjective)
2. Failed to ask patient if patient had any questions
3. Failed to ask about leg edema

Scenario 4: Level 1 Physical Examination

Clearly unsatisfactory. On the 9-point mini-CEX instrument, this trainee should receive a rating of 3 or lower.

Description of Deficiency

Primary

1. Failed to take blood pressure
2. No respiratory rate taken
3. No pulse taken
4. cursory examination of oral cavity, no light or tongue blade
5. No examination of ears in patient with fever
6. No examination of nasal cavity
7. Did not examine sinuses
8. Lymph node examination incomplete and cursory (e.g., does not examine posterior nodes or supraclavicular)
9. Lung examination incomplete—no anterior or right middle lobe (RML) auscultation
10. Lung examination technique incorrect—does not compare from side to side (subtle to detect)
11. Lung examination—no percussion
12. Cardiac examination—no assessment of point of maximal impulse (PMI)
13. Cardiac examination—no use of bell to assess S3 or S4
14. Cardiac examination—does not listen to heart in recumbent position
15. Cardiac examination—no assessment of jugular venous distension (JVD)

Secondary

16. Abdominal examination not performed

17. Carotid examination not performed
18. Thyroid examination not performed

Scenario 5: Level 2 Physical Examination

Marginal performance. This performance was rated a 4 on a 9-point scale by two experts in physical diagnosis.

Description of Deficiency

Primary

1. Pulse not taken
2. Respiratory rate not taken
3. Did not examine ears
4. Did not examine nasal passages
5. Lymph node examination too rapid and cursory (did not examine posterior, supraclavicular nodes completely)
6. Lung examination incomplete—no anterior or RML auscultation
7. Cardiac examination—no assessment of JVD

Secondary

8. Took blood pressure in incorrect position
9. Did not ask if sinuses tender when palpated
10. Thyroid examination not performed
11. Carotid examination not performed
12. Abdominal examination not performed

Scenario 6: Level 3 Physical Examination

High satisfactory to low superior. This trainee should receive a score of 6 or 7 on a 9-point mini-CEX scale.

Description of Deficiency

Primary

1. Respiratory rate not taken
2. Did not examine nasal passages
3. Pulmonary—did not listen anteriorly with stethoscope
4. Cardiac—did not assess PMI

Secondary

5. Did not listen to complete respiratory cycle before moving stethoscope
6. Thyroid examination not performed
7. Posterior lymph node examination not completed
8. Abdominal examination not performed

Scenario 7: Level 1 Counseling

Clearly unsatisfactory. On the 9-point mini-CEX instrument, this trainee should receive a rating of 3 or lower.

Description of Deficiency

Primary

1. No discussion of patient role in decision making
2. No discussion of the risks (side effects)

3. No discussion of alternatives
4. No mention of dose
5. No assessment of patient's understanding (Do you have any questions?)
6. No discussion of the uncertainties of starting the medicine
7. No exploration of patient preference
8. Failed to address patient reluctance to take medicine

Secondary

9. No discussion of the degree/magnitude of benefit for stroke or acute myocardial infarction (AMI) prevention
10. Did not respond to nonverbal cues
11. Follow-up interval too long
12. Never told patient her cholesterol level
13. Never told patient her goal cholesterol level

Scenario 8: Level 2 Counseling

Marginal performance. Although the resident is pleasant, a number of key items necessary for informed decision making are lacking. Two experts in informed decision making rated this performance a 4 on a 9-point scale.

Description of Deficiency

Primary

1. No discussion of patient role in decision making
2. No discussion of alternatives
3. No discussion of other side effects (e.g., myopathy)
4. No mention of dose
5. No exploration of patient preference

6. No discussion of the uncertainties of starting the medicine
7. No response to patient reluctance to take medicine

Secondary

8. No discussion of the degree/magnitude of benefit for stroke or AMI prevention
9. Follow-up interval too long
10. Failed to counsel on need for blood test at 6 weeks
11. Never told patient her goal low-density lipoprotein level
12. Missed opportunity to counsel on tobacco use

Scenario 9: Level 3 Counseling

High satisfactory to low superior. This trainee should receive a score of 6 or 7 on a 9-point mini-CEX scale. The trainee may have provided *too* much information for patient, and was not sufficiently patient-centered.

Description of Deficiency

Primary

1. No discussion of the uncertainties of starting the medicine

Secondary

2. No mention of dose
3. Did not completely respond to patient fear
4. Frames other medications negatively

Contents

1	Evaluation Challenges in the Era of Outcomes-Based Education	1
	John J. Norcini, Eric S. Holmboe, and Richard E. Hawkins	
2	Issues of Validity and Reliability for Assessments in Medical Education.....	10
	Brian E. Clauser, Melissa J. Margolis, and David B. Swanson	
3	Evaluation Forms and Global Rating Scales	24
	Louis Pangaro and Eric S. Holmboe	
4	Using Written Examinations to Assess Medical Knowledge and Its Application.....	42
	Richard E. Hawkins and David B. Swanson	
5	Practice Audit, Medical Record Review, and Chart-Stimulated Recall.....	60
	Eric S. Holmboe	
6	Multisource Feedback (360-Degree Evaluation).....	75
	Jocelyn M. Lockyer and Stephen G. Clyman	
7	Portfolios	86
	Eric S. Holmboe, Margery H. Davis, and Carol Carraccio	
8	Direct Observation: Standardized Patients.....	102
	Richard E. Hawkins and John R. Boulet	
9	Direct Observation by Faculty.....	119
	Eric S. Holmboe	
10	Evaluating Evidence-Based Practice	130
	Michael L. Green	
11	Competence in Improving Systems of Care Through Practice-Based Learning and Improvement	149
	F. Daniel Duffy and Eric S. Holmboe	
12	Simulation-Based Assessment	179
	Ross J. Scalese and S. Barry Issenberg	
13	Working with Problem Residents: A Systematic Approach.....	201
	Catherine R. Lucey and Robert McL. Boote	
14	Constructing an Evaluation System for an Educational Program	216
	Richard E. Hawkins and Eric S. Holmboe	
	Index	239

Evaluation Challenges in the Era of Outcomes-Based Education

John J. Norcini, PhD, Eric S. Holmboe, MD,
and Richard E. Hawkins, MD

DRIVERS OF CHANGE IN ASSESSMENT

Outcomes-Based Education
Accountability and Quality Assurance
Technology
Psychometrics

FRAMEWORK FOR ASSESSMENT

Dimension 1: Competencies
Medical Knowledge
Interpersonal and Communication Skills
Patient Care
Professionalism
Practice-Based Learning and Improvement
Systems-Based Practice
Dimension 2: Levels of Assessment
Miller's Pyramid
The Cambridge Model
Dimension 3: Assessment of Progression

CRITERIA FOR CHOOSING A METHOD

ELEMENTS OF EFFECTIVE FACULTY DEVELOPMENT

FUTURE CHALLENGES

Methods of Assessment
Traditional Measures
Methods Based on Observation
Simulation
Work
New Competencies: Teamwork and
Systems-Based Practice
Systems of Assessment

CONCLUSION

REFERENCES

Through the early 1950s, physicians were assessed in limited ways.¹ Medical knowledge was evaluated with essays and other open-ended question formats that were graded by an instructor. Clinical skill and judgment were tested using an oral examination that often required the student to go to the bedside, gather patient information, and present it along with a diagnostic list and treatment plan to one or more examiners who asked questions. Because these were the only generally accepted methods available, they were applied to most assessment problems even if they were not completely suitable to the task.

From that point to the present, there have been extensive changes in the way assessment is conducted. Methods have proliferated, as has the sophistication of their use. Much progress has been made in the assessment of medical knowledge with a variety of written and computer-based techniques offering reliable and valid results. In the last few decades, considerable

gains have been made in defining and enhancing the psychometric qualities of objective structured clinical examinations (OSCEs), particularly related to their use in high-stakes examinations. However, assessment in the context of clinical education has lagged to some degree, especially in the areas of clinical skills and performance. Equally important, the methods that have been developed to support clinical education often rely on faculty who are inexperienced in their use, do not share common standards, and have not been trained to apply them in a consistent fashion. Faculty development has failed to keep pace with the application of these new educational methods.

This chapter will present an overview of the drivers of change in the assessments used during clinical education, a framework for such assessment, criteria for choosing methods, elements of an effective faculty development effort, and the nature of future challenges.

Drivers of Change in Assessment

The increase in the number of methods and the enhanced sophistication of assessment overall has been motivated by public pressure for accountability and quality improvement. This has been accompanied by curricular changes in the form of outcomes-based education and supported by improvements in technology and psychometrics.

Outcomes-Based Education

Consistent with trends in all of education, the past two decades have seen an evolution in the thinking about how physicians should be trained.^{2,3} A focus on the educational process has given way to an emphasis on what a physician should look like at the end of training and at important junctures during the training process. Outcomes-based education starts with a specification of the competencies expected of a physician, and these requirements drive the content and structure of the curriculum, the selection and deployment of teaching and learning methods, the site of training, and the nature of the teachers. Assessment plays a central role in determining whether students and residents have actually achieved the competencies that have been specified and whether the educational program has been efficacious.

This change in thinking and the need to assess the diverse competencies of the physician has been an important factor in the development of new methods of assessment. Pressure for additional developments will persist as more schools and programs implement changes to attain the goals of an outcomes-based education.

Accountability and Quality Assurance

The movement to outcomes-based education has been accompanied by significant efforts to enhance the accountability of physicians.⁴ Motivated in part by high-profile cases such as those involving Michael Swango in the United States and Howard Shipman in the United Kingdom, the public has pressured medicine to increase its level of oversight and eliminate the “bad apples.”^{5,6} Medical educators are also more keenly aware that too many trainees graduate with substantial deficiencies in knowledge and clinical skills.^{7,8} Promoting trainees who lack competence erodes the trust between the medical profession and the public.

At the same time, there has been a variety of efforts to improve the quality of health care.^{9–11} These efforts have relied on methods devised by workers in the field of quality management science and, in some cases, used successfully in industry for over 50 years. Central to both accountability and quality assurance is assessment. It offers a means of identifying those whose overall performance is well below standards and identifying areas of weakness, helping to drive the continuous quality improvement process. These developments

have helped to fuel the creation of several new methods of assessment and to increase the use of those already available.

Technology

Over the past 50 years, the availability of more sophisticated technology has changed the testing of medical knowledge and judgment in fundamental ways.^{12,13} The introduction of the computer heralded an era of large-scale testing by encouraging the use of multiple-choice questions (MCQs) that could be scanned by machine, turned into scores, and then reported in an efficient and objective fashion.

More recently, the intelligence of the computer has improved assessment in two ways:

1. On the one hand, it has enabled the application of significant psychometric advances to the assessment of medical knowledge. Specifically, the computer’s intelligence has improved efficiency by allowing the selection of questions that are targeted to the ability of particular examinees. Sequential testing and adaptive testing permit gains in efficiency and precision.
2. On the other hand, it has improved the assessment of clinical decision making by permitting the use of interactive item formats that more closely simulate the types of judgments physicians need to make in practice.

The impact of technology on assessment of clinical skills has been slower to develop but there are now a number of tools that recreate aspects of the clinical encounter with considerable fidelity. These methods have a growing impact on assessment, especially in the area of procedural skills.

Psychometrics

At the same time that the technology has improved, there have been significant advances in psychometrics, the basic science of assessment. Classical test theory, prominent from the turn of the 20th century, has gradually given way to measurement models based on strong assumptions about test items and examinees. The family of item response theory models now makes it possible to produce equivalent scores even when examinees take tests made up of different questions.¹⁴ They also support the computer-based administration of examinations that are tailored to the ability level of individual test-takers; this allows tests to be shortened by as much as 40%.¹⁵ The ability to shorten tests has cost and validity implications; less test material exposure decreases the likelihood that future examinees are familiar with examination content.

Generalizability theory, another major development, makes it possible to identify how much error is associated with different facets of measurement (e.g., raters, patients).¹⁶ Based on this information, assessments can be prospectively designed to make the best use of resources, such as faculty time, while maintaining the reliability of the results.

In addition to these major developments, there have been a number of other advances. For example, there is a variety of systematic methods available for setting standards on tests and for identifying when test questions are biased against particular groups of examinees.^{17,18} Test development methods have gotten better, as have the means for judging whether particular items are working properly. Overall, these advances have improved both the quality and efficiency of assessment.

Framework for Assessment

As methods of assessment have proliferated, so has the need to use them efficiently and to combine them into a system of evaluation. Developing, implementing, and sustaining effective systems for the evaluation of clinical competence in medical school, residency, and fellowship programs require consideration of what competencies need to be assessed, how to best assess them, and the level of the trainee being assessed. Consequently, a three-dimensional framework for structuring an assessment system is needed. Along the first dimension are the competencies that need to be assessed, along the second is the level of assessment required, and along the third is the trainees' stage of development.

Dimension 1: Competencies

As shown in Table 1-1, there are several schemes for describing the knowledge, skills, and attributes of the physician.^{4,19–21} The Canadian Medical Education Directions for Specialists (CanMEDS) model, which was developed by the Royal College of Physicians and Surgeons in Canada, describes the competencies in terms of the roles of a physician. Good Medical Practice, which was created by the General Medical Council in the United Kingdom, describes the elements of good practice.

In the United States, two influential groups developed a set of core competencies. The Accreditation Council for Graduate Medical Education (ACGME) and the American Board of Medical Specialties (ABMS) adopted six general competencies. These competencies are the outcomes framework for residency and fellowship training as well as maintenance of certification programs throughout a physician's career in the United States. The Institute of Medicine (IOM) has recommended five core skills, or competencies, that create a framework for evaluating performance and stimulating the reform of education. They are intended to improve professional education and practice with a goal of enhancing the safety and quality of health care. Although there are some differences among the schemes, there is also significant overlap in these descriptions of a physician.

For purposes of this chapter, we will focus on the six ACGME/ABMS competencies: medical knowledge, interpersonal and communication skills, patient care, professionalism, practice-based learning and improvement, and system-based practice. Definitions for each of these follow with more emphasis on the last two, which are relatively new.

These competencies are intended as the first step in identifying the learning objectives of graduate training programs, and it is anticipated that they will be adapted to the content, education, and practice of the particular specialty/subspecialty. The data produced by the assessment of these competencies serve as a basis for judging the quality of the trainees and their training, as well as supporting the continuous improvement of both.

Medical Knowledge

Students, residents, and practicing physicians must possess knowledge of the basic and clinical sciences and be able to apply them to patient care. Moreover, they are expected to demonstrate an appropriate approach to reasoning about clinical problems.

Table 1-1 The Competencies of Physicians as Described by Four Organizations

CANMEDs	GMC	ACGME/ABMS	IOM
Medical expert	Good clinical care	Medical knowledge	Employ evidence-based practice
Communicator	Maintaining good medical practice	Interpersonal and communication skills	Work in interdisciplinary teams
Collaborator	Teaching and training appraising and assessing	Patient care	Provide patient-centered care
Manager	Relationships with patients	Professionalism	—
Health advocate	Working with colleagues	Practice-based learning and improvement	Apply quality improvement
Scholar	Probity	Systems-based practice	Utilize informatics
Professional	Health	—	—

ABMS, American Board of Medical Specialists; ACGME, Accreditation Council for Graduate Medical Education; CanMEDS, Canadian Medical Education Directions for Specialists; GMC, General Medical Council (UK); IOM, Institute of Medicine.

Interpersonal and Communication Skills

Students, residents, and practicing physicians must possess the interpersonal and communication skills that produce effective information exchange and relationships among members of the health care team. With patients, this competence supports the development and maintenance of a therapeutic and ethical relationship.

Patient Care

Good patient care requires that physicians are both compassionate and effective. They must be able to communicate well and demonstrate caring while gathering the data they need. This competence necessitates informed decisions based on medical knowledge and patient preferences, management plans that are carried out fully, and counseling of both patients and their families. Essential procedures must be carried out safely and effectively.

Professionalism

Students, residents, and practicing physicians must be committed to carrying out their professional responsibilities, adhering to ethical principles, and being sensitive to patients. Professionalism is ingrained in overall clinical competence and includes the aspirations to excellence, humanism, accountability, and altruism.

Practice-Based Learning and Improvement

Trainees and practicing physicians are expected to apply scientific evidence and use methods to investigate, evaluate, and continuously improve the quality of care for patients. This requires trainees to identify areas of opportunity for improvement, identify and correct medical errors, and use information technology at the point of care for the benefit of patients. The appropriate knowledge, skills, and attitudes in quality improvement and evidence-based medicine are needed for this competency.²²

Systems-Based Practice

Trainees and practicing physicians require a deep understanding of both the micro- and macro-systems in which health care is provided. Trainees must learn to apply this knowledge to utilize efficiently and effectively the resources, health care providers, and technology to optimize the care delivered not only to individual patients, but to populations of patients as well. Knowledge, skills, and attitudes in effective teamwork are crucial to this general competency.^{22,23}

Dimension 2: Levels of Assessment

The multifaceted nature of the competencies makes it apparent that no single method could provide a

sufficient basis for making judgments about students or residents. In an organized approach to this problem, Miller proposed a classification scheme that stratifies assessment methods based on what they require of the trainee. Often referred to as Miller's pyramid, it is composed of four levels: knows, knows how, shows how, and does.²⁴

Miller's Pyramid

Knows. This is the lowest level of the pyramid and it contains methods that assess what a trainee "knows" in an area of competence. Forming the base of the pyramid, knowledge represents the foundation upon which clinical competence is built. An MCQ-based examination composed of questions focused on ethics and principles of patient confidentiality would provide an assessment of what a trainee "knows" about professionalism.

Knows how. To function as a physician, a good knowledge base is necessary but insufficient. It is important to know how to use this knowledge in the acquisition of data, the analysis and interpretation of findings, and the development of management plans. For example, a method that poses a moral dilemma, asks trainees to reason through it, and evaluates the sophistication of their moral thinking would provide a "knows how" assessment of professionalism.

Shows how. Although trainees may know and know how, they may not be able to integrate these skills into a successful performance with patients. Consequently, certain assessment methods require the trainee to show how they perform with patients. For example, a standardized patient presenting with an ethical challenge would offer the trainee an opportunity to "show how" he or she would respond to a professionalism challenge.

Does. No matter how good traditional assessment methods become, there remains the concern that what happens in a controlled testing environment does not generalize directly to what happens in practice. The highest level of Miller's pyramid, therefore, focuses on methods that provide an assessment of routine performance. For example, the development and use of a critical incident system, such as the one currently used in some medical schools, offers an assessment of what students actually do in terms of professionalism.

Miller's pyramid is a useful framework for considering differences and similarities among assessment methods. However, the fact that it is a pyramid might imply to some that methods addressing the higher levels are better. Instead, superior methods are those best aligned with the purpose of the test. For example, if an assessment of medical knowledge is needed, a method associated with that level (e.g., multiple-choice questions) is better than a method associated with another level (e.g., standardized patients).

The Cambridge Model

As physicians near the end of training and enter practice, external forces come to play a very large role in performance. The Cambridge Model, a variation on Miller's pyramid, proposes that performance in practice (the highest level of the pyramid) is influenced by two large forces beyond competence.²⁵ Systems-related factors, such as government programs, patient expectations, and guidelines, strongly influence what physicians do. Similarly, factors related to the individual physician such as state of mind, physical and mental health, and relationships with peers and family have a significant effect. Consequently, assessment becomes more difficult because it is harder to disentangle the effects of the context of care from the competence of the individual physician. Here, a focus on health care processes and outcomes as a measure of what a physician "does" may provide a more

valid assessment of a physician's ability to integrate multiple competencies within a complex social context.

Dimension 3: Assessment of Progression

Acquiring competence is not an overnight process. Trainees progress through a series of stages that begin in undergraduate medical education and continue throughout their careers. Educators must be able to recognize when a trainee has attained sufficient knowledge, skills, and attitudes to enter the next stage and this requires appropriate standards and benchmarks for the transition. Hubert and Stuart Dreyfus have created a developmental model of learning applicable to the health professions that proposes five stages of educational development (Table 1-2).²⁶

The Dreyfus model proposes that each stage of learning requires a different method of teaching. The novice needs instructor-driven teaching, but the expert needs

Table 1-2 The Stages of Learning as Proposed by Dreyfus

Stage of Learning	Method of Learning (Teaching Style)	Learning Steps	Learner Characteristics
1. Novice	Instruction (instructor) Breaks skill into context-free, discrete tasks, concepts, rules	Recognizes the context-free features Knows rules for determining actions based on these features	Learning occurs in a detached analytic frame of mind
2. Advanced beginner	Practice (coach) Experiences coping with real situations Points out new aspects of material Teaches rules and reasoning techniques for action	Recognizes relevant aspects based on experience that makes sense of the material Learns maxims about actions based on new material	Learning occurs in a detached, analytic frame of mind
3. Competence	Apprenticeship (facilitator) Develops a plan or chooses perspective that separates "important" from "ignored" elements Demonstrates that rules and reasoning techniques for choosing are difficult to come by Role models are also emotionally involved in making decisions	Volume of aspects is overwhelming Performance is exhausting Sense of what's important is lacking Stands alone making correct and incorrect choices Coping becomes frightening, discouraging, elating	Learner is emotionally involved in the task and its outcome Too many subtle differences for rules; student must decide in each case Makes a mistake, then feels remorse Succeeds, then feels elated Emotional learning builds competence
4. Proficiency	Apprenticeship (supervisor) Gains more specific experience with outcomes of one's decisions Applies rules and maxims to decide what to do	Rules and principles are replaced by situational discrimination Emotional responses to success or failure build intuitive responses that replace reasoned ones	Learner immediately sees the goal and salient features Learner reasons how to get to the goal by applying rules and principles
5. Expertise	Independence (mentor) Experiences multiple, small random variations Observes other experts or experiences nonrandom simulations Working through the cases must emotionally matter	Gains experience with increasingly subtle variations in situations Automatically distinguishes situations requiring one response from those requiring another	Immediately sees the goal and what must be done to achieve it Builds on previous learning experiences

From Dreyfus HL: On the Internet. Thinking in Action Series. New York, Routledge, 2001.

independence. Likewise, the characteristics of learners and the steps they must go through to acquire competence will change over the five stages of development. Necessarily, the methods of assessment applied at each developmental level must also evolve. For example, at the level of the novice, an MCQ-based knowledge test might be most appropriate, but a standardized patient-based examination might be better suited to trainees who are in the competence or proficiency stages. Educators need to recognize this developmental sequence when designing an assessment system and it will be critical to ensure that the chosen method is suitable to the task.

Criteria for Choosing a Method

Decisions about which method of assessment to use in a particular circumstance have traditionally rested on validity and reliability. Validity is the degree to which the inferences based on the scores of an assessment are correct. Reliability, a closely related concept, is a measure of the repeatability or consistency of scores, akin to the 95% confidence intervals often provided with medical tests. Valid inferences regarding a particular test score or assessment result are to a large extent dependent upon the reliability of these outcomes. These are certainly critical characteristics of educational tests and they have the further advantage of being quantifiable.

For purposes of assessment in medical education, van der Vleuten and Schuwirth have recently added educational effect, feasibility, and acceptability as factors to be considered in choosing a method of assessment. In terms of educational effect, they argue that trainees will work hard in preparation for an assessment.²⁷ Consequently, the method should direct them to study in the most relevant way. For example, if an educational objective is for trainees to know the differential diagnoses for a particular chief complaint, then assessment using extended matching questions will induce better learning than assessment based on standardized patients.

Feasibility is the extent to which an assessment method is affordable and efficient. Although high-fidelity simulations might be a good way to assess procedural competence, the use of a method such as direct observation of procedural skills (DOPs), which is based on faculty observation, is likely to be more feasible in most graduate training settings.²⁸

Acceptability is the degree to which the trainees and faculty believe that the method produces valid results. This factor will influence motivation of faculty to use the method and reduce the trainees' distrust of the results. It is important that educational leaders not underestimate trainee knowledge and understanding of assessment and their ability to participate in decisions regarding assessment practices.

In addition to these five factors, it is important to consider how a particular method fits into the overall system for assessment. The same method can be used to assess more than one competence. For example,

peer assessment can provide a measure of both professionalism and interpersonal skills. Likewise, two different methods can be used to capture information on the same competence, thereby increasing confidence in the results. For example, patient care can be assessed using both the mini-CEX (clinical evaluation exercise) and monthly ratings by attending physicians.

Educational effect, feasibility, and acceptability are not easily quantifiable, nor is the relationship among methods of assessment in a system. However, these factors plus reliability and validity should be weighed when considering selection of a particular method.

Elements of Effective Faculty Development

Although faculty members are important to evaluation regardless of method, they play a particularly critical role in assessment in the clinical setting because it is often based on observation. Recall that Miller placed "performance," meaning the care of actual patients, at the tip of the pyramid. Envision the pyramid as a spear and at the tip of that spear are patients. Using this metaphor helps faculty appreciate the central role of observation in both assuring trainee competence and guaranteeing that patients receive high-quality, safe care in the context of training.

In many respects, assessment methods based on observation are only as good as the individuals using them. Although there has been substantial progress in creating these new methods, significantly less attention has been paid to the development of approaches to training faculty in how to use them most effectively. This omission continues to occur despite repeated studies demonstrating significant problems with the quality of faculty assessments.²⁹⁻³¹

There are three significant reasons why faculty training is urgently needed. First, in order to perform quality assessment, faculty members must possess sufficient knowledge, skill, and attitudes in that competency. For example, the decline of clinical skills teaching in the 1980s and 1990s resulted in many of today's educators failing to acquire a high level of clinical skills themselves. This limits the degree to which they can validly assess clinical performance.

Second, the competencies will evolve and change over time. Witness the birth of the "new" competencies of practice-based learning and improvement and systems-based practice. The majority of faculty today never received any formal instruction in these competencies during their own training and thus they are often learning new knowledge and skills alongside their trainees.

Finally, assessment is a core tenet of professionalism for medical educators. Too often, faculty members view it as someone else's job, especially when a negative performance appraisal is involved. Faculty development reinforces the importance of assessment and provides medical educators the opportunity to develop common standards for performance.

To make effective use of the methods of assessment based on observation, educational institutions must commit the necessary resources for faculty development. However, too often faculty development translates into a project or a brief workshop. If faculty development is to be truly successful, medical educators need to embrace new strategies that embed faculty development in real-time teaching and clinical activities. Faculty development, like quality improvement and maintenance of competence, must become a continuous process and appropriately rewarded. As noted earlier, the quality and safety of patient care depend on it.

Medical educators must also end their quest for the holy grail of evaluation, the perfect rating form imbued with special powers to solve all evaluation needs. Evaluation is hard work and it requires a multifaceted approach. Landry and Farr, in a landmark article in the performance appraisal field nearly 25 years ago, pleaded with researchers to redirect development efforts from a search for the perfect rating form to training the assessors.³² Researchers in this field subsequently developed a number of validated approaches that can lead to better evaluations. Table 1-3 provides a summary of several approaches with applicability for medical educators. Most, if not all, of these approaches can be used in small, repeated aliquots of time longitudinally, and there is some evidence that they work in the medical education setting.^{33,34} Chapter 9 provides detail on how these training approaches

were modified to create a faculty training program to improve observation skills.

Future Challenges

Although considerable strides have been made in assessment, much work remains to be done. Specifically, effort is needed to continue to (1) refine the different methods of assessment, (2) expand their application to new competencies such as teamwork and systems-based practice, and (3) develop systems that integrate them in support of ongoing quality improvement.

Methods of Assessment

Traditional Measures

Traditional measures will continue to play an important role in the assessment of clinical proficiency. Specifically, written methods such as MCQs and standardized patients will be mainstays of all assessment programs for the near future. All of these methods can be improved and work on each must continue.

Methods Based on Observation

Even though assessment has been woven through the basic science curriculum, historically it has not been as

Table 1-3 Methods to Train Faculty in the General Competencies

Training Method	Description	Example
Performance dimension training (PDT)	Familiarize faculty with appropriate performance dimensions or standards for use in evaluation by reviewing the dimensions of a performance or competency. Faculty members work in small groups to improve their understanding of these definitions with review of actual trainee performance or clinical evaluation vignettes. PDT should focus on optimal performance.	Faculty members discuss the elements of what constitutes a safe and efficient discharge of a patient who needs home assistance and follow-up (systems-based practice).
Frame-of-reference training	Using the results of the PDT exercise, faculty members define what would constitute "satisfactory" performance (the anchor point). Faculty members then practice evaluating trainees performing at various levels of competence using the evaluation instrument of choice. The group discusses reasons for the differences between faculty ratings.	Faculty members are given several vignettes along with examples of the medical record, etc., regarding a discharge performed by a resident. For each vignette, the faculty members rate the level of performance (unsatisfactory, marginal, satisfactory, superior). The vignettes provide examples of different levels of competence in systems-based practice. After each rating, group members discuss their ratings with each other. This exercise helps to "calibrate" faculty to be able to discriminate between different levels of competence.
Rater error training	Faculty members discuss the common errors (such as halo effect or compensation fallacy) in ratings. Each error is described and defined.	Examples of each error are provided for discussion and review. Actual examples from the program could be used.

well integrated with clinical education. Nonetheless, assessment methods based on the observation of routine encounters in the clinical setting offer a rich and feasible target for assessment. Continued refinement of the methods themselves is needed, as is faculty development, which is a key to their successful use. Furthermore, the opportunity for educational feedback as part of these methods is probably as important as their assessment potential.

Simulation

Improvements in technology have spurred the development of a series of simulators that recreate reality with high fidelity. The use of simulation in assessment is in its infancy, the technology remains expensive, and several developments are needed before widespread adoption and use. Researchers will need to continue to focus on identifying appropriate scoring methods, optimizing the generalizability of scores, and ensuring their relevance to performance in practice.³⁵ Particularly in the area of procedural skills, however, these methods will offer the ability to test under a variety of conditions without concern for harm to patients. Educators will confront difficult decisions requiring them to balance the cost, variable fidelity of individual simulation methods, and potential risks to patients (and trainees) in making decisions regarding how best to assess procedural skills.³⁶

Work

The assessment of physicians' performance at work (the "does" level of Miller's pyramid) is a relatively recent development. Despite the need for significant research, the day-to-day performance of physicians is being used increasingly in the settings of continuous quality improvement and physician accountability. Assessment in this context is a matter of identifying the basis for the judgments (e.g., outcomes, process of care), deciding how the data will be gathered, and avoiding threats to validity and reliability (e.g., patient mix, patient complexity, attribution, and numbers of patients).³⁷ Given the pressure to increase quality and decrease costs, it is important that improvements in this form of assessment happen quickly.

New Competencies: Teamwork and Systems-Based Practice

The concepts of systems-based practice and interdisciplinary team education are only now taking shape in clinical practice and medical education. Educators are struggling to determine how to incorporate these new competencies into their curricula, so it is not surprising that the current science around evaluating competencies is in its infancy. However, several groups are defining the specific knowledge, skills, and attitudes required for competent interdisciplinary teamwork and interaction with health care systems.^{23,38} Chapter 11

provides a framework for integrating systems-based thinking and practice into the educational environment and a starter set of evaluation measures and methods.

Systems of Assessment

The movement toward outcomes-based education and assessment presents many challenges for medical educators. Educational leaders will need to integrate traditional and new assessment methods into their educational programs to ensure that individual trainees meet important educational and professional objectives and to inform continued quality improvement of their programs. Assessment approaches must be clearly aligned with educational objectives and congruent with teaching and learning methods. Assessment should be closely intertwined with instructional activities in order to optimize efficient use of resources and to consolidate learning. The assessment system will need to include multiple methods to capture each of the general competencies and ideally to provide for the assessment of different aspects of each competency by different methods. Program and clerkship directors will need to prepare the assessors, through the implementation of robust faculty development programs, and inform and engage trainees in order for the assessment system to succeed.

Beyond the performance of individual trainees, the assessment system will need to support the continuous collection and analysis of aggregate data to provide feedback regarding the quality of the educational program. This includes information from more traditional assessment methods, such as program-level subscores on MCQ examinations or aggregate case-level data from clinical skills examinations, as well as composite scores or ratings from newer methods such as multi-source feedback and computer simulation-based exercises. It also involves collection and analysis of clinical information, such as compliance with evidence-based health care processes or patient health outcomes that can provide the impetus for curricular change or feedback on the quality of educational interventions. Establishing such a connection, at least at the institutional level, will facilitate conduct of needed research to elucidate the relationships between educational activities and health care practices and outcomes.

In addition to compiling aggregate data within programs to inform quality improvement initiatives, assessment systems will need to enable information gathering regarding the performance of program graduates. As with concurrent measures, educational leaders will need to access and incorporate into their assessment systems information about future competence and performance of program graduates in order to guide quality improvement efforts. Some information, such as licensure actions, in-training or board certification examination scores, or program director ratings may not be difficult to obtain. Obtaining other sources of information, such as specific performance measures or clinical data, to provide additional feedback regarding educational program quality will require

more effort. The formation of collaborative projects and networks linking professional and clinical outcomes across the spectrum of education and practice will facilitate understanding and incorporation of information critical to the continuous quality improvement of educational programs.

Conclusion

Public and professional pressure to increase accountability and quality improvement in clinical care has resulted in important changes in medical education and assessment. Delineation of essential physician competencies and a move toward outcomes-based medical education has led to a critical review of the quality and methods used in the assessment of competence and performance. Advances in technology and psychometrics have supported continued refinement of traditional assessment modalities and the development of new approaches. Educational leaders now face difficult challenges in developing and integrating assessment systems into their educational programs. They must understand the psychometric properties of various assessment tools, consider their relevance to trainee level, as well as to instructional methods and educational objectives, and then balance these factors against program culture and resource availability in deciding what methods to use in their assessment system. The chapters that follow are intended to help guide educational leaders in designing their assessment systems to support evaluation of individual trainees and continuous quality improvement of their educational programs.

REFERENCES

1. Norman GR: Research in medical education: Three decades of progress. *BMJ* 2002;324:1560–1562.
2. Harden RM, Crosby JR, Davis M: An introduction to outcome-based education. *Med Teacher* 1999;21(1):7–14.
3. Association of Medical Education in Europe: Education Guide No 14: Outcome-based Education. Dundee, AMEE, 1999.
4. Institute of Medicine: Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC, National Academy Press, 2001.
5. Stewart JB: *Blind Eye: How the Medical Establishment Let a Doctor Get Away with Murder*. New York, Simon & Shuster, 1999.
6. The Final Report of the Shipman Inquiry. Accessed at <http://www.the-shipman-inquiry.org.uk/backgroundinfo.asp>, July 26, 2007.
7. Mangione S, Nieman LZ: Cardiac auscultatory skills of internal medicine and family practice trainees. A comparison of diagnostic proficiency. *JAMA* 1997;278(9):717–722.
8. Reilly BM: Physical examination in the care of medical inpatients: An observational study. *Lancet* 2003;362(9390):1100–1105.
9. Berwick DM, Godfrey AB, Rossener J: *Curing Health Care: New Strategies for Quality Improvement*. San Francisco, Jossey-Bass, 1990.
10. Laffel G, Blumenthal D: The case for using industrial quality management science in health care organizations. *JAMA* 1989;262:2869–2873.
11. Plsek PE: Quality improvement methods in clinical medicine. *Pediatrics* 1999;103(Suppl):203–214.
12. Bunderson CV, Inouye DK, Olsen JB: The four generations of computerized educational measurement. In Linn RL (ed): *Educational Measurement*. Washington, DC, American Council on Education, 1989.
13. Norcini JJ: Computers in physician licensure and certification: New methods of assessment. *J Educ Computing Res* 1994;10:161–171.
14. Hambleton RK, Swaminathan H: *Item response theory: Principles and applications*. Dordrecht, Kluwer, 1985.
15. Green BF: Adaptive testing by computer. In Ekstrom RB (ed): *Principles of Modern Psychological Measurement*. San Francisco, Jossey-Bass, 1983, pp 5–12.
16. Brennan RL: *Generalizability Theory*. New York, Springer-Verlag, 2001.
17. Norcini JJ: Standard setting. In Dent JA, Harden RM (eds): *A Practical Guide for Medical Teachers*. Churchill Livingstone, 2005, pp 293–301.
18. Ekstrom RA (ed): *Handbook of Methods for Detecting Test Bias*. Baltimore, Johns Hopkins Press, 1982.
19. Frank JR, Jabbour M, Tugwell P, et al: Skills for the new millennium: Report of the societal needs working group, CanMEDS 2000 Project. *Ann R Coll Phys Surg Can* 1996;29:206–216.
20. General Medical Council: *Good Medical Practice*. London, General Medical Council, 2001.
21. Leach DC: A model for GME: Shifting from process to outcomes. A progress report from the Accreditation Council for Graduate Medical Education. *Med Educ* 2004;38(1):12–14.
22. American Board of Internal Medicine (ABIM): *Portfolio for Internal Medicine Residency Programs*. Philadelphia, ABIM, 2004.
23. Baker GR, Gelmon S, Headrick L, et al: Collaborating for improvement in health professions education. *Qual Manag Health Care* 1998;6(2):1–11.
24. Miller G: The assessment of clinical skills/competence/performance. *Acad Med* 1990;65(Suppl):S63–S67.
25. Rethans JJ, Norcini JJ, Barón-Maldonado M, et al: The relationship between competence and performance: Implications for assessing practice performance. *Med Educ* 2002;36:901–909.
26. Dreyfus HL: *On the Internet. Thinking in Action Series*. New York, Routledge, 2001.
27. Van der Vleuten CP, Schuwirth LW: Assessing professional competence: From methods to programmes. *Med Educ* 2005;39(3):309–317.
28. National Health Service: *Modernising Medical Careers: The Foundation Programmes*. Accessed at <http://www.mmc.nhs.uk/pages/foundation>, Nov. 17, 2006.
29. Herbers JE Jr, Noel GL, Cooper GS, et al: How accurate are faculty evaluations of clinical competence? *J Gen Intern Med* 1989;4:202–208.
30. Noel GL, Herbers JE Jr, Caplow MP, et al: How well do internal faculty members evaluate the clinical skills of residents? *Ann Intern Med* 1992;117:757–765.
31. Kroboth FJ, Hanusa BH, Parker S, et al: The inter-rater reliability and internal consistency of a clinical evaluation exercise. *J Gen Intern Med* 1992;7:174–179.
32. Landy FJ, Farr JL: *The Measurement of Work Performance: Methods, Theory and Applications*. Orlando, FL, Academic Press, 1983.
33. Holmboe ES, Hawkins RE, Huot SJ: Effects of training in direct observation of medical residents' clinical competence: A randomized trial. *Ann Intern Med* 2004;140(11):874–881.
34. Berbano EP, Browning R, Pangaro L, Jackson JL: The impact of the Stanford Faculty Development Program on ambulatory teaching behavior. *J Gen Intern Med* 2006;21(5):430–434.
35. Boulet JR, Swanson DB: Psychometric challenges of using simulations for high-stakes assessment. In Dunn D (ed): *Simulators in Critical Care Education and Beyond*. Philadelphia, Lippincott, Williams & Wilkins, 2004, pp 119–130.
36. Ziv A, Wolpe RP, Small SD, Click S: *Simulation-Based Medical Education: An Ethical Imperative*. *Acad Med* 2003;78: 783–788.
37. Norcini JJ: Current perspectives in assessment: The assessment of performance at work. *Med Educ* 2005;39:880–889.
38. Ogrinc G, Headrick LA, Foster T: Teaching and assessing resident competence in practice-based learning and improvement. *J Gen Intern Med* 2004;19(5 Pt 2):496–500.