Defining and Assessing Professional Competence

Ronald M. Epstein, MD
Edward M. Hundert, MD

Context Current assessment formats for physicians and trainees reliably test core knowledge and basic skills. However, they may underemphasize some important domains of professional medical practice, including interpersonal skills, lifelong learning, professionalism, and integration of core knowledge into clinical practice.

Objectives To propose a definition of professional competence, to review current means for assessing it, and to suggest new approaches to assessment.

Data Sources We searched the MEDLINE database from 1966 to 2001 and reference lists of relevant articles for English-language studies of reliability or validity of measures of competence of physicians, medical students, and residents.

Study Selection We excluded articles of a purely descriptive nature, duplicate reports, reviews, and opinions and position statements, which yielded 195 relevant citations.

Data Extraction Data were abstracted by 1 of us (R.M.E.). Quality criteria for inclusion were broad, given the heterogeneity of interventions, complexity of outcome measures, and paucity of randomized or longitudinal study designs.

Data Synthesis We generated an inclusive definition of competence: 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 the community being served. Aside from protecting the public and limiting access to advanced training, assessments should foster habits of learning and self-reflection and drive institutional change. Subjective, multiple-choice, and standardized patient assessments, although reliable, underemphasize important domains of professional competence: integration of knowledge and skills, context of care, information management, teamwork, health systems, and patient-physician relationships. Few assessments observe trainees in real-life situations, incorporate the perspectives of peers and patients, or use measures that predict clinical outcomes.

Conclusions In addition to assessments of basic skills, new formats that assess clinical reasoning, expert judgment, management of ambiguity, professionalism, time management, learning strategies, and teamwork promise a multidimensional assessment while maintaining adequate reliability and validity. Institutional support, reflection, and mentoring must accompany the development of assessment programs.

JAMA. 2002;287:226-235 www.jama.com

MEDICAL SCHOOLS, POSTgraduate training programs, and licensing bodies conduct assessments to certify the competence of future practitioners, discriminate among candidates for advanced training, provide motivation and direction for learning, and judge the adequacy of training programs. Standards for professional competence delineate key technical, cognitive, and emotional aspects of practice, including those that may not be measurable.[1,2] However, there is no agreed-upon definition of competence that encompasses all important domains of professional medical practice. In response, the Accreditation Council for Graduate Medical Education defined 6 areas of competence and some means of assessing them[3]: patient care (including clinical reasoning), medical knowledge, practice-based learning and improvement (including information management), interpersonal and communication skills, professionalism, and systems-based practice (including health economics and teamwork).[3]

In this article, we will advance a definition of professional competence of physicians and trainees that expands on these 6 areas, perform an evidence-based critique of current methods of assessing these areas of competence, and propose new means for assessing residents and medical students.

For editorial comment see p 243.

DEFINING PROFESSIONAL COMPETENCE

Building on prior definitions,[1-3] we propose that professional competence is *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.* Competence builds on a foundation of basic clinical skills, scientific knowledge, and moral development. It includes a cognitive function—acquiring and using knowledge to solve

Author Affiliations: Departments of Family Medicine (Dr Epstein), Psychiatry (Drs Epstein and Hundert), and Medical Humanities (Dr Hundert), University of Rochester School of Medicine and Dentistry, Rochester, NY.
Corresponding Author and Reprints: Ronald M. Epstein, MD, University of Rochester School of Medicine and Dentistry, 885 South Ave, Rochester, NY 14620 (e-mail: ronald_epstein@urmc.rochester.edu).

real-life problems; an integrative function—using biomedical and psychosocial data in clinical reasoning; a relational function—communicating effectively with patients and colleagues; and an affective/moral function—the willingness, patience, and emotional awareness to use these skills judiciously and humanely (BOX 1). Competence depends on habits of mind, including attentiveness, critical curiosity, self-awareness, and presence. Professional competence is developmental, impermanent, and context-dependent.

Box 1. Dimensions of Professional Competence

Cognitive
Core knowledge
Basic communication skills
Information management
Applying knowledge to real-world situations
Using tacit knowledge and personal experience
Abstract problem-solving
Self-directed acquisition of new knowledge
Recognizing gaps in knowledge
Generating questions
Using resources (eg, published evidence, colleagues)
Learning from experience

Technical
Physical examination skills
Surgical/procedural skills

Integrative
Incorporating scientific, clinical, and humanistic judgment
Using clinical reasoning strategies appropriately (hypothetico-deductive, pattern-recognition, elaborated knowledge)
Linking basic and clinical knowledge across disciplines
Managing uncertainty

Context
Clinical setting
Use of time

Relationship
Communication skills
Handling conflict
Teamwork
Teaching others (eg, patients, students, and colleagues)

Affective/Moral
Tolerance of ambiguity and anxiety
Emotional intelligence
Respect for patients
Responsiveness to patients and society
Caring

Habits of Mind
Observations of one's own thinking, emotions, and techniques
Attentiveness
Critical curiosity
Recognition of and response to cognitive and emotional biases
Willingness to acknowledge and correct errors

Acquisition and Use of Knowledge

Evidence-based medicine is an explicit means for generating an important answerable question, interpreting new knowledge, and judging how to apply that knowledge in a clinical setting.[4] But Polanyi[5] argues that competence is defined by tacit rather than explicit knowledge. Tacit knowledge is that which we know but normally do not explain easily, including the informed use of heuristics (rules of thumb), intuition, and pattern recognition. The assessment of evidence-based medicine skills is difficult because many of the heuristics used by novices are replaced by shortcuts in the hands of experts,[6] as are other clinical skills.[7]

Personal knowledge is usable knowledge gained through experience.[8] Clinicians use personal knowledge when they observe a patient's demeanor (such as a facial expression) and arrive at a provisional diagnosis (such as Parkinson disease) before eliciting the specific information to confirm it. Because experience does not necessarily lead to learning and competence,[9] cognitive and emotional self-awareness is necessary to help physicians question, seek new information, and adjust for their own biases.

Integrative Aspects of Care

Professional competence is more than a demonstration of isolated competencies[10]; "when we see the whole, we see its parts differently than when we see them in isolation."[11] For example, the student who can elicit historical data and physical findings, who can suture well, who knows the anatomy of the gallbladder and the bile ducts, and who can draw the biosynthetic pathway of bilirubin may not accurately diagnose and manage a patient with symptomatic gallstones. A competent clinician possesses the integrative ability to think, feel, and act like a physician.[6,12-15] Schon[16] argues that professional competence is more than factual knowledge and the ability to solve problems with clear-cut solutions: it is defined by the ability to manage *ambiguous* problems, tolerate uncertainty, and make decisions with limited information.

Competence depends on using expert scientific, clinical, and humanistic judgment to engage in clinical reasoning.[14,15,17,18] Although expert clinicians often use pattern recognition for routine problems[19] and hypothetico-deductive reasoning for complex problems outside their areas of expertise, expert clinical reasoning usually involves working interpretations[12] that are elaborated into branching networks of concepts.[20-22] These networks help professionals initiate a process of problem solving from minimal information and use subsequent information to refine their understanding of the problem. Reflection allows practitioners to examine their own clinical reasoning strategies.

Building Therapeutic Relationships

The quality of the patient-physician relationship affects health and the recovery from illness,[23,24] costs,[25] and outcomes of chronic diseases[26-29] by altering patients' understanding of their illnesses and reducing patient anxiety.[26] Key measurable patient-centered[28] (or relationship-centered)[30,31] behaviors include responding to patients' emotions and participatory decision making.[29]

Medical errors are often due to the failure of health systems rather than individual deficiencies.[32-34] Thus, the assessment of teamwork and institutional self-assessment might effectively complement individual assessments.

Affective and Moral Dimensions

Moral and affective domains of practice may be evaluated more accurately by patients and peers than by licensing bodies or superiors.[35] Only recently have validated measures captured some of the intangibles in medicine, such as trust[36,37] and professionalism.[38,39] Recent neurobiological research indicates that the emotions are central to all judgment and decision making,[13] further emphasizing the importance of assessing emotional intelligence and self-awareness in clinical practice.[1,40-42]

Habits of Mind

Competence depends on habits of mind that allow the practitioner to be attentive, curious, self-aware, and willing to recognize and correct errors.[43] Many physicians would consider these habits of mind characteristic of good practice, but they are especially difficult to objectify. A competent physician, for example, should be able to judge his or her level of anxiety when facing an ambiguous clinical presentation and be aware of how the anxiety of uncertainty may be influencing his or her clinical judgment. Errors in medicine, according to this view, may result from overcertainty that one's impressions are beyond doubt.[41,43,44]

Context

Competence is context-dependent. Competence is a statement of relationship between an ability (in the person), a task (in the world),[45] and the ecology of the health systems and clinical contexts in which those tasks occur.[46,47] This view stands in contrast to an abstract set of attributes that the physician possesses—knowledge, skills, and attitudes—that are assumed to serve the physician well in all the situations that he or she encounters. For example, rather than assessing a student's competence in diagnosing and treating heart disease (a disease-specific domain) by dividing it into competencies (physical examination, interpretation of electrocardiogram, and pharmacology of β-blockers), our view is that competence is defined by the interaction of the task (the concrete process of diagnosing and treating Mrs Brown, a 52-year-old business executive who is now in the emergency department because of new-onset chest pain), the clinician's abilities (eliciting information, forming a therapeutic relationship, performing diagnostic maneuvers, and making judgments about treatment), and the health system (good insurance and ready access to care). Caring for Mrs Brown requires different skills than caring for Ms Hall, a 52-year-old uninsured homeless woman who has similar symptoms and receives episodic care at an inner-city clinic.

Development

Competence is developmental. There is debate about which aspects of competence should be acquired at each stage of training. For example, early clinical experiences and problem-based learning formats encourage clinical reasoning skills formerly relegated to the final years of medical school. But students tend to use the same cognitive strategy for solving all problems, whereas experts draw on several strategies,[6] which raises the question of whether assessment of practicing physicians should be qualitatively different from the assessment of a student. Determining how and at what level of training the patient-physician relationship should be assessed is also difficult. For example, participatory decision making correlates with clinical outcomes,[25,29] but it is unclear when trainees should be assessed on this skill. Although a third-year resident might be expected to counsel a fearful diabetic patient about the need to start insulin, a third-year student might be expected only to elicit the patient's preferences, emotions, and expectations. Changes in medical practice and the context of care invite redefinitions of competence; for example, the use of electronic communication media[48] and changes in patient expectations.[49,50]

CURRENT MEANS OF ASSESSMENT

Assessment must take into account what is assessed, how it is assessed, and the assessment's usefulness in fostering future learning. In discussing validity of measures of competence in an era when reliable assessments of core knowledge, abstract problem solving, and basic clinical skills have been developed,[45,51-56] we must now establish that they encompass the qualities that define a good physician: the cognitive, technical, integrative, contextual, relational, reflective, affective, and moral aspects of competence. We distinguish between expert opinion, in-

termediate outcomes, and the few studies that show associations between results of assessments and actual clinical performance.[57-60]

We consider how the process of assessment might foster future learning. Too often, practitioners select educational programs that are unlikely to influence clinical practice.[61] Good assessment is a form of learning and should provide guidance and support to address learning needs. Finally, we address concerns that the medical profession still lacks adequate accountability to the public[62] and has not done enough to reduce medical errors.[32,63]

Within each domain of assessment, there are 4 levels at which a trainee might be assessed (FIGURE).[64] The *knows* level refers to the recall of facts, principles, and theories. The *knows how* level involves the ability to solve problems and describe procedures. The *shows how* level usually involves human (standardized patient), mechanical, or computer simulations that involve demonstration of skills in a controlled setting. The *does* level refers to observations of real practice. For each of these levels, the student can demonstrate the ability to imitate or replicate a protocol, apply principles in a familiar situation, adapt principles to new situations, and associate new knowledge with previously learned principles.[65]

METHODS

Using the MEDLINE database for 1966 to 2001, we searched for articles that studied the reliability or validity of measures of clinical or professional competence of physicians, medical students, and residents. An initial search using the following Medical Subject Headings of the National Library of Medicine yielded 2266 references: *educational measurement*, *patient simulation*, *clinical competence* OR *professional competence* AND *reproducibility of results*, *validity* OR *research*, OR the text word *reliability*. This set was reduced by including any of 20 text words describing assessment techniques; we used words such as *OSCE*, *oral examination*, *peer assessment*, *triple jump*, *essay*, *portfolio*, and *CEX* (clinical evaluation exercise), yielding 430 references. Articles of a purely descriptive nature, reviews that offered no new data, and opinions and position statements were excluded, yielding 101 English-language references. We surveyed the

Figure. A Framework for Assessment

		LEVEL OF ASSESSMENT				CONTEXT OF CARE				
		Knows	Knows how	Shows how	Does	New problem	Chronic illness	Emergency	Preventive	Acute hospital
CLINICAL TASKS	Self-assessment and reflection									
	Information gathering from patients and families									
	Relationship-building and professionalism									
	Sharing information, behavior change, and patient involvement									
	Physical examination									
	Patient procedural skills (suturing, drawing blood)									
	Interpretation of diagnostic tests (electrocardiogram, mental status, imaging)									
	Diagnostic reasoning: Psychosocial issues									
	Diagnostic reasoning: Biomedical issues									
	Diagnostic reasoning: Diagnostic uncertainty									
	Clinical judgment: Planning further diagnostic workup									
	Clinical judgment: Generating therapeutic plan									
	Accessing, interpreting, and applying the medical literature									
	Presenting data to colleagues (referral letter, chart note)									
KNOWLEDGE CONTENT AREAS	Basic mechanisms (anatomy, immunology, microbiology)									
	Pathophysiology of disease (dermatology, renal, gastrointestinal)									
	Social science (epidemiology, psychology, culture/diversity)									
	Special topics (spirituality, ethics, economics)									

The grid is filled out according to the type of assessment conducted, ie, standardized patient or simulation, video, postencounter probe, essay, or computer exercises. Each category can be combined with a number designating a category such as the name of a patient, a type of computer exercise, or a team exercise.

first 200 of the 2165 references excluded and found none that met our search criteria. Quality criteria for inclusion were broad, given the small number of controlled trials of assessment interventions and the complexity of outcome measures. Because we knew that MEDLINE search strategies would not capture all relevant studies, we searched reference lists in the 101 articles, other review articles, and books and did additional literature searches using the key authors of recent reviews; we gathered 94 additional relevant references. Of the 195 references, 124 presented new data on assessment of physicians.

Summary of Studies

The 3 most commonly used assessment methods are subjective assessments by supervising clinicians, multiple-choice examinations to evaluate factual knowledge and abstract problem solving,[66] and standardized patient assessments of physical examination and technical and communication skills.[67-69] Although curricular designs increasingly integrate core knowledge and clinical skills, most assessment methods evaluate these domains in isolation. Few assessments use measures such as participatory decision making[70] that predict clinical outcomes in real practice. Few reliably assess clinical reasoning, systems-based care, technology, and the patient-physician relationship.[3,69] The literature makes important distinctions between criteria for licensing examinations and program-specific assessments with mixed formative and summative goals.

Evaluation of factual knowledge and problem-solving skills by using multiple-choice questions offers excellent reliability[71-75] and assesses some aspects of context and clinical reasoning. Scores on Canadian licensing examinations, which include standardized patient assessment and multiple-choice tests, correlated positively with subsequent appropriate prescribing, mammographic screening, and referrals,[38] and multiple-choice certification examination scores correlated with subsequent faculty[76] and peer[77] ratings. Many have questioned the validity of multiple-choice examinations, though.[78-81] For example, compared with Florida family physicians who are not board-certified, those who are have nearly twice the risk of being sued.[82] Standardized test scores have been inversely correlated with empathy, responsibility, and tolerance.[83] Also, because of lack of expertise and resources, few medical school examinations can claim to achieve the high psychometric standards of the licensing boards.

The Objective Structured Clinical Examination (OSCE) is a timed multistation examination often using standardized patients (SPs) to simulate clinical scenarios. The roles are portrayed accurately[56,84] and simulations are convincing; the detection rate of unannounced SPs in community practice is less than 10%.[57,59,85-89] Communication, physical examination, counseling, and technical skills can be rated reliably if there is a sufficiently large number of SP cases[67,90-100] and if criteria for competence are based on evidence.[101] Although few cases are needed to assess straightforward skills, up to 27 cases may be necessary to assess interpersonal skills reliably in high-stakes examinations.[102,103] Although SPs' ratings usually correlate with those of real patients,[104] differences have been noted.[105-107]

Defining pass/fail criteria for OSCEs has been complex.[54,108-111] There is debate about who should rate student performance in an OSCE.[112] Ratings by the SP are generally accurate[52] but may be hampered by memory failure, whereas external raters, either physicians or other SPs, may be less attuned to affective aspects of the interview and significantly increase the cost of the examination.

Checklist scores completed by physician-examiners in some studies improve with expertise of the examinees[113] and with the reputation of the training program.[90,114] But global rating scales of interpersonal skills may be more valid than behavioral checklists.[7,115,116] The OSCE scores may not correlate with multiple-choice examinations and academic grades,[90,100,117] suggesting that these tools measure different skills. Clinicians may behave differently in examination settings than in real practice,[106,118] and short OSCE stations can risk fragmentation and trivialization of isolated elements of what should be a coherent whole.[119] The OSCE also has low test reliability for measuring clinical ethics.[120]

There are few validated strategies to assess actual clinical practice, or Miller's *does* level. Subjective evaluation by residents and attending physicians is the major form of assessment during residency and the clinical clerkships and often includes the tacit elements of professional competence otherwise overlooked by objective assessment instruments. Faculty ratings of humanism predicted patient satisfaction in one study.[121] However, evaluators often do not observe trainees directly. They often have different standards[122,123] and are subject to halo effects[124] and racial and sex bias.[125,126] Because of interpatient variability and low interrater reliability, each trainee must be subject to multiple assessments for patterns to emerge. Standardized rating forms for direct observation of trainees[127-132] and structured oral examination formats have been developed in response to this criticism.[133,134]

The Royal College of General Practitioners, dissatisfied with the capability of the OSCE to evaluate competence for the final professional licensing examination, developed a format in which candidates for certification present several best-case videotapes of their performance in real clinical settings to a trained examiner who uses specified criteria for evaluation.[135] Although the face validity of such a measure is high and the format is well accepted by physicians,[136] the number of cases that should be presented to achieve adequate reliability is unclear.[137-139]

Profiling by managed-care databases is increasingly used as an evaluation measure of clinical competence. However, data abstraction is complex[140] and defining competence in

terms of cost and value is difficult. The underlying assumptions driving such evaluation systems may not be explicit. For example, cost analyses may favor physicians caring for more highly educated patients.[141]

Peer ratings are accurate and reliable measures of physician performance.[77,142] Peers may be in the best position to evaluate professionalism; people often act differently when not under direct scrutiny.[143] Anonymous medical student peer assessments of professionalism have raised awareness of professional behavior, fostered further reflection, helped students identify specific mutable behaviors, and been well accepted by students.[35] Students should be assessed by at least 8 of their classmates. The composite results should be edited to protect the confidentiality of the raters.

Self-assessments have been used with some success in standardized patient exercises[144] and in programs that offer explicit training in the use of self-assessment instruments.[145] Among trainees who did not have such training, however, self-assessment was neither valid nor accurate. Rather, it was more closely linked to the trainee's psychological sense of self-efficacy and self-confidence than to appropriate criteria, even among bright and motivated individuals.

Box 2. Some Purposes of Assessment

For the Trainee

Provide useful feedback about individual strengths and weaknesses that guides future learning
Foster habits of self-reflection and self-remediation
Promote access to advanced training

For the Curriculum

Respond to lack of demonstrated competence (denial of promotion, mandated remediation)
Certify achievement of curricular goals
Foster course or curricular change
Create curricular coherence
Cross-validate other forms of assessment in the curriculum
Establish standards of competence for trainees at different levels

For the Institution

Guide a process of institutional self-reflection and remediation
Discriminate among candidates for further training or promotion
Express institutional values by determining what is assessed and how assessment is conducted
Develop shared educational values among a diverse community of educators
Promote faculty development
Provide data for educational research

For the Public

Certify competence of graduates

COMMENT

Aside from the need to protect the public by denying graduation to those few trainees who are not expected to overcome their deficiencies, the outcomes of assessment should foster learning, inspire confidence in the learner, enhance the learner's ability to self-monitor, and drive institutional self-assessment and curricular change. Given the difficulty in validating tests of basic skills, it is not surprising that there is scant literature on the assessment of learning, professionalism, teamwork, and systems-based care or on the ability of assessment programs to drive curricular change or reduce medical errors.

Assessment serves personal, institutional, and societal goals (BOX 2). Distinctions between these goals often are blurred in practice. For example, formative feedback is intended to foster individual reflection and remediation[146] but may be perceived as having evaluative consequences. Summative evaluation is a powerful means for driving curricular content and what students learn. Assessment provides information to allow institutions to choose among candidates for advanced training. The public expects greater self-monitoring, communication, and teamwork from health care practitioners.[147] The decline of public trust in medicine may reflect a growing concern that physicians are not achieving these goals.[36]

Assessment is also a statement of institutional values. Devoting valuable curricular time to peer assessment of professionalism, for example, can promote those values that are assessed while encouraging curricular coherence and faculty development, especially if there are corresponding efforts at the institution toward self-assessment and change.

Whereas performance is directly measurable, competence is an inferred quality.[148] Performance on a multiple-choice test may exceed competence, as in the case of a trainee with a photographic memory but poor clinical judgment. Conversely, competence may exceed test performance, as in the case of a trainee with severe test anxiety. Correlation with National Board scores and feedback on graduates' performance can be useful in validating some assessment instruments but should be done with caution. For example, efficiency is highly valued in residents but less so in medical students.

Future Directions

Medical schools in Canada, the United Kingdom, Australia, Spain, the Netherlands, and the United States have made commitments to developing innovative assessments of professional competence, some of which we describe. These assessments are increasingly multimodal and tailored to the goals and context in which they will be

Box 3. Innovations in Assessing Professional Competence

Multimethod assessment
Clinical reasoning in situations that involve clinical uncertainty
Standardized patient exercises linked to postencounter probes of pathophysiology and clinical reasoning
Exercises to assess use of the medical literature
Long-station standardized patient exercises
Simulated continuity
Teamwork exercises
Unannounced standardized patients in clinical settings
Assessments by patients
Peer assessment of professionalism
Portfolios of videotapes
Mentored self-assessment
Remediation based on a learning plan

used. Large-scale licensure examinations must use computer-gradable formats, but comprehensive examinations using structured direct observation,[107] OSCE stations, real patient cases,[107] case-based questions,[79] peer assessments, and essay-type questions[149] are reliable as well. Proponents of the new formats argue that they provide more useful feedback and are more efficient at the medical school or residency level (Box 1 and Box 3) than traditional formats.[81,150] They target core knowledge and clinical skills in different contexts and at different levels of assessment. Because of their complexity, a matrix (Figure) can be useful to display the domains assessed.

Comprehensive assessments link content across several formats. Postencounter probes immediately after SP exercises using oral, essay, or multiple-choice questions test pathophysiology and clinical reasoning in context.[151,152] Triple-jump exercises[152]—consisting of a case presentation, an independent literature search, and then an oral or written postencounter examination—test the use and application of the medical literature. Validated measures of reflective thinking[153] have been developed that use patient vignettes followed by questions that require clinical judgment. These measures reflect students' capacity to organize and link information; also, they predict clinical reasoning ability 2 years later.[153] Combining formats appears to have added value with no loss in reliability.[150,154] Ongoing educational outcomes research will show whether composite formats help students learn how to learn more effectively, develop habits of mind that characterize exemplary practice,[43] and provide a more multidimensional picture of the examinee than the individual unlinked elements. Two examples of comprehensive assessment formats follow.

Genetics, Evidence-Based Medicine, Screening, and Communication. A student is instructed to perform a literature search about genetic screening test for Alzheimer disease in anticipation of an SP encounter later that day. Assessment instruments include a structured evaluation of the search strategy and a communication rating scale, completed by an SP, that assesses the clarity of the student's presentation and the student's ability to involve the patient in the decision-making process. Next, the student completes an essay about the ethics of genetic screening and the genetics of Alzheimer disease. This exercise assesses the student's communication skills, clinical reasoning, ability to acquire and use new knowledge, and contextualized use of knowledge of genetics, health economics, and medical ethics.

Cognitive and Affective Challenges of Clinical Uncertainty. A rating scale is used to assess a resident on her ability to agree on a plan of action with an SP who portrays an outpatient demanding a computed tomographic scan for headaches without neurological signs. In a postencounter exercise, the resident creates a rank-order differential diagnosis and then answers a series of script concordance[153,155] questions in which the examinee is presented hypothetical additional data (for example, numbness in the left hand) and then asked to judge how her diagnostic hypotheses or therapeutic actions would change. Failure to include a key diagnostic possibility or the overestimation or underestimation of probability are criteria for evaluation. The goal of the exercise is to demonstrate emotional intelligence[40] and self-awareness in the context of conflict and ambiguity. Similar observations might be made with trainees' video portfolios of real clinical encounters.

Well-functioning health systems are characterized by continuity, partnership between physicians and patients, teamwork between health care practitioners, and communication between health care settings.[156,157] The use of time in a continuity relationship can be assessed with a series of SP or real-patient exercises. To assess partnership, patient assessment, currently used to assess physicians in practice,[158] is being tested for students and residents.[159,160] These efforts are guided by data showing that patients' ratings of communication and satisfaction correlate well with biomedical outcomes,[24,29] emotional distress,[161] health care use,[25] and malpractice litigation.[162] Patient ratings also have the potential to validate other measures of competence.[163] Several institutions assess teamwork by using peer assessments. Others use sophisticated mannequins to simulate acute cardiovascular physiological derangements found in intensive care settings[164-169]; trainees are graded on teamwork as well as individual problem solving, and statistical adjustments can account for team composition. Communication between health settings could be assessed at the student level, for example, by grading of their written referral letters.[170]

Although it could be argued that licensing boards do not have the mandate to remediate examinees who perform poorly or modify educational curricula, medical schools and residency programs do. Tests that demonstrate students' strengths or weaknesses may not provide the student with the opportunity to reflect on actual behaviors and patterns of thought that

should be changed. To foster reflection and action, some institutions require a learning plan in which trainees chart their learning needs, the means of achieving them, expected time of completion, and means of verification[146,171] as a required outcome of an assessment.

A strong mentoring system should accompany any comprehensive assessment program. An inadequate system for feedback, mentoring, and remediation will subvert even the most well-conceived and validated examination. Curricular change also can be guided by results of assessments but requires a parallel process of institutional reflection, feedback, and remediation.

These new assessment formats are feasible, and several institutions have invested significant time and resources to develop them. The promise that a more comprehensive assessment of professional competence might improve practice, change medical education, and reduce medical errors should be studied in controlled trials. The public's trust in the medical profession and the ability of medical practitioners to learn from mistakes depends on valid and reliable means of assessment. Medical educators, professional societies, and licensing boards should view professional competence more comprehensively to improve the process of assessment.

Author Contributions: *Study concept and design, critical revision of the manuscript for important intellectual content, and administrative, technical, or material support:* Epstein, Hundert.
Acquisition of data, analysis and interpretation of data, and drafting of the manuscript: Epstein.
Acknowledgment: We would like to express thanks to Francesc Borrell-Carrio, MD, Daniel Federman, MD, Brian Hodges, MD, Daniel Klass, MD, Larry Mauksch, CSW, Timothy Quill, MD, Andres Sciolla, MD, and Kevin Volkan, PhD, for their critical review of the manuscript. Also we would like to acknowledge Anthony LaDuca, PhD, and Albert Oriol-Bosch, MD, for their contributions to our formulation of professional competence.

REFERENCES

1. American Board of Internal Medicine. *Guide to Evaluation of Residents in Internal Medicine*. Philadelphia, Pa: American Board of Internal Medicine; 1999.
2. Norman GR. Defining competence: a methodological review. In: Neufeld VR, Norman GR, eds. *Assessing Clinical Competence*. New York, NY: Springer; 1985:15-35.
3. ACGME Outcome Project. Accreditation Council for Graduate Medical Education Web site. Available at: http://www.acgme.org. 2000. Accessed October 1, 2001.
4. Sackett DL. *Evidence-Based Medicine: How to Practice and Teach EBM*. New York, NY: Churchill Livingstone; 1997.
5. Polanyi M. *Personal Knowledge: Towards a Post-Critical Philosophy*. Chicago, Ill: University of Chicago Press; 1974.
6. Benner P. *From Novice to Expert*. Menlo Park, Calif: Addison-Wesley; 1984.
7. Hodges B, Regehr G, McNaughton N, Tiberius R, Hanson M. OSCE checklists do not capture increasing levels of expertise. *Acad Med*. 1999;74:1129-1134.
8. Dewey J. *Experience and Nature*. 1932 ed. New York, NY: Dover; 1958.
9. Langer EJ. *The Power of Mindful Learning*. Reading, Mass: Perseus Books; 1997.
10. Eraut M. Learning professional processes: public knowledge and personal experience. In: Eraut M, ed. *Developing Professional Knowledge and Competence*. London, England: Falmer Press; 1994:100-122.
11. Polanyi M. The logic of tacit inference. In: Grene M, ed. *Knowing and Being: Essays*. Chicago, Ill: University of Chicago; 1969:123-158.
12. Gale J, Marsden P. Clinical problem solving: the beginning of the process. *Med Educ*. 1982;16:22-26.
13. Damasio AR. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York, NY: GP Putnam's Sons; 1994.
14. Mandin H, Jones A, Woloschuk W, Harasym P. Helping students learn to think like experts when solving clinical problems. *Acad Med*. 1997;72:173-179.
15. Friedman MH, Connell KJ, Olthoff AJ, Sinacore JM, Bordage G. Medical student errors in making a diagnosis. *Acad Med*. 1998;73(suppl):S19-S21.
16. Schon DA. *The Reflective Practitioner*. New York, NY: Basic Books; 1983.
17. Feinstein AR. "Clinical Judgment" revisited: the distraction of quantitative models. *Ann Intern Med*. 1994;120:799-805.
18. Downie RS, Macnaughton J, Randall R. *Clinical Judgement: Evidence in Practice*. Oxford, England: Oxford University Press; 2000.
19. Dunn MM, Woolliscroft JO. Assessment of a pattern-recognition examination in a clinical clerkship. *Acad Med*. 1994;69:683-684.
20. Schmidt HG, Norman GR, Boshuizen HP. A cognitive perspective on medical expertise: theory and implication [published correction appears in *Acad Med*. 1992;67:287]. *Acad Med*. 1990;65:611-621.
21. Bordage G, Zacks R. The structure of medical knowledge in the memories of medical students and general practitioners: categories and prototypes. *Med Educ*. 1984;18:406-416.
22. Bordage G. Elaborated knowledge: a key to successful diagnostic thinking. *Acad Med*. 1994;69:883-885.
23. Starfield B, Wray C, Hess K, Gross R, Birk PS, D'Lugoff BC. The influence of patient-practitioner agreement on outcome of care. *Am J Public Health*. 1981;71:127-131.
24. The Headache Study Group of the University of Western Ontario. Predictors of outcome in headache patients presenting to family physicians—a one year prospective study. *Headache*. 1986;26:285-294.
25. Stewart M, Brown JB, Donner A, et al. The impact of patient-centered care on outcomes. *J Fam Pract*. 2000;49:796-804.
26. Di Blasi Z, Harkness E, Ernst E, Georgiou A, Kleijnen J. Influence of context effects on health outcomes: a systematic review. *Lancet*. 2001;357:757-762.
27. Stewart M. Effective physician-patient communication and health outcomes: a review. *CMAJ*. 1995;152:1423-1433.
28. Stewart M, Brown JB, Boon H, Galajda J, Meredith L, Sangster M. Evidence on patient-doctor communication. *Cancer Prev Control*. 1999;3:25-30.
29. Kaplan SH, Greenfield S, Ware JE Jr. Assessing the effects of physician-patient interactions on the outcomes of chronic disease [published correction appears in *Med Care*. 1989;27:679]. *Med Care*. 1989;27:S110-S127.
30. Tresolini C, and the Pew-Fetzer Task Force. *Health Professions Education and Relationship-Centered Care: Report of the Pew-Fetzer Task Force on Advancing Psychosocial Health Education*. San Francisco, Calif: Pew Health Professions Commission; 1994.
31. Candib LM. *Medicine and the Family: A Feminist Perspective*. New York, NY: Basic Books; 1995.
32. Ternov S. The human side of medical mistakes. In: Spath PL, ed. *Error Reduction in Healthcare: A Systems Approach to Improving Patient Safety*. San Francisco, Calif: Jossey Bass; 1999:97-138.
33. Chopra V, Bovill JG, Spierdijk J, Koornneef F. Reported significant observations during anaesthesia: a prospective analysis over an 18-month period. *Br J Anaesth*. 1992;68:13-17.
34. Cooper JB, Newbower RS, Long CD, McPeek B. Preventable anesthesia mishaps: a study of human factors. *Anesthesiology*. 1978;49:399-406.
35. Asch E, Saltzberg D, Kaiser S. Reinforcement of self-directed learning and the development of professional attitudes through peer- and self-assessment. *Acad Med*. 1998;73:575.
36. Safran DG, Montgomery JE, Chang H, Murphy J, Rogers WH. Switching doctors: predictors of voluntary disenrollment from a primary physician's practice. *J Fam Pract*. 2001;50:130-136.
37. Safran DG, Taira DA, Rogers WH, Kosinski M, Ware JE, Tarlov AR. Linking primary care performance to outcomes of care. *J Fam Pract*. 1998;47:213-220.
38. American Board of Internal Medicine. *Project Professionalism*. Philadelphia, Pa: American Board of Internal Medicine; 1998.
39. Peabody F. The care of the patient. *JAMA*. 1927;88:877-882.
40. Goleman D. *Emotional Intelligence*. New York, NY: Bantam Books; 1995.
41. Novack DH, Suchman AL, Clark W, Epstein RM, Najberg E, Kaplan C. Calibrating the physician: personal awareness and effective patient care. *JAMA*. 1997;278:502-509.
42. Ramsey PG, Wenrich M, Carline JD, Inui TS, Larson EB, LoGerfo JP. Use of peer ratings to evaluate physician performance. *JAMA*. 1993;269:1655-1660.
43. Epstein RM. Mindful practice. *JAMA*. 1999;282:833-839.
44. Novack DH, Epstein RM, Paulsen RH. Toward creating physician-healers: fostering medical students' self-awareness, personal growth, and well-being. *Acad Med*. 1999;74:516-520.
45. Klass D. Reevaluation of clinical competency. *Am J Phys Med Rehabil*. 2000;79:481-486.
46. LaDuca A, Taylor DD, Hill IK. The design of a new physician licensure examination. *Eval Health Prof*. 1984;7:115-140.
47. Kane M. Model-based practice analysis and test specifications. *Appl Meas Educ*. 1997;10:18.
48. Dugdale DC, Epstein R, Pantilat SZ. Time and the patient-physician relationship. *J Gen Intern Med*. 1999;14(suppl 1):S34-S40.
49. Emanuel EJ, Dubler NN. Preserving the physician-patient relationship in the era of managed care. *JAMA*. 1995;273:323-335.
50. Kravitz RL, Cope DW, Bhrany V, Leake B. Internal medicine patients' expectations for care during office visits. *J Gen Intern Med*. 1994;9:75-81.
51. LaDuca A. Validation of professional licensure ex-

aminations: professions theory, test design, and construct validity. *Eval Health Prof*. 1994;17:178-197.
52. De Champlain AF, Margolis MJ, King A, Klass DJ. Standardized patients' accuracy in recording examinees' behaviors using checklists. *Acad Med*. 1997; 72(suppl 1):S85-S87.
53. Kopelow ML, Schnabl GK, Hassard TH, et al. Assessing practicing physicians in two settings using standardized patients. *Acad Med*. 1992;67(suppl):S19-S21.
54. Margolis MJ, De Champlain AF, Klass DJ. Setting examination-level standards for a performance-based assessment of physicians' clinical skills. *Acad Med*. 1998;73(suppl):S114-S116.
55. Ross LP, Clauser BE, Margolis MJ, Orr NA, Klass DJ. An expert-judgment approach to setting standards for a standardized-patient examination. *Acad Med*. 1996;71(suppl):S4-S6.
56. Tamblyn RM, Klass DJ, Schnabl GK, Kopelow ML. The accuracy of standardized patient presentation. *Med Educ*. 1991;25:100-109.
57. McLeod PJ, Tamblyn RM, Gayton D, et al. Use of standardized patients to assess between-physician variations in resource utilization. *JAMA*. 1997;278: 1164-1168.
58. Tamblyn R, Abrahamowicz M, Brailovsky C, et al. Association between licensing examination scores and resource use and quality of care in primary care practice. *JAMA*. 1998;280:989-996.
59. Tamblyn RM. Use of standardized patients in the assessment of medical practice. *CMAJ*. 1998;158: 205-207.
60. Abrahamowicz M, Tamblyn RM, Ramsay JO, Klass DK, Kopelow ML. Detecting and correcting for rater-induced differences in standardized patient tests of clinical competence. *Acad Med*. 1990;65(suppl):S25-S26.
61. Davis D. Does CME work? an analysis of the effect of educational activities on physician performance or health care outcomes. *Int J Psychiatry Med*. 1998;28: 21-39.
62. Kassirer JP. Pseudoaccountability. *Ann Intern Med*. 2001;134:587-590.
63. Kohn LT, Corrigan JM, Donaldson MS. *To Err Is Human: Building a Safer Health System*. Washington, DC: National Academy Press; 2000.
64. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med*. 1990;65(suppl): S63-S67.
65. Broudy HS, Smith BO, Burnett J. *Democracy and Excellence in American Secondary Education*. Chicago, Ill: Rand McNally; 1964.
66. Newble D, Dawson B, Dauphinee D. Guidelines for assessing clinical competence. *Teaching Learning Med*. 1994;6:213-220.
67. Colliver JA, Swartz MH. Assessing clinical performance with standardized patients. *JAMA*. 1997;278: 790-791.
68. Barrows HS. An overview of the uses of standardized patients for teaching and evaluating clinical skills: AAMC. *Acad Med*. 1993;68:443-451.
69. Kassebaum DG, Eaglen RH. Shortcomings in the evaluation of students' clinical skills and behaviors in medical school. *Acad Med*. 1999;74:842-849.
70. Greenfield S, Kaplan S, Ware JE Jr. Expanding patient involvement in care: effects on patient outcomes. *Ann Intern Med*. 1985;102:520-528.
71. Case SM, Becker DF, Swanson DB. Relationship between scores on NBME basic science subject tests and the first administration of the newly designed NBME Part I examination. *Acad Med*. 1992;67(suppl): S13-S15.
72. Norcini JJ, Swanson DB, Grosso LJ, Shea JA, Webster GD. A comparison of knowledge, synthesis, and clinical judgment: multiple-choice questions in the assessment of physician competence. *Eval Health Prof*. 1984;7:485-499.
73. Norcini JJ, Swanson DB, Grosso LJ, Webster GD. Reliability, validity and efficiency of multiple choice question and patient management problem item formats in assessment of clinical competence. *Med Educ*. 1985;19:238-247.
74. Case SM, Ripkey DR, Swanson DB. The relationship between clinical science performance in 20 medical schools and performance on Step 2 of the USMLE licensing examination: 1994-95 Validity Study Group for USMLE Step 1 and 2 Pass/Fail Standards. *Acad Med*. 1996;71(suppl):S31-S33.
75. Case SM, Swanson DB. Validity of NBME Part I and Part II scores for selection of residents in orthopaedic surgery, dermatology, and preventive medicine. *Acad Med*. 1993;68(suppl):S51-S56.
76. Downing SM. The assessment of clinical competence on the emergency medicine specialty certification examination: the validity of clinically relevant multiple-choice items. *Ann Emerg Med*. 1980;9:554-556.
77. Ramsey PG, Wenrich MD, Carline JD, Inui TS, Larson EB, LoGerfo JP. Use of peer ratings to evaluate physician performance. *JAMA*. 1993;269:1655-1660.
78. Holsgrove GJ. Guide to postgraduate exams: multiple-choice questions. *Br J Hosp Med*. 1992;48:757-761.
79. Schuwirth LW, Verheggen MM, van der Vleuten C, Boshuizen HP, Dinant GJ. Do short cases elicit different thinking processes than factual knowledge questions do? *Med Educ*. 2001;35:348-356.
80. Veloski JJ, Rabinowitz HK, Robeson MR, Young PR. Patients don't present with five choices: an alternative to multiple-choice tests in assessing physicians' competence. *Acad Med*. 1999;74:539-546.
81. van der Vleuten C. Validity of final examinations in undergraduate medical training. *BMJ*. 2000;321: 1217-1219.
82. Ely JW, Dawson JD, Young PR, et al. Malpractice claims against family physicians: are the best doctors sued more? *J Fam Pract*. 1999;48:23-30.
83. Tutton PJ. Psychometric test results associated with high achievement in basic science components of a medical curriculum. *Acad Med*. 1996;71:181-186.
84. Tamblyn RM, Klass DK, Schanbl GK, Kopelow ML. Factors associated with the accuracy of standardized patient presentation. *Acad Med*. 1990;65(suppl):S55-S56.
85. Tamblyn R, Berkson L, Dauphinee WD, et al. Unnecessary prescribing of NSAIDs and the management of NSAID-related gastropathy in medical practice. *Ann Intern Med*. 1997;127:429-438.
86. Carney PA, Allen JD, Eliassen MS, Owen M, Badger LW. Recognizing and managing depression in primary care: a standardized patient study. *J Fam Pract*. 1999;48:965-972.
87. Carney PA, Dietrich AJ, Freeman DH Jr, Mott LA. The periodic health examination provided to asymptomatic older women: an assessment using standardized patients. *Ann Intern Med*. 1993;119:129-135.
88. Carney PA, Dietrich AJ, Freeman DH Jr, Mott LA. A standardized-patient assessment of a continuing medical education program to improve physicians' cancer-control clinical skills. *Acad Med*. 1995;70:52-58.
89. Tamblyn RM, Klass DJ, Schnabl GK, Kopelow ML. The accuracy of standardized patient presentation. *Med Educ*. 1991;25:100-109.
90. van der Vleuten C, Sawnson DB. Assessment of clinical skills with standardized patients: state of the art. *Teaching Learning Med*. 1990;2:58-76.
91. Grand'Maison P, Lescop J, Rainsberry P, Brailovsky CA. Large-scale use of an objective, structured clinical examination for licensing family physicians. *CMAJ*. 1992;146:1735-1740.
92. Joorabchi B. Objective structured clinical examination in a pediatric residency program. *AJDC*. 1991; 145:757-762.
93. Latif A. An examination of the examinations: the reliability of the objective structured clinical examination and clinical examination. *Med Teacher*. 1992;14: 179-183.
94. MacRae HM, Cohen R, Regehr G, Reznick R, Burnstein M. A new assessment tool: the patient assessment and management examination. *Surgery*. 1997; 122:335-343.
95. Matsell DG, Wolfish NM, Hsu E. Reliability and validity of the objective structured clinical examination in paediatrics. *Med Educ*. 1991;25:293-299.
96. Vu NV, Barrows HS, Marcy ML, Verhulst SJ, Colliver JA, Travis T. Six years of comprehensive, clinical, performance-based assessment using standardized patients at the Southern Illinois University School of Medicine. *Acad Med*. 1992;67:42-50.
97. Vu NV, Distlehorst LH, Verhulst SJ, Colliver JA. Clinical performance-based test sensitivity and specificity in predicting first-year residency performance. *Acad Med*. 1993;68(suppl):S41-S45.
98. Newble DI, Swanson DB. Psychometric characteristics of the objective structured clinical examination. *Med Educ*. 1988;22:325-334.
99. Petrusa ER, Blackwell TA, Ainsworth MA. Reliability and validity of an objective structured clinical examination for assessing the clinical performance of residents. *Arch Intern Med*. 1990;150:573-577.
100. Gomez JM, Prieto L, Pujol R, et al. Clinical skills assessment with standardized patients. *Med Educ*. 1997;31:94-98.
101. Gorter S, Rethans JJ, Scherpbier A, et al. Developing case-specific checklists for standardized-patient-based assessments in internal medicine: a review of the literature. *Acad Med*. 2000;75:1130-1137.
102. Swartz MH, Colliver JA, Bardes CL, Charon R, Fried ED, Moroff S. Validating the standardized-patient assessment administered to medical students in the New York City Consortium. *Acad Med*. 1997; 72:619-626.
103. Carline JD, Paauw DS, Thiede KW, Ramsey PG. Factors affecting the reliability of ratings of students' clinical skills in a medicine clerkship. *J Gen Intern Med*. 1992;7:506-510.
104. Sanson-Fisher RW, Poole AD. Simulated patients and the assessment of medical students' interpersonal skills. *Med Educ*. 1980;14:249-253.
105. Tamblyn R, Abrahamowicz M, Schnarch B, et al. Can standardized patients predict real-patient satisfaction with the doctor-patient relationship? *Teaching Learning Med*. 1994;6:36-44.
106. Pieters HM, Touw-Otten FW, De Melker RA. Simulated patients in assessing consultation skills of trainees in general practice vocational training: a validity study. *Med Educ*. 1994;28:226-233.
107. Wass V, Jones R, van der Vleuten C. Standardized or real patients to test clinical competence? the long case revisited. *Med Educ*. 2001;35:321-325.
108. Cusimano MD, Rothman A, Keystone J. Defining standards of competent performance on an OSCE. *Acad Med*. 1998;73(suppl):S112-S113.
109. Kaufman DM, Mann KV, Muijtjens AM, van der Vleuten C. A comparison of standard-setting procedures for an OSCE in undergraduate medical education. *Acad Med*. 2000;75:267-271.
110. Travis TA, Colliver JA, Robbs RS, et al. Validity of simple approach to scoring and standard setting for standardized-patient cases in an examination of clinical competence. *Acad Med*. 1996;71(suppl):S84-S86.
111. Norcini JJ, Stillman PL, Sutnick AI, et al. Scoring and standard setting with standardized patients. *Eval Health Prof*. 1993;16:322-332.
112. Martin JA, Reznick RK, Rothman A, Tamblyn RM, Regehr G. Who should rate candidates in an objective structured clinical examination? *Acad Med*. 1996; 71:170-175.
113. Blane CE, Calhoun JG. Objectively evaluating student case presentations. *Invest Radiol*. 1985;20:121-123.
114. Stillman P, Swanson D, Regan MB, et al. Assessment of clinical skills of residents utilizing stan-

dardized patients: a follow-up study and recommendations for application. *Ann Intern Med*. 1991;114:393-401.
115. Cohen R, Rothman AI, Poldre P, Ross J. Validity and generalizability of global ratings in an objective structured clinical examination. *Acad Med*. 1991; 66:545-548.
116. Swartz MH, Colliver JA, Bardes CL, Charon R, Fried ED, Moroff S. Global ratings of videotaped performance versus global ratings of actions recorded on checklists: a criterion for performance assessment with standardized patients. *Acad Med*. 1999;74:1028-1032.
117. Mumford E, Schlesinger H, Cuerdon T, Scully J. Ratings of videotaped simulated patient interviews and four other methods of evaluating a psychiatry clerkship. *Am J Psychiatry*. 1987;144:316-322.
118. Tamblyn RM, Abrahamowicz M, Berkson L, et al. First-visit bias in the measurement of clinical competence with standardized patients. *Acad Med*. 1992; 67(suppl):S22-S24.
119. Ferrell BG, Thompson BL. Standardized patients: a long-station clinical examination format. *Med Educ*. 1993;27:376-381.
120. Singer PA, Robb A, Cohen R, Norman G, Turnbull J. Performance-based assessment of clinical ethics using an objective structured clinical examination. *Acad Med*. 1996;71:495-498.
121. McLeod PJ, Tamblyn R, Benaroya S, Snell L. Faculty ratings of resident humanism predict patient satisfaction ratings in ambulatory medical clinics. *J Gen Intern Med*. 1994;9:321-326.
122. Irby D. Teaching and learning in ambulatory care settings: a thematic review of the literature. *Acad Med*. 1995;70:898-931.
123. Kalet A, Earp JA, Kowlowitz V. How well do faculty evaluate the interviewing skills of medical students? *J Gen Intern Med*. 1992;7:499-505.
124. Rowland-Morin PA, Burchard KW, Garb JL, Coe NP. Influence of effective communication by surgery students on their oral examination scores. *Acad Med*. 1991;66:169-171.
125. Wang-Cheng RM, Fulkerson PK, Barnas GP, Lawrence SL. Effect of student and preceptor gender on clinical grades in an ambulatory care clerkship. *Acad Med*. 1995;70:324-326.
126. Herbers JE Jr, Noel GL, Cooper GS, Harvey J, Pangaro LN, Weaver MJ. How accurate are faculty evaluations of clinical competence? *J Gen Intern Med*. 1989; 4:202-208.
127. Norcini JJ, Blank LL, Arnold GK, Kimball HR. The mini-CEX (clinical evaluation exercise): a preliminary investigation. *Ann Intern Med*. 1995;123:795-799.
128. Dunnington GL, Wright K, Hoffman K. A pilot experience with competency-based clinical skills assessment in a surgical clerkship. *Am J Surg*. 1994;167: 604-606.
129. Dunnington G, Reisner L, Witzke D, Fulginiti J. Structured single-observer methods of evaluation for the assessment of ward performance on the surgical clerkship. *Am J Surg*. 1990;159:423-426.
130. Keynan A, Friedman M, Benbassat J. Reliability of global rating scales in the assessment of clinical competence of medical students. *Med Educ*. 1987;21: 477-481.
131. 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.
132. McKinley RK, Fraser RC, van der Vleuten C, Hastings AM. Formative assessment of the consultation performance of medical students in the setting of general practice using a modified version of the Leicester Assessment Package. *Med Educ*. 2000;34:573-579.
133. Yang JC, Laube DW. Improvement of reliability of an oral examination by a structured evaluation instrument. *J Med Educ*. 1983;58:864-872.
134. Stillman RM, Lane KM, Beeth S, Jaffe BM. Evaluation of the student: improving validity of the oral examination. *Surgery*. 1983;93:439-442.
135. Lockie C. *The Examination for Membership in the Royal College of General Practitioners*. London, England: Royal College of General Practitioners; 1990. Occasional Paper 46.
136. Ram P, van der Vleuten C, Rethans JJ, Grol R, Aretz K. Assessment of practicing family physicians: comparison of observation in a multiple-station examination using standardized patients with observation of consultations in daily practice. *Acad Med*. 1999; 74:62-69.
137. Ram P, van der Vleuten C, Rethans JJ, Schouten B, Hobma S, Grol R. Assessment in general practice: the predictive value of written-knowledge tests and a multiple-station examination for actual medical performance in daily practice. *Med Educ*. 1999;33:197-203.
138. Ram P, Grol R, Rethans JJ, Schouten B, van der Vleuten C, Kester A. Assessment of general practitioners by video observation of communicative and medical performance in daily practice: issues of validity, reliability and feasibility. *Med Educ*. 1999;33: 447-454.
139. Tate P, Foulkes J, Neighbour R, Campion P, Field S. Assessing physicians' interpersonal skills via videotaped encounters: a new approach for the Royal College of General Practitioners Membership examination. *J Health Commun*. 1999;4:143-152.
140. Luck J, Peabody JW, Dresselhaus TR, Lee M, Glassman P. How well does chart abstraction measure quality? a prospective comparison of standardized patients with the medical record. *Am J Med*. 2000; 108:642-649.
141. Fiscella K, Franks P. Influence of patient education on profiles of physician practices. *Ann Intern Med*. 1999;131:745-751.
142. Tornow W, London M. *Maximizing the Value of 360-Degree Feedback: A Process for Successful Individual and Organizational Development*. San Francisco, Calif: Jossey-Bass Inc; 1998.
143. Hundert EM, Hafferty F, Christakis D. Characteristics of the informal curriculum and trainees' ethical choices. *Acad Med*. 1996;71:624-642.
144. Kaiser S, Bauer JJ. Checklist self-evaluation in a standardized patient exercise. *Am J Surg*. 1995;169: 418-420.
145. Gordon MJ. A review of the validity and accuracy of self-assessments in health professions training. *Acad Med*. 1991;66:762-769.
146. Westburg J, Jason H. *Collaborative Clinical Education: The Foundation of Effective Patient Care*. New York, NY: Springer Publishing Co; 1993.
147. Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academy Press; 2001.
148. Gonzi A. *The Development of Competence-Based Assessment Strategies for the Professions*. Canberra: Australian Government Publishing Service; 1993. National Office of Everseas Skills Recognition Research Paper # 8.
149. Wass V, McGibbon D, van der Vleuten C. Composite undergraduate clinical examinations: how should the components be combined to maximize reliability? *Med Educ*. 2001;35:326-330.
150. van der Vleuten C CP, Norman GR, De Graaff E. Pitfalls in the pursuit of objectivity: issues of reliability. *Med Educ*. 1991;25:110-118.
151. Hodges B. *Assessment Across Basic Science and Clinical Skills: Using OSCEs With Postencounter Basic Science Probes*. 2001. Manual located at the University of Toronto, Toronto, Ontario.
152. Smith RM. The triple-jump examination as an assessment tool in the problem-based medical curriculum at the University of Hawaii. *Acad Med*. 1993; 68:366-372.
153. Brailovsky C, Charlin B, Beausoleil S, Cote S, van der Vleuten C. Measurement of clinical reflective capacity early in training as a predictor of clinical reasoning performance at the end of residency: an experimental study on the script concordance test. *Med Educ*. 2001;35:430-436.
154. Verhoeven BH, Hamers JG, Scherpbier AJ, Hoogenboom RJ, van der Vleuten C. The effect on reliability of adding a separate written assessment component to an objective structured clinical examination. *Med Educ*. 2000;34:525-529.
155. Charlin B, Roy L, Brailovsky C, Goulet F, van der Vleuten C. The Script Concordance test: a tool to assess the reflective clinician. *Teaching Learning Med*. 2000;12:189-195.
156. Risser DT, Simon R, Rice MM, Salisbury ML. A structured teamwork system to reduce clinical errors. In: Spath PL, ed. *Error Reduction in Health Care: A Systems Approach to Improving Patient Safety*. New York, NY: John Wiley & Sons; 1999:235-277.
157. Suchman AL, Botelho RJ, Hinton-Walker P. *Partnerships in Healthcare: Transforming Relational Process*. Rochester, NY: University of Rochester Press; 1998.
158. American Board of Internal Medicine. *Patient and Physician Peer Assessment and the ABIM Program of Continuous Professional Development*. Philadelphia, Pa: American Board of Internal Medicine; 2001.
159. Dauphinee WD. Assessing clinical performance. Where do we stand and what might we expect? *JAMA*. 1995;274:741-743.
160. Violato C, Marini A, Toews J, Lockyer J, Fidler H. Feasibility and psychometric properties of using peers, consulting physicians, co-workers, and patients to assess physicians. *Acad Med*. 1997;72 (suppl 1):S82-S84.
161. Roter DL, Hall JA, Kern DE, Barker LR, Cole KA, Roca RP. Improving physicians' interviewing skills and reducing patients' emotional distress—a randomized clinical trial. *Arch Intern Med*. 1995;155:1877-1884.
162. Beckman HB, Markakis KM, Suchman AL, Frankel RM. The doctor-patient relationship and malpractice: lessons from plaintiff depositions. *Arch Intern Med*. 1994;154:1365-1370.
163. Tamblyn R, Benaroya S, Snell L, McLeod P, Schnarch B, Abrahamowicz M. The feasibility and value of using patient satisfaction ratings to evaluate internal medicine residents. *J Gen Intern Med*. 1994;9: 146-152.
164. Helmreich RL. On error management: lessons from aviation. *BMJ*. 2000;320:781-785.
165. Helmreich RL, Davies JM. Anaesthetic simulation and lessons to be learned from aviation. *Can J Anaesth*. 1997;44:907-912.
166. Schaefer HG, Helmreich RL. The importance of human factors in the operating room. *Anesthesiology*. 1994;80:479.
167. Davies JM, Helmreich RL. Virtual reality in medical training. *CMAJ*. 1997;157:1352-1353.
168. Helmreich RL, Musson DM. Surgery as team endeavour. *Can J Anaesth*. 2000;47:391-392.
169. Sexton JB, Thomas EJ, Helmreich RL. Error, stress, and teamwork in medicine and aviation: cross sectional surveys. *BMJ*. 2000;320:745-749.
170. Epstein RM. Communication between primary care physicians and consultants. In: Suchman AL, Botelho RJ, Hinton-Walker P, eds. *Partnerships in Health Care: Transforming Relational Process*. Rochester, NY: University of Rochester Press; 1998:171-183.
171. Frankford DM, Konrad TR. Responsive medical professionalism: integrating education, practice, and community in a market-driven era. *Acad Med*. 1998; 73:138-145.