Defining and Assessing Professional Competence

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

MEDICAL SCHOOLS, POST-GRADUATE 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. HOWEVER, THERE IS NO AGREEd-UPON DEFINITION OF COMPETENCE THAT ENCOMPASSSA 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: 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.

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.
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.

**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. But Polanyi 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, as are other clinical skills.

Personal knowledge is usable knowledge gained through experience. 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, 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; “when we see the whole, we see its parts differently than when we see them in isolation.” 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. Schon 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.

©2002 American Medical Association. All rights reserved.
Compliance with the new standards requires public reporting of errors that result in serious harm to patients. This creates the potential for an expert’s judgment to be questioned, particularly if the expert is not a general practitioner. The report must reflect a reasonable level of judgment. The expert’s judgment may be questioned if the expert does not have the required knowledge or skill. In such cases, the expert’s judgment may be questioned if the expert does not have the required knowledge or skill.

### Habits of Mind


### Context

Context is context-dependent. Competence is a statement of relationship between an ability (in the person), a task (in the world), and the ecology of the health systems and clinical contexts in which those tasks occur. In this view, 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. Key measurable patient-centered behaviors include responding to patients’ emotions and participatory decision making. Medical errors are often due to the failure of health systems rather than individual deficiencies. Thus, the assessment of teamwork and institutional self-assessment might effectively complement individual assessments.

### 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, 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.\textsuperscript{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.\textsuperscript{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\textsuperscript{62} and has not done enough to reduce medical errors.\textsuperscript{32,63}

Within each domain of assessment, there are 4 levels at which a trainee might be assessed (Figure).\textsuperscript{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.\textsuperscript{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**

<table>
<thead>
<tr>
<th>LEVEL OF ASSESSMENT</th>
<th>CONTEXT OF CARE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knows</td>
<td>Chronic illness</td>
</tr>
<tr>
<td>Knows how</td>
<td>Emergency</td>
</tr>
<tr>
<td>Shows how</td>
<td>Preventive</td>
</tr>
<tr>
<td>Does</td>
<td>Acute hospital</td>
</tr>
</tbody>
</table>

- 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)
- 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, and standardized patient assessments of physical examination and technical and communication skills. 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 that predict clinical outcomes in real practice. Few reliably assess clinical reasoning, systems-based care, technology, and the patient-physician relationship. 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 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 and multiple-choice certification examination scores correlated with subsequent faculty and peer ratings. Many have questioned the validity of multiple-choice examinations, though. For example, compared with Florida family physicians who are not board-certified, those who are have nearly twice the risk of being sued. Standardized test scores have been inversely correlated with empathy, responsibility, and tolerance. 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 and simulations are convincing; the detection rate of unannounced SPs in community practice is less than 10%. Communication, physical examination, counseling, and technical skills can be rated reliably if there is a sufficiently large number of SP cases and if criteria for competence are based on evidence. 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. Although SPs’ ratings usually correlate with those of real patients, differences have been noted.

Defining pass/fail criteria for OSCEs has been complex. There is debate about who should rate student performance in an OSCE. Ratings by the SP are generally accurate 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 examiners and with the reputation of the training program. But global rating scales of interpersonal skills may be more valid than behavioral checklists. The OSCE scores may not correlate with multiple-choice examinations and academic grades suggesting that these tools measure different skills. Clinicians may behave differently in examination settings than in real practice and short OSCE stations can risk fragmentation and trivialization of isolated elements of what should be a coherent whole. The OSCE also has low test reliability for measuring clinical ethics.

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. However, evaluators often do not observe trainees directly. They often have different standards and are subject to halo effects and racial and sex bias. 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 and structured oral examination formats have been developed in response to this criticism.

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. Although the face validity of such a measure is high and the format is well accepted by physicians, the number of cases that should be presented to achieve adequate reliability is unclear. Profiling by managed-care databases is increasingly used as an evaluation measure of clinical competence. However, data abstraction is complex 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 exercises144 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.

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 remediation146 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 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
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 questions149 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 exercises132—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 thinking133 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.130,134 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 unlabeled 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 outpatients 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 concordance135 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 intelligence60 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.24,25 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 settings164-169; 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 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-Carrió, MD, Daniel Federman, MD, Brian Hodges, MD, Daniel Klass, MD, Larry Mauksch, MD, and Kevin Volkman, PhD, for their critical review of the manuscript. Also we would like to acknowledge Anthony LaDuca, PhD, and Albert Orlit-Bosch, MD, for their contributions to our formulation of professional competence.

REFERENCES


50. LaDuca A. Validation of professional licensure ex-
DEFINING AND ASSESSING PROFESSIONAL COMPETENCE

Downloaded From: http://jama.jamanetwork.com/ by a University of Cincinnati Libraries User  on 02/19/2015
DEFINING AND ASSESSING PROFESSIONAL COMPETENCE


Downloaded From: http://jama.jamanetwork.com/ by a University of Cincinnati Libraries User on 02/19/2015