

The role of assessment in competency-based medical education

ERIC S. HOLMBOE¹, JONATHAN SHERBINO², DONLIN M. LONG³, SUSAN R. SWING⁴ & JASON R. FRANK⁵, FOR THE INTERNATIONAL CBME COLLABORATORS

¹American Board of Internal Medicine, USA, ²McMaster University, Hamilton, Canada, ³Johns Hopkins University, Baltimore, USA, ⁴Accreditation Council for Graduate Medical Education, USA, ⁵Royal College of Physicians and Surgeons of Canada and University of Ottawa, Canada

Abstract

Competency-based medical education (CBME), by definition, necessitates a robust and multifaceted assessment system. Assessment and the judgments or evaluations that arise from it are important at the level of the trainee, the program, and the public. When designing an assessment system for CBME, medical education leaders must attend to the context of the multiple settings where clinical training occurs. CBME further requires assessment processes that are more continuous and frequent, criterion-based, developmental, work-based where possible, use assessment methods and tools that meet minimum requirements for quality, use both quantitative and qualitative measures and methods, and involve the wisdom of group process in making judgments about trainee progress. Like all changes in medical education, CBME is a work in progress. Given the importance of assessment and evaluation for CBME, the medical education community will need more collaborative research to address several major challenges in assessment, including “best practices” in the context of systems and institutional culture and how to best to train faculty to be better evaluators. Finally, we must remember that expertise, not competence, is the ultimate goal. CBME does not end with graduation from a training program, but should represent a career that includes ongoing assessment.

Introduction

Competency-based medical education (CBME), by definition, necessitates a robust and multifaceted assessment system (Norcini et al. 2008). Assessment and the judgments or evaluations that arise from it are important at the level of the trainee, the program, and the public. For trainees, CBME requires enhanced attention to formative assessment to ensure they receive frequent and high-quality feedback to guide their development and the acquisition of the necessary competencies (Carraccio et al. 2002; Bing-You & Trowbridge 2009). For those trainees with deficiencies in certain knowledge areas, skills, or attitudes, CBME can provide an “early warning system” to guide remedial action; for the few trainees who do not and will not ever possess the minimum level of competence required for medical practice, early identification will facilitate an earlier and fair exit from medical education. On the other end of the spectrum, more advanced trainees can receive frequent, formative assessment that allows their training to be focused more effectively, thus potentially facilitating their more rapid advancement – to the ultimate benefit of patients, society, and the trainees themselves.

At the program level, effective assessment provides the information and judgment necessary to enable program-level decisions about trainee advancement to be made reliably and fairly (Hawkins & Holmboe 2008). Effective assessment also potentially reduces dependence on educational “dwell time” as a proxy for competence – a characteristic that describes

Practice points

- A competency-based approach to medical education relies on continuous, comprehensive, and elaborate assessment and feedback systems.
- Ideally, a major portion of the assessments should be performed in the context of the clinical workplace and should be criterion-referenced.
- Assessment facilitates the developmental progression of competence.
- A number of useful assessment methods already exist; work should focus on helping training programs use such methods more effectively.
- New assessment tools and approaches will need to be developed for “new” competencies such as teamwork, systems, and quality improvement, among others, to fully realize the promise of CBME.

most current medical education programs (Carraccio et al. 2002). The aggregation of assessment information across trainees provides valuable feedback on the training program’s curriculum as part of continuous quality improvement. In the United States, aggregated measurement of competence in training programs has been proposed as a way to allow the accreditation system to evolve in a manner that places more emphasis on the attainment of educational outcomes and less

Correspondence: Eric S. Holmboe, MD, American Board of Internal Medicine, 510 Walnut Street, Suite 1700, Philadelphia, PA 19106. Tel: 215-446-3606; fax: 215-446-3633; email: eholmboe@abim.org

on process and structure; such a system would thus be focused on continuous quality improvement (Goroll et al. 2004; Nasca 2008). This is not to say that program evaluation should be based solely on aggregate assessment data, but simply that without using aggregate information as part of the “synthesis” in judging programs, it is hard to imagine how we can fulfill any of the promises of outcomes-based education.

Finally, robust, accurate assessment is essential to professional self-regulation, a privilege granted to medical education but increasingly viewed with skepticism and cynicism worldwide. For example, the governments in Australia, Canada, and the United Kingdom have become more directly involved in the regulation of medical education, and similar conversations are beginning to occur in the United States (Chantler & Ashton 2009; Shaw et al. 2009; Medicare Payment Advisory Commission 2009). Training a physician is a very expensive enterprise for which, in almost every country, substantial financial support is provided from the public purse. CBME therefore provides an opportunity to regain public trust by using precious resources more wisely and efficiently, ensuring that all trainees attain high standards of knowledge, skills, and attitudes in the key competencies expected of them.

The setting of training and assessment

After the early years of medical school, most education occurs through the care of real patients in clinical settings. Although various forms of simulation are becoming standard (Issenberg et al. 2005; Cleland et al. 2009), learning and assessment will occur predominantly in the clinical workspace for the foreseeable future. This requires that any evaluation system incorporate a robust and effective work-based assessment program. Traditional approaches to measurement, based in the psychometric imperative, have been leery of work-based assessment, given the biases inherent in the clinical setting and the challenges of “adjusting” for contextual factors that make it difficult to determine the “true” score, or rating, of competence (Rethans et al. 2002; Williams et al. 2003; Govaerts et al. 2007).

The implications of these considerations for CBME is that this approach to medical education must account for and incorporate contextual factors arising from the clinical setting into assessment processes.

Clinical microsystems

The predominant clinical units where trainees work and learn – for example, ambulatory clinics, hospital wards, surgical suites, and intensive care units – are microsystems. As defined by Nelson and colleagues, a clinical microsystem is “a small group of people who work together on a regular basis to provide care to discrete subpopulations of patients. It has clinical and business aims, linked processes, and a shared information environment, and produces performance outcomes” (Nelson et al. 2007).

Microsystems provide the context for work-based training and assessment. Although it follows logically that a prerequisite for CBME would be that trainees work and learn in functional microsystems to enhance the attainment of competency, the assessment system is also inevitably embedded in the microsystems of the training program, making it important for educators to carefully consider how the culture and functionality of these multiple microsystems affect assessment processes (Rethans et al. 2002). Unfortunately, there is substantial evidence that trainees too often learn and work in dysfunctional microsystems. This reality may be a major impediment to CBME in general and to assessment in particular (Bowen et al. 2005; Reddy et al. in press; Hafferty & Levinson 2008). However, beyond the contextual issues, our understanding of clinical microsystems can help to inform our decisions about how an effective assessment system for CBME should move forward.

Necessary components of an effective assessment system

Assessment should be viewed in the context of a complex adaptive system (McDaniel & Driebe 2001; Nelson et al. 2007). Complex adaptive systems share several important characteristics. First, they consist of multiple interconnected elements, including individuals who have the capacity to learn from one another, to adapt, and therefore to change (Suchman 2006). Assessment systems consist of multiple “agents” (e.g., faculty members, peers, patients, and other non-physician health care providers) using multiple assessment methods and tools (e.g., exams, mini-CEX, audit, multi-source feedback, simulation, etc.) in collaboration with the trainee in a competency-based training model. Understanding these interactions and how they adapt and change is crucial to creating, maintaining, and constantly improving a CBME assessment system. Table 1 correlates the nine factors identified for successful clinical

Table 1. Microsystem success factors and assessment system correlates.

Microsystem success characteristic	Assessment system correlates
Information and information technology	Portfolio, preferably electronic
Leadership of microsystem	Clerkship and program directors
Macrosystem support of microsystem	Support and resources from department chair and institution
Patient focus	Appropriate clinical experiences; measuring patient experience
Staff focus	Faculty development in assessment; involvement of non-physicians in assessment
Interdependence of care team	Working in interdisciplinary teams; teamwork competence
Process improvement	Continuous quality improvement of assessment methods and training tools
Education and training	Competency-based; developmental clinical experiences; milestones and benchmarks
Performance results	Outcomes of training; at minimum, competence needed to advance to next stage

microsystems with what these factors might look like for a successful assessment system. With this framework in mind, we will now explore six key features and components of effective assessment in CBME.

1. Assessment needs to be more continuous and frequent

As Carraccio and colleagues have outlined (2002), a competency-based education program emphasizes formative over summative assessment. This is not to say that summative assessment is unimportant; indeed, the medical education community has a professional obligation to the public to ensure that its trainees are ultimately competent for unsupervised practice. A greater emphasis on formative assessment, while supported by educational theory (McCowan 1998; Hodge 2007), is also consistent with work on the development of expertise through “deliberate practice” (Ericsson 2006, 2007). The deliberate practice concept highlights the need for effective coaching, mentoring, and feedback. Feedback is only as good as the assessment that informs it: inaccurate assessment leads to ineffective feedback and potentially delayed development. However, effective feedback can be a powerful tool for professional development (Hattie & Timperley 2007).

As noted by Hattie and Timperley in their extensive review across the continuum of education, feedback may be the most potent “intervention” in helping learners progress (Hattie & Timperley 2007). Feedback in clinical education is a complex process involving specific skills that must be tightly integrated into the assessment system (van der Ridder et al. 2008). We now know that, when performed in isolation, self-assessment is not only ineffective but is potentially dangerous (Davis et al. 2006; Eva & Regehr 2008). Furthermore, feedback is a key component that guides trainees in more meaningful self-directed assessment-seeking behaviour that is critical in a competency-based system (Eva & Regehr 2008). An effective CBME system must continuously link robust assessment with equally robust feedback on a continuous basis.

2. Assessment must be criterion-based, using a developmental perspective

A normative approach to assessment, based on comparable trainees within an institution, makes the attainment of true outcomes very difficult. As a result, standards are too often set below appropriate expectations. A simple example might help to illustrate the problem. The insertion of central venous lines is an important procedure for many trainees to know how to perform, but is associated with potentially serious patient complications. A growing number of residency programs in the United States have mandated simulator training before allowing residents to perform the procedure on patients in hospital. One of these programs compared performance on central line insertion at baseline and after training using a criterion-based approach. At the baseline assessment, essentially all the residents failed to meet the criteria for minimal safety in independently performing central line insertion: in other words, simply getting the line into the right vessel was

not enough. In fact, the baseline performance among the residents was remarkably similar, making the point that a normative approach to assessment in this situation could have led to a mistaken judgment that most members of the group were competent, when in fact everyone was incompetent to insert central lines safely (Barsuk et al. 2009).

Criteria should also to be developmental in nature, where appropriate. Defining the criteria in developmental terms, commonly called milestones or benchmarks, allows programs to determine whether the trainee is on an appropriate “trajectory” (Green et al. 2009). Milestones provide specific guidance on trainee progress throughout the continuum of their training program. For example, the milestone for effective counselling by an intern at 12 months would be the effective use of the basic elements of informed decision-making for uncomplicated issues (e.g., starting a medication with known risks), but by 24 months he or she must be able to engage patients and family members in shared decision-making for complicated diagnostic and therapeutic scenarios. Milestones, in effect, become the blueprint for assessment and help to guide the appropriate selection of assessment methods and tools, and can help to create the holistic narratives or “stories” of where trainees should be developmentally (Green et al. 2009).

3. Competency-based medical education, with its emphasis on preparation for what the trainee will ultimately do, requires robust work-based assessment

Simulation and other non-work-based forms of assessment will continue to grow in importance, as they should. Simulation, in particular, provides a venue for deliberate practice, including immediate assessment and feedback during the early stages of learning, while protecting patients from potential harm (Issenberg et al. 2005). Nonetheless, assessment must also be based on “authentic” encounters and frequent direct observation (Carraccio et al. 2002; Williams et al. 2003; Govaerts et al. 2007). Although some have noted the lack of strong evidence that work-based assessments are better than more traditional forms (Norcini 2003), we believe that work-based assessment is an essential component of CBME, especially given the greater need for formative assessment and feedback.

As a result, a CBME assessment system places more, not fewer, demands on faculty. Faculty work side by side with trainees on a daily basis and are therefore in an excellent position to provide real-time evaluation and feedback. They need to be keen and accurate observers of trainee performance, but despite this central role, we know very little about effective faculty observation skills and behaviours (Williams et al. 2003; Govaerts et al. 2007). Of the limited studies performed to date on observation, all have demonstrated that faculty frequently fail to identify deficiencies in trainees’ clinical skills (Herbers et al. 1989; Kalet et al. 1992; Holmboe 2004). Few studies have described an attempt to improve direct observation skills through faculty development (Noel et al. 1992; Holmboe et al. 2004; Cook et al. 2009); the only one that showed any benefit concerned assessment conducted

in a controlled setting rather than in clinical environments with real patients (Holmboe et al. 2004). One of the major challenges will be how best to train faculty to be more accurate observers and better assessors of performance, especially with respect to the complex interactions and contextual factors involved in actual patient care that often cannot be reproduced and measured with simulated patients. In addition, faculty corroboration of trainee findings and judgments through other supervisory activities beyond direct observation are also important inputs into effective assessment (Kennedy et al. 2007).

4. Training programs must use assessment tools that meet minimum standards of quality

The community needs to move away from developing multiple “home-grown” assessment tools and work instead toward the adoption of a core set of assessment tools that will be used across all programs within a country or region. Medical education has suffered from too much variability in the choice and use of assessment tools, akin to the variability seen in the delivery and quality of health care (Fisher et al., 2003). Several frameworks are available to guide the evaluation of the quality of assessment tools. One of these, the utility index (van der Vleuten 1996), is a simple but useful formula:

$$\text{Utility} = \text{validity} \times \text{reliability} \times \text{educational impact} \\ \times \text{acceptability} \times \text{cost effectiveness.}$$

Another framework, recently developed by the ACGME's Advisory Committee on Educational Outcome Assessment, uses a hybrid approach that, combining elements of the utility index and a grading system based on clinical guidelines, produces a “report card” summary for a specific assessment tool (Swing et al. in press). However, a word of caution is in order: we cannot wait for the “perfect” assessment tools but, rather, must use the best combination of tools available for the purpose. It is also important to highlight the fact that being “good enough” does not depend only on whether a tool has satisfactory psychometric characteristics. A number of assessment experts are arguing for a broader conception of measurement that considers constructivist approaches and incorporates, instead of adjusting for, context into the assessment process (Govaerts et al. 2007).

A number of tools have been studied for work-based assessment, but too many have been studied only within single institutions, or have not been sufficiently investigated for validity, reliability, and other attributes. For example, the best-studied assessment tool for direct observation is the mini-CEX; although at least 20 studies of this tool are now in print, we still lack a full understanding of how best to utilize it (Kogan et al. 2009). The primary reason for this state of affairs is the lack of recognition that any work-based assessment tool is only as good as the individual using it (Landy & Farr 1980; Murphy & Cleveland 1995). For CBME to be ultimately successful, we need not only a combination of better assessment tools but also more skilled faculty and other assessors who will use them.

5. We must be willing to incorporate more “qualitative” approaches to assessment

Qualitative approaches to assessment could include written narrative and the synthesis of conversations that occur during evaluation sessions. Research has shown that valuable and defensible information can be obtained during evaluation sessions, especially with respect to difficult competencies such as professionalism (Hemmer et al. 2000; Battistone et al. 2001), and that qualitative methods can be used reliably to judge portfolios (Driessen et al. 2005). In fact, as a synthetic and comprehensive approach to assessment, portfolios require a mixed approach to judging overall competence not only as part of an evaluation system (Holmboe et al., 2006) but also as an important component of continuous professional development and maintenance of certification (Holmboe 2008).

Some have argued that there is too much emphasis on the “objectification” of assessment when judgment can just as effectively be expressed in words instead of numbers (Govaerts et al. 2007). For example, the results of a direct observation assessment by faculty could be synthesized into a number on a rating scale, a categorization using words of judgment (e.g., “satisfactory”), or a narrative description (e.g., “the trainee appropriately began the patient interview with an open-ended question and effectively gathered key information for diagnosis”). All three have the capacity to provide a judgment, but the narrative example provides the level of specificity needed by the trainee to make improvements and develop learning plans.

6. Assessment needs to draw upon the wisdom of a group and to involve active engagement by the trainee

No single individual should make judgments about the competence of a trainee in isolation, especially for summative decisions (Swing et al. 2010). Assessment in a CBME system must actively engage the resident in the assessment process. The concept of “self-directed assessment seeking” for practising physicians is an equally important concept for trainees (Eva & Regehr 2008). CBME demands active involvement by the trainee, and programs must empower trainees in assessment. When it is done well, the assessment process of CBME should prepare trainees to maintain, at a minimum, competence over the course of their careers. Multiple studies suggest that too many practising physicians in the past were unsuccessful in this pursuit (Choudhry et al. 2005). Ensuring that all physicians have the skills to seek and perform reliable and valid assessments of their own practice performance is essential to the maintenance of competence (Duffy et al. 2008). It is conceivable that milestones and other descriptive performance criteria developed as a part of CBME will provide helpful guidance for self-assessment during a physician's training and career.

Future concepts for assessment

Future assessment approaches will need to focus more on the *interactions* involved in competence and clinical practice than

simply on the *tasks* of being a physician. This is especially true with regard to the interaction between trainees and the microsystems where they train, which includes the interactions trainees have with all individuals working in the microsystem. Until recently, we have viewed the “system” mainly as context for assessment. However, we are now beginning to recognize that physicians need specific knowledge, skills, and attitudes to work successfully within microsystems. Examples of such microsystem competencies include working effectively as a member of a team and effective interprofessional interactions with non-physician health care providers. The dividing line between systems as providing a “context” for competency and as a specific facet of “competency” is increasingly blurred, for competency is not only demonstrated within the specific context of a system but also pertains to engagement with the system itself. In other words, one element of competency is how effectively a trainee or physician interacts with the system, either to get a task or process done well, or to change the system in order to improve a clinical care process.

Trainees will need these skills when they become responsible for working and leading microsystems of their own. For CBME, an essential philosophical question for assessment will be what and how much “adjustment” for the microsystem should be part of the assessment of trainees, and what aspect of trainees’ interactions with their microsystems is itself a competency.

Finally, evaluation – the judgment aspect of the assessment system – must be integrative and synthetic. One of the major criticisms that has been made of CBME is that it has a propensity to reduce learning and assessment to a series of “checkboxes” (Leung 2002; Talbot 2004). However, we do not believe that the philosophy and theories underpinning CBME are at all inconsistent with integrative and interactional assessment. The power of robust, multifaceted assessment facilitates a process that can synthesize the results of longitudinal and developmental assessment into a more comprehensive, holistic evaluation that is more than the sum of its parts. Human judgment, whether applied in the construction of multiple-choice questions or the use of standardized patients and simulation, or as applied in direct observations by faculty, patients, and other health care providers, will be part of the assessment process for the foreseeable future. The challenge for the work-based assessment aspects of the CBME system is to maximize the quality of human observation and judgment.

Research agenda

Like all changes in medical education, CBME is a work in progress. Given the importance of assessment and evaluation for CBME, the medical education community will need to embark on more collaborative research to identify best practices of assessment in the context of systems and institutional cultures. This research will require consideration of new strategies that incorporate the best of both quantitative and qualitative methods and deliberately include context as a key component of the research. “Realistic evaluation” and the (UK) Medical Research Council’s strategy on studying complex interventions are two models that warrant attention for medical

education research (Pawson & Tilley 1997; Campbell et al. 2007).

The other urgent area of need is to determine how to train faculty to be better evaluators. Despite the central role of faculty in both teaching and assessment, we still know soberingly little about how faculty members conduct their assessments and how best to improve their evaluation skills. Although assessments by others such as patients, peers, and other health care providers are also critical, and despite the fact that simulation has much to offer, faculty cannot and should not be removed from the process. At a minimum, it is part of a faculty member’s professional responsibility to perform evaluations.

Finally, our assessment frameworks need to account for expertise. As we study approaches to assessment within the competency-based model of medical education, we must remember that CBME does not seek competence as an ultimate *state*, but rather recognize that expertise is the end *goal*. CBME does not end with graduation from a training program, but should be integral to a career that includes ongoing assessment.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

Notes on contributors

ERIC S. HOLMBOE, MD, is Senior Vice President and Chief Medical Officer at the American Board of Internal Medicine and American Board of Internal Medicine Foundation. He is also Professor Adjunct of Medicine at Yale University, and Adjunct Professor at the Uniformed Services University of the Health Sciences.

JONATHAN SHERBINO, MD, MEd, FRCPC, is the Director of Continuing Professional Education, Division of Emergency Medicine, McMaster University, Hamilton, Ont., and a Clinician Educator with the Royal College of Physicians and Surgeons of Canada.

DONLIN M. LONG, MD, PhD, is Distinguished Service Professor, The Johns Hopkins University School of Medicine, Baltimore, Md., USA.

SUSAN R. SWING, PhD, is Vice President of Outcome Assessment at the Accreditation Council for Graduate Medical Education. She is a co-developer of the ACGME/ABMS Toolbox of Assessment Methods and is working on collaborative projects to evaluate the quality of assessment tools and develop performance milestones for residents.

JASON R. FRANK, MD, MA(Ed), FRCPC, is the Associate Director of the Office of Education, Royal College of Physicians and Surgeons of Canada, and the Director of Education in the Department of Emergency Medicine, University of Ottawa.

References

- Barsuk JH, McGaghie WC, Cohen ER, Balachandran JS, Wayne DB. 2009. Use of simulation-based mastery learning to improve the quality of central venous catheter placement in a medical intensive care unit. *J Hosp Med* 4(7):397–403.
- Battistone MJ, Pendleton B, Milne C, Battistone ML, Sande MA, Hemmer PA, Shomaker TS. 2001. Global descriptive evaluations are more responsive than global numeric ratings in detecting students’ progress during the inpatient portion of an internal medicine clerkship. *Acad Med* 76(10):S105–S107.
- Bing-You RG, Trowbridge RL. 2009. Why medical educators may be failing at feedback. *JAMA* 302(12):1330–1331.

- Bowen JL, Salerno SM, Chamberlain JK, Eckstrom E, Chen HL, Brandenburg S. 2005. Changing habits of practice. Transforming internal medicine residency education in ambulatory settings. *J Gen Intern Med* 20(12):1181–1187.
- Campbell NC, Murray E, Darbyshire J, Emery J, Farmer A, Griffiths F, Guthrie B, Lester H, Wilson P, Kinmouth AL. 2007. Designing and evaluating complex interventions to improve health care. *BMJ* 334(7591):455–459.
- Carraccio C, Wolfsthal SD, Englander R, Ferentz K, Martin C. 2002. Shifting paradigms: From Flexner to competencies. *Acad Med* 77(5):361–367.
- Chantler C, Ashton R. 2009. The purpose and limits to professional self-regulation. *JAMA* 302(18):2032–2033.
- Choudhry NK, Fletcher RH, Soumerai SB. 2005. Systematic review: The relationship between clinical experience and quality of health care. *Ann Intern Med* 142(4):260–273.
- Cleland JA, Abe K, Rethans JJ. 2009. AMEE Guide No. 42: The use of simulated patients in medical education. *Med Teach* 31(6):477–486.
- Cook DA, Dupras DM, Beckman TJ, Thomas KG, Pankratz VS. 2009. Effect of rater training on reliability and accuracy of mini-CEX scores: A randomized, controlled trial. *J Gen Intern Med* 24(1):74–79.
- Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. 2006. Accuracy of physician self-assessment compared with observed measures of competence: A systematic review. *JAMA* 296(9):1094–1102.
- Drissen E, van der Vleuten C, Schuwirth L, van Tartwijk C, Vermunt J. 2005. The use of qualitative research criteria for portfolio assessment as an alternative to reliability evaluation: A case study. *Med Educ* 39(2):214–220.
- Duffy FD, Lynn LA, Didura H, Hess B, Caverzagie K, Grosso L, Lipner RA, Holmboe ES. 2008. Self assessment of practice performance: Development of the ABIM Practice Improvement Module (PIM). *J Contin Ed Health Prof* 28(1):38–46.
- Ericsson KA. 2006. The influence of expertise and deliberate practice on the development of superior expert performance. In: Ericsson KA, Charness N, Hoffman RR, Feltovich PJ, editors. *Cambridge Handbook of Expertise and Expert Performance*. Cambridge, UK: Cambridge University Press. pp 683–704.
- Ericsson KA. 2007. An expert-performance perspective of research on medical expertise: The study of clinical performance. *Med Educ* 41(12):1124–1130.
- Eva KW, Regehr G. 2008. “I’ll never play professional football” and other fallacies of self-assessment. *J Contin Educ Health Prof* 28(1):14–19.
- Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ, Lucas FL, Pinder EL. 2003. The implications of regional variations in Medicare spending. Part 1: The content, quality, and accessibility of care. *Ann Intern Med* 138(4):273–287.
- Goroll AH, Sirio C, Duffy FD, Le Blond RF, Alguire P, Blackwell TA, Rodak WE, Nasca T, Residency Review Committee for Internal Medicine. 2004. A new model for accreditation of residency programs in internal medicine. *Ann Intern Med* 140(11):902–909.
- Govaerts MJ, van der Vleuten CP, Schuwirth LW, Muijtens AM. 2007. Broadening perspectives on clinical performance assessment: Rethinking the nature of in-training assessment. *Adv Health Sc Educ Theory Pract* 12(2):239–260.
- Green ML, Aagaard EM, Caverzagie KJ, Chick DA, Holmboe E, Kane G, Smith CD, Iobst W. 2009. Charting the road to competence: Developmental milestones for internal medicine residency training. *J Grad Med Educ* 1:5–20.
- Hawkins RE, Holmboe ES. 2008. Constructing an evaluation system for an educational program. In: Holmboe ES, Hawkins RE, editors. *Practical Guide to the Evaluation of Clinical Competence*. Philadelphia, PA: Mosby/Elsevier. pp 216–238.
- Hafferty FW, Levinson D. 2008. Moving beyond nostalgia and motives: Towards a complexity science view of medical professionalism. *Perspect Biol Med* 51(4):599–615.
- Hattie J, Timperley H. 2007. The power of feedback. *Rev Educ Res* 77(1):81–112.
- Herbers JE Jr, Noel GL, Cooper GS, Harvey J, Pangaro LN, Weaver MJ. 1989. How accurate are faculty evaluations of clinical competence? *J Gen Intern Med* 4(3):202–208.
- Hemmer PA, Hawkins R, Jackson JL, Pangaro LN. 2000. Assessing how well three evaluation methods detect deficiencies in medical students’ professionalism in two settings of an internal medicine clerkship. *Acad Med* 75(2):167–173.
- Hodge S. 2007. The origins of competency-based training. *Aust J Adult Learn* 47(2):179–209.
- Holmboe ES. 2004. Faculty and the observation of trainees’ clinical skills: Problems and opportunities. *Acad Med* 79(1):16–22.
- Holmboe ES. 2008. Assessment of the practicing physician: Challenges and opportunities. *J Contin Educ Health Prof* 28(Suppl 1):S4–S10.
- Holmboe ES, Hawkins RE, Huot SJ. 2004. Effects of training in direct observation of medical residents’ clinical competence: A randomized trial. *Ann Intern Med* 140(11):874–881.
- Holmboe ES, Rodak W, Mills G, McFarlane MJ, Schultz HJ. 2006. Outcomes-based evaluation in resident education: Creating systems and structured portfolios. *Am J Med* 119(8):708–714.
- Issenberg SB, McGaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ. 2005. Features and uses of high-fidelity medical simulations that lead to effective learning: A BEME systematic review. *Med Teach* 27(1):10–28.
- Kalet A, Earp JA, Kowlowitz V. 1992. How well do faculty evaluate the interviewing skills of medical students? *J Gen Intern Med* 7(5):499–505.
- Kennedy TJ, Lingard L, Baker GR, Kitchen L, Regehr G. 2007. Clinical oversight: Conceptualizing the relationship between supervision and safety. *J Gen Intern Med* 22(8):1080–1085.
- Kogan JR, Holmboe ES, Hauer KE. 2009. Tools for direct observation and assessment of clinical skills of medical trainees: A systematic review. *JAMA* 302(12):1316–1326.
- Landy FJ, Farr JL. 1980. Performance rating. *Psychol Bull* 87(1):72–107.
- Leung WC. 2002. Competency based medical training: Review. *BMJ* 325(7366):693–696.
- McCowan RJ. 1998. Origins of competency-based training. Buffalo (NY): Center for Development of Human Services. <http://www.bsc-cdhs.org>
- McDaniel RR, Driebe DJ. 2001. Complexity science and healthcare management. In: Blair JD, Fottler MD, Savage GT, editors. *Advances in Health Care Management*, Vol. 2. Stamford, CN: JAI Press. pp 11–36.
- Medicare Payment Advisory Commission (MedPAC). 2009. *Medical Education in the United States: Upporting long-term delivery system reforms*. In: *Report to the Congress: Improving Incentives in the Medicare Program*. June, 2009; pp 3–38. Washington (DC): MedPAC. http://www.medpac.gov/documents/Jun09_EntireReport.pdf
- Murphy KR, Cleveland JN. 1995. *Understanding performance appraisal: Social, organizational, and goal based perspectives*. Thousand Oaks, CA: Sage Publications.
- Nasca TJ. 2008. Where will the “milestones” take us? The next accreditation system. ACGME [Accreditation Council for Graduate Medical Education] Bulletin (September): 3–5. http://www.acgme.org/acWebsite/bulletin/ACG11_BulletinSep08_F.PDF
- Nelson EC, Batalden PB, Godfrey MM. (Eds.) 2007. *Quality by design: A clinical microsystems approach*. San Francisco (CA): Jossey-Bass.
- Noel GL, Herbers JE Jr, Caplow MP, Cooper GS, Pangaro LN, Harvey J. 1992. How well do internal medicine faculty members evaluate the clinical skills of residents? *Ann Intern Med*. 117:757–765.
- Norcini JJ. 2003. Work-based assessment. *BMJ* 326(7392):753–755.
- Norcini JJ, Holmboe ES, Hawkins RE. 2008. Evaluation challenges in the era of outcomes-based education. In: Holmboe ES, Hawkins RE, editors. *Practical Guide to the Evaluation of Clinical Competence*. Philadelphia, PA: Mosby/Elsevier. pp 1–9.
- Pawson R, Tilley N. 1997. *Realistic evaluation*. London, UK: Sage Publications.
- Reddy SG, Babbott SF, Beasley BW, Nadkarni M, Gertner E, Holmboe ES. In press. Prevalence and functionality of electronic health records in Internal Medicine continuity clinics. *Acad Med*.
- Rethans JJ, Norcini JJ, Barón-Maldonado M, Blackmore D, Jolly BC, LaDuca T, Lew S, Page GG, Southgate LH. 2002. The relationship between competence and performance: Implications for assessing practice performance. *Med Educ* 36(10):901–909.
- Shaw K, Cassel CK, Black C, Levinson W. 2009. Shared medical regulation in a time of increasing calls for accountability and transparency: Comparison of recertification in the United States, Canada, and the United Kingdom. *JAMA* 302(18):2008–2014.

- Suchman AL. 2006. A new theoretical foundation for relationship-centered care. Complex responsive processes of relating. *J Gen Intern Med* 21(Suppl 1):S40–S44.
- Swing SR, Clyman SG, Holmboe E, Williams RG. 2010. Advancing resident assessment in graduate medical education. *J Grad Med Educ*. In press.
- Talbot M. 2004. Monkey see, monkey do: A critique of the competency model in graduate medical education. *Med Educ* 38(6):587–592.
- van de Ridder JM, Stokking KM, McGaghie WC, ten Cate OT. 2008. What is feedback in clinical education? *Med Educ* 42(2):189–197.
- van der Vleuten CPM. 1996. The assessment of professional competence: Developments, research and practical implications. *Adv Health Sci Educ* 1(1):41–67.
- Williams RG, Klamen DA, McGaghie WC. 2003. Cognitive, social and environmental sources of bias in clinical performance ratings. *Teach Learn Med* 15(4):270–292.