Evolving concepts of assessment in a competency-based world

Peter Harris, Farhan Bhanji, Maureen Topps, Shelley Ross, Steven Lieberman, Jason R. Frank, Linda Snell, Jonathan Sherbino & on behalf of the ICBME Collaborators

Published online: 09 Jun 2017.

Submit your article to this journal ⧉

View related articles ⧉

View Crossmark data ⧉

MEDICAL TEACHER

Taylor & Francis
Taylor & Francis Group

Evolving concepts of assessment in a competency-based world

Peter Harris[a], Farhan Bhanji[b,c], Maureen Topps[d], Shelley Ross[e] [ID], Steven Lieberman[f], Jason R. Frank[b,g], Linda Snell[b,c] and Jonathan Sherbino[h]; on behalf of the ICBME Collaborators

[a]Office of Medical Education, University of New South Wales, Sydney, Australia; [b]Royal College of Physicians and Surgeons of Canada, Ottawa, Canada; [c]Centre for Medical and Department of General Internal Medicine, McGill University, Montreal, Quebec, Canada; [d]Cumming School of Medicine, University of Calgary, Calgary, Canada; [e]Department of Family Medicine, University of Alberta, Edmonton, Canada; [f]Office of the Dean of medicine, University of Texas Medical Branch, Galveston, TX, USA; [g]Department of Emergency Medicine, University of Ottawa, Ottawa, Canada; [h]Division of Emergency Medicine, Department of Medicine, McMaster University, Hamilton, Canada

ABSTRACT

Competency-based medical education (CBME) is an approach to the design of educational systems or curricula that focuses on graduate abilities or competencies. It has been adopted in many jurisdictions, and in recent years an explosion of publications has examined its implementation and provided a critique of the approach. Assessment in a CBME context is often based on observations or judgments about an individual's level of expertise; it emphasizes frequent, direct observation of performance along with constructive and timely feedback to ensure that learners, including clinicians, have the expertise they need to perform entrusted tasks. This paper explores recent developments since the publication in 2010 of Holmboe and colleagues' description of CBME assessment. Seven themes regarding assessment that arose at the second invitational summit on CBME, held in 2013, are described: competency frameworks, the reconceptualization of validity, qualitative methods, milestones, feedback, assessment processes, and assessment across the medical education continuum. Medical educators interested in CBME, or assessment more generally, should consider the implications for their practice of the review of these emerging concepts.

Introduction

Competency-based medical education (CBME) is an approach to curriculum design and trainee assessment whose fundamental aim is to improve the training of health care professionals so that they deliver consistent, high-quality patient care (Carraccio et al. 2002; Frank et al. 2010). CBME strives to systematically enhance abilities in trainees and practitioners, not only ensuring that competence is achieved and maintained, but also encouraging excellence. Assessment plays a critical role in this process (Holmboe et al. 2010). Competency-based assessment is designed to improve the quantity and quality of feedback to learners, including clinicians, while supporting the practice of reflection and the development of skills for lifelong learning; to utilize assessment data as part of a continuous quality improvement process; and, perhaps most importantly, to ensure that patient care is delivered by providers with demonstrated competence in the relevant domains. Substantial progress has been made over the past five years with the development and dissemination of milestones based on competency frameworks (Carraccio & Englander 2014; Frank et al. 2014; Holmboe et al. 2015), along with the exploration of related concepts such as entrustable professional activities (EPAs) (ten Cate et al. 2016). In this issue, Englander et al. (2017) describe the interrelationships of EPAs, milestones, and competencies. This paper reviews evolving concepts in CBME assessment, including competency frameworks, validity, qualitative methods, milestones, feedback, assessment processes, and the learning continuum.

Practice points

- Competency-based medical education requires frameworks that provide an organized structure for learning, reflect the authentic work of practitioners, and provide opportunities for assessment at the "performance" level.
- Qualitative assessment by collectives of supervisors can provide a rich source of data to inform judgment points and summative decisions about progress.
- Validity relates to the degree to which the "evidence" supports the appropriateness of the interpretations and actions made on the basis of the assessment program. It is not a "number."
- Assessment programs set achievement milestones and then determine which unachieved milestones represent barriers to progression for the individual learner.
- Assessment in continuing professional development recognizes that many competencies are truly achieved only through independent practice.

Competency frameworks and CBME assessment

A competency framework is an organized schema composed of statements of the abilities required for effective professional practice. It is designed to reflect the real work of practitioners, encompassing acceptable local practice and aspirations for future practice. It provides a structure

within which learning activities should relate to one or more of the competencies defined by that framework. Medical education premised on competency frameworks not only engages learners in authentic tasks but also emphasizes the "doing" assessment situated at the pinnacle of Miller's classic description (1990) of clinical competence.

Achieving the goals of CBME requires a robust, programmatic approach to assessment that focuses on workplace-based formative assessment (Holmboe et al. 2010). The approach emphasizes *assessment for learning* (Schuwirth & van der Vleuten 2011) rather than isolated, high stakes, point-in-time, summative examinations of learning. This approach fits well with the Govaerts (2008) description of integrated, context-dependent, demonstrable choices linked to professional domains, and emphasizes the interdependency of tasks and the multiplicity of assessments required for competency. The format of these "assessments" might not reflect traditional examination methods, relying more on tools aligned with clinical practice, and it is essential that learners and clinicians "buy in" to assessment methods that utilize the direct observation and constructive feedback of experts and peers.

In contrast to traditional practice, assessment in CBME presents the challenge of requiring multiple observers of multiple encounters at multiple times with diverse patients (for further details, see Lockyer et al. 2017a in this issue). Summative judgments of competence should not be made in isolation (Holmboe et al. 2010); rather, multiple observations and ratings need to be combined in a manner that conveys a real picture of the trainee to be judged collectively. "Decision moments" need to be disconnected from "assessment moments" (Schuwirth & van der Vleuten 2011), and data collection should continue until it is rich enough to support a summative decision. These data should retain their detail up to the decision point and not be reduced to a series of pass/fail decisions, or even to a series of rankings (Schuwirth & Ash 2013).

Assessment in CBME should focus on improving competence, not simply on identifying incompetence (Schuwirth & Ash 2013). This can be achieved through the practice of *assessment for learning*, in which frequent affirmations of progress are kept separate from higher-stakes summative decision points. There are arguments in favor of continuing the practice of controlled, standardized testing in the "knows," "knows how," and "shows how" domains (Miller 1990), since we can use traditional assessments to support the development of underlying knowledge and skills while still acknowledging that "competence is specific, not generic" (van der Vleuten et al. 2010). To examine Miller's "does" domain, however, assessment has to move to the workplace and incorporate authentic interactions in clinical environments.

Conceptualizing validity in CBME assessment

The validity of an assessment, particularly a summative assessment, is of paramount concern. During the mid-20th century, the adoption of psychometrics in medical education offered a systematic approach to the rigorous interrogation of assessment data in order to determine the accuracy of a judgment. However, it is now understood that traditional representations of validity (i.e. content, criterion, and construct validity) can result in a limited and superficial understanding of the accuracy of a judgment. Specifically, a traditional psychometric approach allows only quantitative data to inform a judgment of competence (Hodges 2013), often focusing on knowledge or skills that are easy to measure, and leading to overly reductionist surrogates of competence (Govaerts et al. 2007). Many educators, because of issues of feasibility, have studied validity only in relation to a single instrument rather than a program of assessment; this can result in a limited conceptualization of validity (Cook et al. 2014).

More than 25 years ago, Messick (1989, p. 5) proposed a definition of validity that moved past the statistical accuracy of quantitative scores: "Validity is an integrated evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment." Building on this conceptualization, Kane (1992) construed validity as a web of inferences about the "truth" of a summative judgment. By this conception, validity is not a "number" but, rather, an argument that supports the final judgment about "true" physician competence.

Essentially, four key inferences are made in building a validity argument: moving from observation to score; moving from a single to a global score; extrapolating to a summative judgment; and linking these data to the constructs that inform the assessment program (Schuwirth & van der Vleuten 2012; Cook et al. 2015). Ultimately, the truth of the attestation that a physician is competent requires more than a statistic demonstrating the correlation between one number and another; it also requires the examination of each of these four key inferences. We need demonstration of a trainee's development in a robust assessment program, based on sufficient data to make a valid summative judgment with an acceptably low measurement error.

Assessment in CBME requires the deliberate design of a program of assessment by a group of educators and the use of assessment tools by teaching faculty in a manner that both supports learning and documents performance. Individual assessors may be required to add to the narrative of a trainee's performance rather than simply recording a mark. These multiple assessment-based judgments will indicate the readiness of the learner to progress in the training program.

Qualitative assessment

By broadening assessment from its traditional focus on the learner's knowledge to his or her performance in more complex professional activities, CBME requires approaches that integrate the assessment of knowledge, skills, behaviors, and attitudes across multiple domains. Many sources of potentially rich qualitative assessment information are available, such as multi-source feedback, workplace-based assessment, reflection, and portfolio assessment, and are likely to become increasingly prominent as competency-based assessment systems evolve (van der Vleuten et al. 2010). Qualitative approaches are especially useful in assessing what van der Vleuten et al. (2010, p. 709) refer to as "domain-independent competencies" such as

professionalism and in providing actionable feedback to guide learners in their development.

To date, the implementation of qualitative assessment has been hampered not only by the time and effort it demands of faculty and often learners – for example, in the compilation and assessment of portfolios (Donato & George 2012) – but also by an emphasis on the traditional psychometric constructs of reliability and validity, under which qualitative data may fare poorly in comparison with scores obtained through standardized testing. The so-called "psychometric discourse" (Hodges 2006) has inhibited the use and refinement of methods of qualitative assessment in medical education, but as the use of these methods has increased the traditional emphases on reliability and validity has been augmented by new constructs such as trustworthiness and authenticity (Govaerts et al. 2007). For example, van der Vleuten et al. (2010) described three strategies to establish trustworthiness in qualitative assessment by linking each to the qualitative research strategies that support them and potential assessment strategies in medical education. Similarly, Driessen et al. (2005) described five qualitative research strategies used to achieve credibility and dependability in portfolio assessment.

The need for multiple types and points of assessment across a broad range of domains is broadly accepted (van der Vleuten & Schuwirth 2005). Integrating these multiple points of assessment – many of which are qualitative in nature – is itself an exercise in qualitative assessment. We can increase the rigor of this exercise by applying methods of structural coherence and peer examination that have already been established in the practice of qualitative research (van der Vleuten et al. 2010).

Milestones and assessment

CBME promotes the development of milestones that can reflect both progression and exit levels of performance. A milestone is a "defined, observable marker of an individual's ability along a developmental continuum" (Englander et al. 2017, in this issue). They are criterion-based and, when taken together, chart the developmental journey of a learner through a particular competency or performance. Milestones that incorporate progressive steps align well with an emphasis on deliberate practice and feedback. They provide a clearly articulated basis for expected, sequential behaviors and encourage a shared mental model for learners and assessors that can guide learning. Promotion to the next level of training or responsibility may depend on the achievement of a well-defined exit milestone, and may be accomplished at different rates by different trainees.

Programmatic approaches to assessment rely on milestones set by the training organization to map progress. They are informative for both programs and trainees, and collectively describe an individual's status at a point in time. They also offer the opportunity for a program to reach its educational goal of "optimising the learning effect for all students" (Schuwirth & van der Vleuten 2014, p. 125), by focusing on individual progress and observing how an individual trainee performs (and, one hopes, improves) on a given task or domain over time. Such repeated personalized data allow a trainee's educational trajectory to be tracked and enable the early detection of substandard performance or the option of acceleration where achievement in an area is demonstrated early. Programs need to decide which unachieved milestones should be barriers to a trainee's promotion and which may be attained later with increased clinical experience.

In addition to program decisions about the achievement of milestones and preparedness to move to the next level of training, trainees require assessment systems to provide them with the tools to develop skill in self-assessment. Current systems of training and assessment have struggled with the challenge of teaching trainees how to assess themselves accurately (Davis et al. 2006) and to carry out reflection-on-practice or reflection-in-practice (Eva & Regehr 2005). CBME approaches, with their focus on evidence of progress, have the potential to develop frameworks and processes for guided self-assessment (Duffy & Holmboe 2006; Galbraith et al. 2008). CBME requires regular communication between teacher and learner to facilitate guided self-assessment; the ultimate goal is to develop enhanced self-reflection skills that will support lifelong learning during independent practice.

Feedback and student outcomes

In comparison with the traditional approach, CBME provides greater understanding and transparency of achievement milestones, and thus improved accountability for both trainees and training organizations. The clearer and more practice-focused competency frameworks used in CBME can also foster improved attention to trainees, given the more frequent use of feedback and assessments. Assessments should identify and encourage desired practice, informing the ongoing development of an educational plan with the goal of reducing the differences between observed and desired practice. Strategies such as the regular review of an individualized learning plan focused on achievements can support this development.

CBME's substantial focus on the use of formative feedback promotes *assessment for learning* (Black & William 1998; Brown 2004). Formative feedback shared with trainees can help them correct behavior that is ineffective or unsafe, and reinforces behaviors that are effective (Bazrafkan et al. 2013). Regular, high-quality formative feedback that is informative, behavior- and task-specific, based on direct observation, and timely provides trainees with essential information by which they direct their behavior (Ramani & Krakov 2012). The sharing of formative feedback creates a safe environment where learners can gain awareness of their strengths and weaknesses (Bazrafkan et al. 2013). An interesting comparison with respect to giving and receiving feedback can be found in the learning culture of music: medical learners regard self-assessment as a skill to develop, while music students recognize the continuing need for external feedback (Watling et al. 2013).

In the CBME approach, faculty roles in assessment emphasize workplace observations, judgments about the progression of expertise, and a renewed emphasis on deliberate feedback. Assessment needs to link to (and be embedded in) clinical workflow, resonating with practising clinicians so that their participation is maximized and sustained. The increased emphasis on workplace assessment in

CBME should create the opportunity for an open and rich dialogue between learners and their assessors to promote development. Faculty must be skilled in conveying this feedback and in creating a positive and safe environment, and learners will need to be willing to accept critical feedback. In most current training programs, the majority of learners progress successfully. The enhanced clarity offered by milestones allows this progress to be recognized but also encourages learners in difficulty to be part of their own "solution." A collaborative approach fosters repeated assessment once an area of dyscompetence (Frank et al. 2010) has been identified and remediated. This allows "borderline" residents not to be pushed through as an effect of dichotomous pass/fail judgments, but to truly achieve each significant milestone. Remediation and enrichment are the two sides of the CBME coin (Schuwirth & van der Vleuten 2011).

Many assessment systems currently focus substantial effort and energy on the borderline or failing trainee with a culture that is not adequately prepared to overcome the "failure to fail" (Dudek et al. 2005). CBME challenges educators to assist (or exclude) those progressing more slowly, while shifting more attention to accelerating and challenging learners who are satisfying the standards. In view of the risk that learners may infer that their learning in a particular area is complete on achieving "competence" and fail to appreciate the need for continual learning throughout their career, assessment in CBME needs to develop a language around excellence and not just around the achievement, or failure to achieve, competence.

Assessment in the authentic environment

Holmboe et al. (2010) describe the clinical microsystems in which learning and assessment occur. Work-based assessment related to the deliberate practice of both individual and sets of competencies requires support by frequent formative assessment and multiple observations. Judgments based on multiple observations by multiple assessors have strong face validity. Recent authors, including Moonen-van Loon et al. (2013), describe compilations that reach adequate levels of defensibility.

Recent work exploring the gap between assessors' observation of performance and interpretation of a rating scale (Crossley & Jolly 2012) noted enhanced agreement when clear clinical anchors were used in an assessment tool. One of the difficulties of assessor training noted by Holmboe et al. (2010) may be reduced by the greater use of recognizable clinical ratings (that is, use the language of clinical assessors) in assessment tools. For the assessors, these multiple observations ought to be shared among the members of a properly constituted competency committee that can function in a "safe place" legally (along the lines of peer-review or morbidity meetings) to allow free discussion about the needs of individual trainees. A flow of information between trainees and a committee of assessors will provide valuable feedback to both and optimize the environment of *assessment for learning*.

There is emerging evidence that the "wisdom of crowds" or a group process in making judgments of competence is important. Since no single method or combination of methods can assess all of the competencies required of a physician, group judgment can produce the best possible decisions about competence. Additionally, the group process may be better able to identify issues concerning professionalism than cumulative single-assessor tools (Hemmer et al. 2000). Such an approach is already familiar to many medical educators through the use of grading committees for the summative assessment of student or resident clinical performance at binary decision points, such as progression or readiness for independent practice. Faculty support to ensure that adequate time is dedicated for completing work-based assessments and participating in groups that "judge" the overall progression of the candidate is critical.

The fragmented learning environment created by a change of rotations every one or two months results in a lack of continuity in assessment. The perpetual cycle of each assessor (or assessor team) starting from scratch with each learner should not be acceptable. In the clinical context, we do not hand over a patient without sharing clinical details essential for care, and we should expect a similar forwarding of information in the handover of learners. An incomplete handover can create unnecessary risk. Although ownership of the developmental trajectory rests with the learner, it also needs to be shared with faculty to determine whether the appropriate milestones or EPAs are being achieved. In feeding performance results forward, we must balance any risk of biasing subsequent assessors against the value of such information to their entrustment decisions.

The learning and assessment continuum

The untapped promise of CBME is its application beyond undergraduate and postgraduate training and into clinical practice. Medical knowledge is expanding faster than ever with the publication of over 75 new randomized controlled trials and 11 systematic reviews every day (Bastian et al. 2010). It should not surprise us if established physicians are not integrating new medical knowledge as consistently as their junior colleagues (Lipner et al. 2011). Additionally, the decay in knowledge and skills that occurs naturally with infrequent use and the effects of age can be mitigated or at least recognized with a CBME approach to continuing learning and appropriate ongoing assessment (Baxter et al. 2014). As Lockyer et al. (2017b) discuss in this issue, assessment needs to extend beyond the completion of residency training to become routine within clinicians' continuing professional development (CPD).

Continuing professional development has long recognized that many competencies are realized only with independent practice, and that graduation is simply a waypoint on the journey. Assessment structured around real tasks can motivate continuing learning and provide useful guidance about expected standards of practice. Self-assessment alone is inadequate for identifying learning and performance needs (Eva & Regehr 2005).

CBME can provide a logical structure for information on learning development to flow forward to the next training stage and into CPD. An informative appraisal of achievements can help structure the CPD profile or possibly influence areas of subspecialization or practice focus. These data would, at least, identify milestones that were lagging at graduation and that could benefit from enrichment

during the clinical years immediately after graduation. As this process is a continuum, some modules developed for trainees may be usefully employed in such structured CPD and vice versa.

Conclusions

The rationale for moving to CBME includes a clearer understanding of levels of achievement and improved accountability of both trainees and training organizations. One implication of these clearer and more practice-focused competency frameworks is greater attention to all trainees. Assessment within CBME should be based on competency frameworks and develop an educational plan based on the differences between desired and observed performance. Milestones provide a clear articulation of expected, sequentially acquired behaviors and encourage the formation of a shared mental model to guide learners and assessors. The focus for assessment in CBME is on individual development based on clear, regular, and expected feedback. Assessment in CBME is not simply about meeting bare competence, but is also intended to inspire learners and programs to strive for excellence.

Validity in the context of CBME requires a shift away from the statistical analysis of single tools toward the evaluation of inferences and actions that derive from a *program* of assessment. Authentic workplace-based assessment targeted at the "does" level of Miller's classic description is critical to this notion of validity. Judgments of progression or competence need to involve the "wisdom of crowds" and to be conducted away from the teacher–learner encounter and away from individual assessment tasks. Qualitative (or narrative) data may be critically important to these discussions. Interaction and supportive feedback promotes *assessment for learning* and creates a culture where they are not only accepted but also desired. Finally, the CBME approach to assessment has considerable potential in CPD, where despite its current underutilization it has the potential to make the largest impact on physician development and patient care.

Many assessment issues related to CBME remain unresolved, and more data on the effective application of this approach are needed. Holmboe et al.'s (2010) description of assessment in CBME has been followed by substantial contributions to the theory and practice of CBME-related assessment. New publications evaluating the impact of competency-based programs have been reported (see Ferguson et al. 2017, in this issue) and we anticipate that further contributions in the near future will help guide subsequent implementation.

Disclosure statement

Notes on contributors

Peter Harris, MB BS, is a Senior Lecturer in the Office of Medical Education, University of New South Wales, Australia.

Farhan Bhanji, MD, is an Associate Director of Assessment and Clinician Educator at the Royal College of Physicians and Surgeons of Canada, and Professor of Pediatrics and Director of Pediatric Simulation at McGill University, Canada

Maureen Topps, MB ChB, is an Associate Dean, Postgraduate Medical Education, and Associate Professor, Department of Family Medicine, University of Calgary, Canada.

Shelley Ross, PhD, is an Associate Professor, and Director of Research and Innovation, Competency-Based Achievement System Program, in the Department of Family Medicine, Faculty of Medicine & Dentistry, University of Alberta, Canada.

Steven A. Lieberman, MD, is a Professor and Senior Dean for Administration, Office of the Dean of Medicine, Department of Internal Medicine, Division of Endocrinology, University of Texas Medical Branch at Galveston, USA.

Jason R. Frank, MD, is the Director, Specialty Education, Strategy and Standards in the Office of Specialty Education at the Royal College of Physicians and Surgeons of Canada, and the Director of Educational Research & Development in the Department of Emergency Medicine, University of Ottawa, Canada.

Linda Snell, MD, is a Professor of Medicine and Core Faculty member, Centre for Medical and Department of General Internal Medicine, McGill University, Canada, and Senior Clinician Educator, Royal College of Physicians and Surgeons of Canada, Canada.

Jonathan Sherbino, MD, is an Associate Professor in the Department of Medicine, McMaster University, and an Assistant Dean with Program for Education Research and Development, McMaster University, Canada.

ORCID

Shelley Ross http://orcid.org/0000-0001-9581-3191

References

Bastian H, Glasziou P, Chalmers I. 2010. Seventy-five trials and eleven systematic reviews a day: How will we ever keep up? PLoS Med. 7:e1000326.

Baxter AD, Boet S, Reid D, Skidmore G. 2014. The aging anesthesiologist: a narrative review and suggested strategies. Can J Anaesth. 61:865–875.

Bazrafkan L, Ghassemi GH, Nabeiei P. 2013. Feedback is good or bad? Medical residents' point of view on feedback in clinical education. J Adv Med Educ Prof. 1:51–54.

Black P, William D. 1998. Assessment and classroom learning. Assess Educ Princ Policy Pract. 5:7–74.

Brown S. 2004. Assessment for learning. Learn Teach Higher Educ. 1:81–89.

Carraccio C, Englander RA. 2014. A milestone for the pediatrics milestones. Acad Pediatr. 14:S1–S3.

Carraccio C, Wolfsthal SD, Englander R, Ferentz K, Martin C. 2002. Shifting paradigms: from Flexner to competencies. Acad Med. 77:361–367.

Cook DA, Brydges R, Ginsburg S, Hatala R. 2015. A contemporary approach to validity arguments: a practical guide to Kane's framework. Med Educ. 49:560–575.

Cook DA, Zendejas B, Hamstra SJ, Hatala R, Brydges R. 2014. What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. Adv in Health Sci Educ. 19:233–250.

Crossley J, Jolly B. 2012. Making sense of work-based assessment: ask the right questions, in the right way, about the right things, of the right people. Med Educ. 46:28–37.

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. 296:1094–1102.

Donato AA, George DL. 2012. A blueprint for implementation of a structured portfolio in an internal medicine residency. Acad Med. 87:185–191.

Driessen E, van der Vleuten CPM, Schuwirth LWT, van Tartwijk J, Vermunt J. 2005. The use of qualitative research criteria for portfolio assessment as an alternative to reliability evaluation: a case study. Med Educ. 39:214–220.

Dudek NL, Marks MB, Regehr G. 2005. Failure to fail: the perspectives of clinical supervisors. Acad Med. 80:S84–S87.

Duffy FD, Holmboe ES. 2006. Self-assessment in lifelong learning and improving performance in practice: physician know thyself. JAMA. 296:1137–1139.

Englander R, Frank JR, Carraccio C, Sherbino J, Ross S, Snell L; ICBME Collaborators. 2017. Toward a shared language for competency-based medical education. Med Teach. 39:582–587.

Eva KW, Regehr G. 2005. Self-assessment in the health professions: A reformulation and research agenda. Acad Med. 80:S46–S54.

Ferguson PC, Caverzagie KJ, Nousiainen MT, Snell L; ICBME Collaborators. 2017. Changing the culture of medical training: an important step toward the implementation of competency-based medical education. Med Teach. 39:599–602.

Frank JR, Snell LS, ten Cate O, Holmboe ES, Carraccio C, Swing SR, Harris P, Glasgow NJ, Campbell C, Dath D, et al. 2010. Competency-based medical education: theory to practice. Med Teach. 32:638–645.

Frank JR, Snell LS, Sherbino J, editors. 2014. Draft CanMEDS 2015 Milestones Guide – September 2014. Ottawa: Royal College of Physicians and Surgeons of Canada.

Galbraith RM, Hawkins RE, Holmboe ES. 2008. Making self-assessment more effective. J Contin Educ Health Prof. 28:20–24.

Govaerts MJB, van der Vleuten CPM, Schuwirth LWT, Muijtjens AMM. 2007. Broadening perspectives on clinical performance assessment: rethinking the nature of in-training assessment. Adv Health Sci Educ Theory Pract. 12:239–260.

Govaerts MJB. 2008. Educational competencies or education for professional competence? Med Educ. 42:234–236.

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. 75:167–173.

Hodges B. 2006. Medical education and the maintenance of incompetence. Med Teach. 28:690–696.

Hodges B. 2013. Assessment in the post-psychometric era: learning to love the subjective and collective. Med Teach. 35:564–568.

Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR; International CBME Collaborators. 2010. The role of assessment in competency-based medical education. Med Teach. 32:676–682.

Holmboe ES, Yamazaki K, Edgar L, Conforti L, Yaghmour N, Miller RS, Hamstra SJ. 2015. Reflections on the first 2 years of milestone implementation. J Grad Med Educ. 7:506–511.

Kane MT. 1992. An argument-based approach to validation. Psychol Bull. 112:527–535.

Lipner R, Song H, Biester T, Rhodes R. 2011. Factors that influence general internists' and surgeons' performance on maintenance of certification exams. Acad Med. 86:53–58.

Lockyer J, Bursey F, Richardson D, Frank JR, Snell L, Campbell C; ICBME Collaborators. 2017a. Competency-based medical education and continuing professional development: a conceptualization for change. Med Teach. 39:617–622.

Lockyer J, Carraccio C, Chan MK, Hart D, Smee S, Touchie C, Holmboe ES, Frank JR; ICBME Collaborators. 2017b. Core principles of assessment in competency-based medical education. Med Teach. 39:609–616.

Messick S. 1989. Validity. In: RL Linn, editor. Educational measurement, 3rd edition. New York: American Council on Education and Macmillan.

Miller GE. 1990. The assessment of clinical skills/competence/performance. Acad Med. 65:S63–S67.

Moonen-van Loon JMW, Overeem K, Donkers HHLM, van der Vleuten CPM, Driessen EW. 2013. Composite reliability of a workplace-based assessment toolbox for postgraduate medical education. Adv in Health Sci Educ. 18:1087–1102.

Ramani S, Krakov SK. 2012. Twelve tips for giving feedback effectively in the clinical environment. Med Teach. 34:787–791.

Schuwirth LW, van der Vleuten CPM. 2011. Programmatic assessment: from assessment of learning to assessment for learning. Med Teach. 33:478–485.

Schuwirth LWT, van der Vleuten C. 2012. Programmatic assessment and Kane's validity perspective. Med Educ. 46:38–48.

Schuwirth L, Ash J. 2013. Assessing tomorrow's learners: in competency-based education only a radically different holistic method of assessment will work. Six things we could forget. Med Teach. 35:555–559.

Schuwirth LWT, van der Vleuten C. 2014. Assessing competence: extending the approaches to reliability. In: Hodges B, Lingard L, editors. The question of competence. Ithaca (NY): Cornell; p. 125.

ten Cate O, Hart D, Ankel F, Busari J, Englander R, Glasgow N, Holmboe E, Iobst W, Lovell E, Snell LS, et al. 2016. Entrustment decision making in clinical training . Acad Med. 91:191–198.

van der Vleuten CPM, Schuwirth LWT. 2005. Assessing professional competence: from methods to programmes. Med Educ. 39:309–317.

van der Vleuten CPM, Schuwirth LWT, Scheele F, Driessen EW, Hodges B. 2010. The assessment of professional competence: building blocks for theory development. Best Pract Res Clin Obstet Gynaecol. 24:703–719.

Watling C, Driessen E, van der Vleuten CP, Vanstone M, Lingard L. 2013. Music lessons: revealing medicine's learning culture through a comparison with that of music. Med Educ. 47:842–850.