

Competency-Based Education: Milestones or Millstones¹?

GEOFF NORMAN, PHD
JOHN NORCINI, PHD
GEORGES BORDAGE, MD, PHD

In the past decade, there has been increasing interest in competency-based education (CBE), the notion that an expert physician is defined by a broad set of identified competencies. The idea has been advanced, nearly simultaneously, in several countries—Canada (CanMEDS roles),¹ the United States (Accreditation Council for Graduate Medical Education [ACGME] competencies),² the United Kingdom (Tomorrow's Doctor),³ and Scotland (the Scottish Doctor)⁴—and adopted by others. CanMEDS competencies have been adopted and adapted by 16 countries, including the Netherlands, Denmark, and Mexico.⁵ Moreover, the Carnegie Foundation's influential Flexner centenary strongly recommended adopting CBE and made the claim that “Adoption of OBE [outcome-based or competency-based education] would better equip medical graduates to respond effectively in complex situations and efficiently continue to expand the depth and breadth of the requisite competencies.”^{6,7} Similar promises emerge from many of these foundational documents. Harden^{8(p666)} ascribes a number of advantages to OBE/CBE:

OBE is a sophisticated strategy for curriculum planning that offers a number of advantages. It is an intuitive approach that engages the range of stakeholders . . . it encourages a student-centred approach and at the same time supports the trend for greater accountability and quality assurance . . . [it] highlights areas in the curriculum which may be neglected . . . such as ethics and attitudes . . .

Regrettably, these declarations appear to be more a matter of faith than of evidence.

The primary intent of CBE is, we believe, transparency, so that the profession and the public can be confident that a training program is producing competent physicians who

BOX MILESTONE CONCERNS AND POTENTIAL BENEFITS

Concerns

Conceptual

- Breadth versus precision of competencies
- Differential focus on more procedural competencies
- Evidence of effectiveness

Psychometric

- Number of competencies
- Sampling error
- Growth curve anomalies

Logistical

- Ability to identify attainment
- Implications for length of training
- Resource implications

Potential benefits

Curriculum

- Clear specification of competencies to be achieved
- Potential to guide curriculum implementation

Assessment in course

- Increased observation and feedback
- More specific direction for learning

In-course assessment

- Greater and more systematic evidence for summative evaluation in course

Licensure and certification

- Greater consistency between curriculum and board examination
- Greater accountability of programs
- More accurate and useful accreditation

are equipped with the knowledge and skills for practice. It is hard to challenge that premise; the issue is whether the proposed mechanisms can deliver on the promise. A corollary common to both CBE and its predecessor—behavioral objectives—is the notion that different learners will achieve different competencies at different rates, so that residents may be certified competent in starting an intravenous line early in their career but may take longer to achieve competency in intubation. The individual need not take additional time practicing skills for which he or she is competent and can, therefore, learn more efficiently by focusing on those skills for which competence has not yet been achieved. An extrapolation of that notion is that some residents may well achieve all the competencies available on a particular rotation earlier or later than others, and so, can progress through graduate medical education (GME) at a different pace.

Whether the approach can be operationalized satisfactorily at the level of precision required to implement CBE, the fact remains that it may have positive side benefits, such as increased observations of residents, greater attention to

¹ The title was first used by Trudy Roberts of the University of Leeds in an Association for Medical Education in Europe symposium but occurred independently to the first author.

Geoff Norman, PhD, is Professor, Department of Clinical Epidemiology and Biostatistics, Program for Educational Research and Development, McMaster University, Hamilton, Ontario, Canada; **John Norcini, PhD**, is President, Foundation for Advancement of International Medical Education; and **Georges Bordage, MD, PhD**, is Emeritus Professor, Department of Medical Education, University of Illinois at Chicago.

Corresponding author: Geoff Norman, PhD, Department of Clinical Epidemiology and Biostatistics, Program for Educational Research and Development, McMaster University, Room 3519, MDCL, 1200 Main Street West, Hamilton, ON L8N 3Z5, Canada, 905.525.9140 ex. 23114, norman@mcmaster.ca

DOI: <http://dx.doi.org/10.4300/JGME-D-13-00445.1>

the GME curriculum, and so forth. To ensure that those goals are met and the implementation of CBE is not sidelined by a ponderous administrative superstructure, this editorial is intended to elaborate potential problems at 3 levels: conceptual, psychometric, and logistic.

Conceptual Problems

One purported advantage of CBE is that the competencies appear, on the surface, to have the virtue of simplicity. They are generally few in number, at least at the top level. CanMEDS lists 8 roles, ACGME has 6 competencies, and the Scottish Doctor lists 7. Remarkably, there appears to be little overlap among many of the specifications.⁹ For example, mapping CANMEDs roles of “scholar,” “collaborator,” or “advocate” onto the ACGME competencies of practice-based learning, systems-based practice, or professionalism is not at all straightforward. Moreover, although an economy of roles is touted as 1 feature that sets CBE apart from its historical predecessor, behavioral objectives, this parsimony is rapidly lost in the further, more detailed specification of each role or competency. For example, the Scottish Doctor lists 86 subcompetencies under 12 competencies. Similarly, the ACGME competencies have been adapted to various specialties; in internal medicine, that has led to 142 subcompetencies.¹⁰ Yet apparently, that specificity is insufficient; for example, one competency is “the ability to perform and interpret common laboratory tests,” but another document, admittedly somewhat dated, related to the practice of internal medicine in the United States lists 39 different laboratory tests that a graduating internist should master.¹¹

The argument that competencies are somehow different from behavioral objectives is difficult to maintain in the face of this proliferation of competencies. Indeed, Morcke and colleagues⁷ described a direct lineage between the behavioral objectives movement of the 1970s and current interest in outcomes or competencies. In view of the similarity between behavioral objectives and competencies, one should consider the factors that led to the ultimate demise of behavioral objectives: lack of evidence of clear educational benefits, the effort required to specify objectives and then implement educational and assessment approaches, and the dominance of low-level objectives. The fate of behavioral objectives might serve as a cautionary note to advocates of CBE. That, however, does not appear to be the case; the historical precedent appears to have been dealt with by attempting to distance CBE from behavioral objectives.¹² Such a strategy does not survive anything but the most cursory inspection.

Moreover, some of the detailed critiques of behaviorism have been revisited in challenges to CBE. In particular, as discussed by Harden,⁸ a purported advantage of CBE is that

it encourages a more holistic view of the nature of clinical competence than simple mastery of a body of knowledge evidenced by passing an examination. That broad perspective is exemplified by the CanMEDs roles, in which the “expert” role is surrounded (one might even say “protected”) by 7 additional competencies.² However, paradoxically, the requirement for detailed specification of competencies may mitigate against such a holistic view, as pointed out by Morcke et al,⁷ since qualities, such as humanism, altruism, professionalism, and scholarship, are difficult to define in objective, measurable terms, and thus may receive short shrift in assessment. Indeed, a recent study showed that supervisors are averse to attempting to assess such qualities in students, and when they do, the assessments are so unreliable as to be nearly useless.¹³

This discussion does not begin to exhaust the numerous theoretical challenges to CBE. However, CBE will be adopted despite its detractors. Consequently, it may be more valuable to turn to a critical examination of the application and implementation of CBE.

Psychometric Considerations

Although the long lists of competencies belie the notion that CBE is distinguishable from behavioral objectives, the 2 approaches do diverge at the level of measurement. Consistent with the behaviorist research agenda, achievement of behavioral objectives was viewed as binary. An objective was or was not achieved, and developers of objectives were admonished to ensure that the objectives were framed in such a way that such judgments were unequivocal. Whether such objectivity was achieved, or could be achieved, is debatable.¹⁴ That is not relevant to the present discussion.

Perhaps appropriately, the CBE movement has taken a more realistic perspective on the achievement of competencies. There is recognition that, however atomistic a competency is specified (a point we return to in the next section), describing its achievement as a step function is unrealistic. A child does not crawl one day and walk the next: Walking is a gradual process of mastery that occurs over weeks or months.

This notion of the gradual acquisition of competence has been represented as a nonlinear growth curve, where growth of competence is initially rapid, but then decelerates (FIGURE 1).¹⁵ These hypothetical curves have been described in several documents and appear a plausible way to represent the growth of competence or skill in a domain.¹⁵

There is some evidence to support the hypothesis that simple motor skills do follow a power law that resembles those curves. However, such curves are derived from a

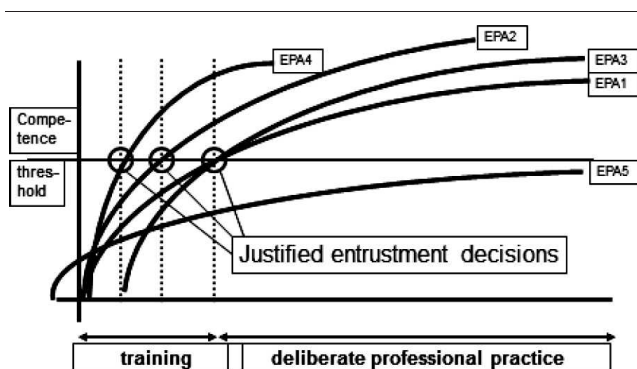


FIGURE 1 **HYPOTHETICAL GROWTH CURVES FOR A SINGLE TRAINEE FOR DIFFERENT ENTRUSTABLE PROFESSIONAL ACTIVITIES (EPAs)**

From Ten Cate et al.¹⁵ Medical competence: the interplay between individual ability and the health care environment. *Med Teach*. 2010;32(8):669–675. Adapted with permission.

simple repetitive task (in one study, cigar-rolling) and a single session. Studies of more representative tasks, such as experiments in which individuals seek to master a simple task (pursuit rotor task, in which the subject is required to keep a flexible metal wand on a rotating metal dot) over successive days show that performance at each new session shows some regression over the end of the previous session, but there is monotonic growth across sessions (FIGURE 2).¹⁶ The implication is that a particular level of competence—a point on the y-axis—can be achieved and “unachieved,” and there may be multiple points where competence is achieved. For example, if we define mastery as a time on target of 7 seconds, the achievement of mastery in rotor pursuit would occur after 30, 44, or 72 trials, depending on the session.¹⁶ Further, although these curves may be a more valid representation of the growth in specific competencies in the residency setting, which occurs in fits and starts over long periods, it should be emphasized that “. . . learning curves of almost every conceivable shape can and have been found.”¹⁶

We are unaware of any data describing the growth of competency in GME; however, some sense of the magnitude of the measurement problem can be derived from experience with progress tests, such as the 180-item, multiple-choice test administered to McMaster undergraduates every 4 months (FIGURE 3). The growth curve from the successive administration of those tests is similar to those postulated for Milestones (FIGURE 1). If we were to set competency for independent practice at 50%, it could be achieved on an individual test at any time from 60 to 120 months. However, these actual data are derived from class means with a sample size 100, and the variability in

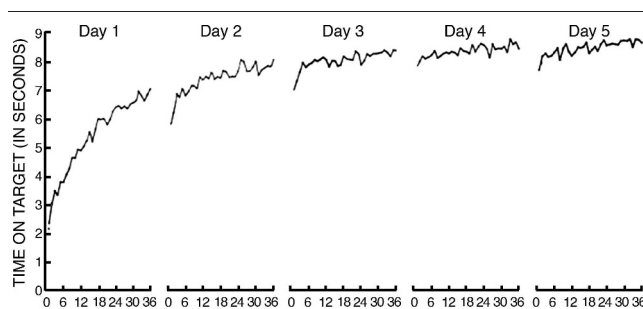


FIGURE 2 **GROWTH CURVE**

Adapted from *American Journal of Psychology*. Copyright 1952 by the Board of Trustees of the University of Illinois. Used with the permission of the University of Illinois Press.

scores at any time point is derived almost entirely from differences in test difficulty, despite the fact that the individual tests are based on 180 items randomly sampled from a fixed-item pool.

It is apparent from these examples that detailed specification of the acquisition of any ACGME competency, to the point that it can be identified with the attainment of a level consistent with independent practice, at a particular point in time, will require an unrealistic sample of observations.

An alternative, of course, is to abandon any attempt at detailed specification of competencies and proceed on the basis of the general competencies. However, doing so appears to introduce additional issues. Sherbino et al,¹³ examining implementation of a rating of CanMEDS roles in an emergency medicine clerkship, showed a reliability of 0.12 between ratings, which would require a minimum of 17 ratings per competency.¹³ Equally problematic, the internal consistency of the scale of CanMEDS competencies was 0.94, and a factor analysis revealed one g factor accounting for 87% of the variance. Raters were unable to distinguish among those broad competencies and had poor agreement. Similarly, an evaluation of a global rating based on the CanMEDS roles,¹⁷ although it yielded 5 factors in a factor analysis, had an overall internal consistency of 0.97, again showing a lack of discrimination between scales.

It may be that these competencies are so broad that the lack of discrimination simply reflects the inability of assessors to adequately define and internalize the meaning of the terms. It may be likened to the difference between a question like “How well do you know European geography?” and “What are the capitals of the following western European countries?” It is possible then, that improved overall assessment may result from observations at a more detailed level.

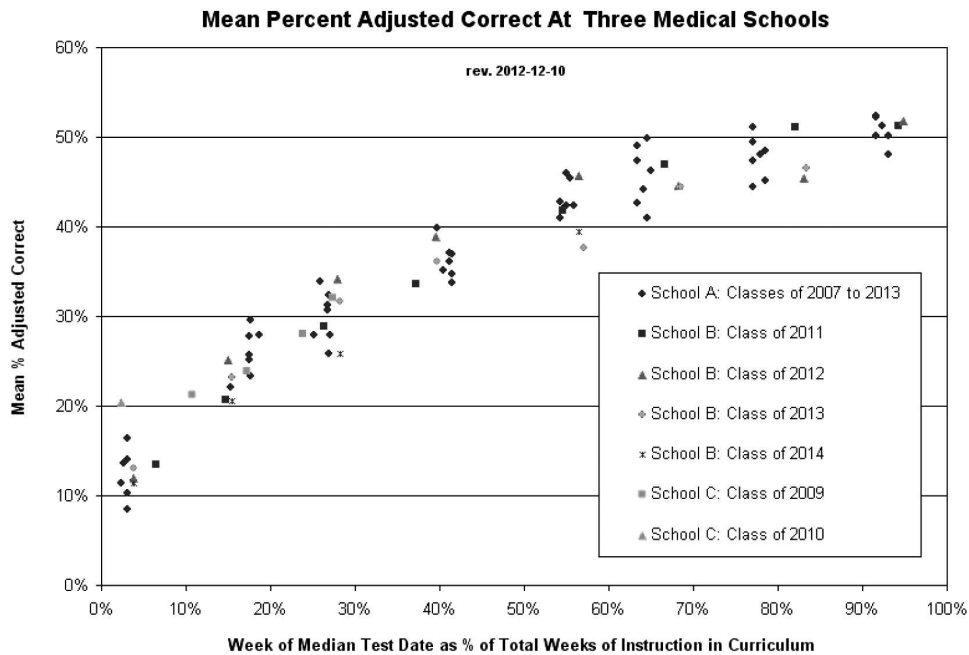


FIGURE 3 | **PROGRESS TEST SCORE BY YEAR OF TRAINING IN 3 INSTITUTIONS**

G. Norman, PhD. Unpublished data, 2013.

Of course, it may also be that at this broad level, the competencies overlap to a substantial degree or there are very few students with focal deficits. The latter would not be terribly surprising given that the entire educational enterprise is aimed at producing graduates with competence across a number of domains.

One further consideration is the issue of sampling. Although there is utility in the specification of competencies at a detailed level, it remains the case that most assessment methods show low, but consistent, correlations across domains. That being so, resident assessment can be derived from a deliberate sampling strategy, based on research to determine how many samples and what sampling strategy may suffice.

On the one hand, detailed specification of competencies so that relatively objective documentation can be obtained results in an overwhelming number of assessments per resident. On the other hand, lumping competencies together precludes accurate and helpful feedback and leads to questioning precisely what it means to achieve competence. Moreover, it appears that judges cannot discriminate individual competencies even at this global level of specification, and many observations are needed to achieve even minimal levels of reliability.

It follows then that reasonable summative assessment is possible at a broad level, but that formative assessment must necessarily be based on observations at a more detailed level.

Logistic Problems

The evidence we have presented to date casts serious doubt on the ability of any GME curriculum based on attainment of individual competency to actually identify, with any degree of precision, a time point when an individual learner has achieved a prescribed level of competence, such as needed for specialty Milestones. If that could be achieved, however, with a feasible number of assessments, it is unclear what the consequences might be.

Competency-based education has all the logistical problems of mastery learning and many more. However attractive it may be to presume that some learners may attain competency more quickly than others or, more precisely, that some learners may achieve *some competencies* more quickly than others do, those activities take place in a setting where the primary responsibility is to patient care. Some individualization may be possible; one resident may “sign off” on a particular procedure early and then redirect time to a different procedure. However, it is not practical to have faculty available to assess particular competencies at any point in time. Moreover, it is simply not practical to have variable training times.

Contracts are given to residents for specific periods (usually a year), and it would be difficult to create a contract to allow residents to leave a program earlier or later. Currently, institutions cannot plan schedules and budgets based on residents having individualized lengths of training. Currently, those few residents who do not achieve

competencies within that year are given x number of months additional time for remediation. That disrupts residency program schedules but is manageable because the numbers are small. At present, individualizing training periods for each resident is not feasible, thus the claims that Milestones can save money or time by possibly graduating residents earlier appear specious.

The implication of this viewpoint is unsettling. Competency-based education promises that different individuals will achieve competency in different domains at different times. Although that may lead to hypothetical efficiencies in allocation of educational resources and resident time, it is incompatible with the current realities of GME.

On the other hand, the detailed specification of competencies may provide a basis to critically examine the GME curriculum and refocus training on mastery of specific competencies rather than continued exposure to competencies that have been clearly achieved.

Is There Any Advantage for CBE?

The problems we just reviewed point to serious difficulties associated with any attempt to precisely link GME to defined levels of performance outcomes achieved at different time points. However, CBE may still offer real, indirect benefits over the traditional, time-based approach to GME.

One can envision at least 4 distinct roles for specified competencies:

1. Guiding learning for a student or resident
2. Giving daily feedback to residents and students at the bedside and in clinic
3. Making pass/fail decisions at the end of a rotation and the end of a program
4. Making pass/fail decisions for licensure and certification

Guiding Learning

A detailed specification of competency based on explicit unpacking of the broad CanMEDS or ACGME competencies, as has been attempted by many specialties in the United States, may have utility on a number of levels. First, it can be used in program planning to ensure that each resident is at least minimally exposed to every competency. That may lead to more attention on the curriculum and deliberate supplementing of actual clinical experiences with simulations, for example, to ensure that those competencies are addressed.

Second, such a description could also easily be operationalized for specific skills, such as diagnostic procedures,

by introducing a procedure log that measures degrees of competence rather than counts the number of procedures, as is currently mandated in procedural specialties.

Assessment and Feedback

At minimum, the requirement for certification of specific competencies must inevitably lead to increased observation of resident performance, an area that is historically underserved. Just as the mini-CEX (clinical evaluation exercise) has led to greater numbers of observations, a CBE framework may have this desirable outcome. Although it may never be feasible to conduct sufficient numbers of observations to “sign off” on all competencies, increased observation would bring its own benefits.

Determining Pass and Fail in Programs

To the extent that individual programs devise standardized tests of knowledge and performance, the competency framework provides a readily available blueprint for test sampling. Again, the impossibility of observing all competencies sufficiently is recognized, but the blueprint can provide increased confidence on the representativeness of the content of in-training assessments and examinations. In turn, that may increase the confidence with which programs can certify achievement of residency program objectives.

Accreditation and Certification

Similarly, the competency framework is a natural blueprint for specialty boards to use in examination specification. It is self-evident, we believe, from the earlier discussion that CBE is unlikely to replace board examinations. However, the competency statement is a natural starting point for design of board examinations, again directed at assurance of content representativeness (validity), and ultimately, improved assessment of resident performance.

These goals are achievable without implementation of a detailed assessment program to attempt, however optimistically, to document achievement of every Milestone at every level. It may well be sufficient to implement the detailed competencies at a curriculum level, while striving for assessment of residents at an intermediate level of specificity, accompanied by detailed and constructive feedback based on direct observation.

A further advantage is that certifying boards will have more systematic evidence on which to base accreditation procedures. Some ability to track areas within programs where objectives are and are not attained may provide useful diagnostic information to improve programs.

However, to achieve a defensible CBE curriculum at a postgraduate level, implementation must be accompanied by ongoing research to identify acceptable levels of assessment activity that provide defensible assessment of

individuals while avoiding crippling the system with excessive assessment activity.

Competency-based education may never achieve the many virtues attributed to it by its most enthusiastic proponents. A century of educational research clearly demonstrates that there are no educational “wonder drugs.” Nevertheless, the change of focus toward achievement of competency, rather than putting in time, is likely to have demonstrable, if incremental, benefits. The challenge is to reallocate resources strategically to maximize benefit while avoiding creating an administrative elephant.

References

- 1 Frank JR, Snell LS, ten Cate O, Holmboe ES, Carraccio C, Swing SR, et al. Competency-based medical education: theory to practice. *Med Teach*. 2010;32(8):638–645.
- 2 Batalden P, Leach D, Swing S, Dreyfus H, Dreyfus S. General competencies and accreditation in graduate medical education. *Health Aff (Millwood)*. 2002;21(5):103–111.
- 3 Rubin P, Franchi-Christopher D. New edition of tomorrow's doctors. *Med Teach*. 2002;24(4):368–369.
- 4 Simpson JG, Furnace J, Crosby J, Cumming AD, Evans PA, Friedman Ben David M, et al. The Scottish doctor—learning outcomes for the medical undergraduate in Scotland: a foundation for competent and reflective practitioners. *Med Teach*. 2002;24(2):136–143.
- 5 Whitehead CR, Austin Z, Hodges BD. Flower power: the armoured expert in the CanMEDS competency framework? *Adv Health Sci Educ Theory Pract*. 2011;16(5):681–694.
- 6 Cooke M, Irby DM, O'Brien BC. *Educating Physicians: A Call for Reform of Medical School and Residency*. Hoboken, NJ: Jossey-Bass; 2010.
- 7 Morcke AM, Dornan T, Elka B. Outcome (competency) based education: an exploration of its origins, theoretical basis and empirical evidence. *Adv Health Sci Educ Theory Pract*. 2013;18(4):851–863.
- 8 Harden RM. Outcome-based education: the ostrich, the peacock and the beaver. *Med Teach*. 2007;29(7):666–671.
- 9 Albanese MA, Nejicano G, Mullan P, Kokotailo P, Gruppen L. Defining characteristics of educational competencies. *Med Educ*. 2008;42(3):248–255.
- 10 Green ML, Aagaard EM, Caverzagie KJ, Chick DA, Holmboe E, Kane G, et al. Charting the road to competence: developmental milestones for internal medicine residency training. *J Grad Med Educ*. 2009;1(1):5–20.
- 11 Wigton RS, Nicolas JA, Blank LL. Procedural skills of the general internist: a survey of 2500 physicians. *Ann Intern Med*. 1989;111:1023–1034.
- 12 Harden RM. Learning outcomes and instructional objectives: is there a difference? *Med Teach*. 2002;24(2):151–155.
- 13 Sherbino J, Kulsegaram K, Worster A, Norman GR. The reliability of encounter cards to assess the CanMEDS roles. *Adv Health Sci Educ Theory Pract*. 2013;18(5):987–996.
- 14 Van der Vleuten CP, Norman GR, De Graaf ED. Pitfalls in the pursuit of objectivity: issues of reliability. *Med Educ*. 1991;25(2):110–118.
- 15 ten Cate O, Snell L, Carraccio C. Medical competence: the interplay between individual ability and the health care environment. *Med Teach*. 2010;32(8):669–675.
- 16 Newell KM, Liu YT, Mayer-Kress G. Time scales in motor learning and development. *Psychol Rev*. 2001;108(1):57–82.
- 17 Kassam A, Donnon T, Rigby I. Validity and reliability of an in-training evaluation report to measure the CanMEDS Roles in emergency medicine residents. *CJEM*. 2013;15(0):1–7.