

Two Cheers for Milestones

LOUIS N. PANGARO, MD, MACP

This editorial will explore the implementation of milestones across graduate medical education (GME) from 2 perspectives. The first is my perspective as a clinician, who often asks, “How do I make decisions with a patient when there isn’t evidence to use as a guideline?” The second is my perspective as a department chair who asks a different question: “What resources are needed for milestone implementation?”

In medical education, the broader question that calls for our clinician judgment is how do we make decisions when we perceive a need for action, but have incomplete scholarship for evidence-based practice? When do we need an Institutional Review Board to decide whether our activities are research or just practical decision making? How can we have educational innovations (like the milestones) in response to new challenges and social pressure for public oversight of GME, if we are wholly committed to evidence-based practice?

An implicit question is whether the implementation of milestones is justified. As the reader can see from the title of this article, my own judgment is in favor of milestones, but I believe that we need to withhold the third cheer for now. In any case, I do not write as a cheerleader, but as someone responding to the practical problems facing my faculty.

Milestones are defined as “competency-based developmental outcomes (eg, knowledge, skills, attitudes, and performance) that can be demonstrated progressively by residents and fellows from the beginning of their education through graduation to the unsupervised practice of their specialties.”¹ The purpose for their implementation is: “First and foremost, the milestones are designed to help all residencies and fellowships produce highly competent physicians to meet the health and health care needs of the public.”¹

My first cheer is because the milestones are more useful to my teachers than the competencies have been, and milestone development has been based on communities of practice.²

My second cheer is for the fact that milestones represent an advance over the abstraction of the compe-

tencies. The third cheer is withheld because the evidence base remains weak, and resources for systematic studies—of efficacy, implementation cost, and opportunity cost, which clinicians expect of diagnostic tools—are not yet sufficiently clear.

The First Cheer

The milestones are a practical solution to the challenge underlying the desire of the Accreditation Council for Graduate Medical Education (ACGME) to demonstrate through outcomes that GME is meeting the needs of our society.³ With so much public money, especially from Medicare, poured into GME, the stakes are very high for professional oversight of GME. Even more important, the Next Accreditation System³ (NAS) represents an unmistakable message to society, and to academic medicine, that we serve others, not ourselves. If the milestones need regulatory power to give force to this message, this should be supported. There is no clearer statement about fulfilling our educational promise of duty and expertise (to paraphrase Pellegrino⁴), than how seriously we take this task.

Over time, it may become clear to trainees that milestones have the potential to minimize interrater differences in evaluations across teachers; it may be less of a “crapshoot” as to who is grading. Fairness to society, learners, and teachers has to be a prime characteristic of our profession’s assessment system and, in fact, *evaluation equals professionalism*.

Impressive effort from colleagues across the country has gone into the development of each specialty’s subcompetencies and milestones. Although the directive came from above, the process has been bottom-up. In addition to the hours of committee time, there are now several hundred articles available through the PubMed database (US National Library of Medicine). To me, this has been an essential part of the process: milestones have been “developed through engagement of the specialty community.”⁵

As in clinical situations, expert consensus and national discussion around milestone development and implementation lends legitimacy. This is similar to sources I rely on in discussions with patients when no empirical evidence is available on which to base one’s recommendation.

The Second Cheer

The second cheer is for the theoretic advance that the milestones represent over the competencies. The analogy

Louis N. Pangaro, MD, MACP, is Professor and Chair, Department of Medicine, Uniformed Services University of the Health Sciences.

The opinions herein are the author’s own and do not reflect the Uniformed Services University of the Health Sciences or the US Department of Defense.

Corresponding author: Louis N. Pangaro, MD, MACP, Uniformed Services University of the Health Sciences, 4301 Jones Bridge Road, Bethesda, MD 20814-4799, 301.295.2010, fax 301.295.5792, louisipangaro@aol.com

DOI: <http://dx.doi.org/10.4300/JGME-D-14-00738.1>

here is that, in making clinical decisions without applicable evidence, we rely—usually successfully—on basic principles (physiology, pharmacology, etc) to ground our decisions. Theoretic constructs underlying educational frameworks⁶ are important.

The 6 general competencies are phrased as abstractions. A good deal of writing on the milestones has reflected the need for something more explicit. Our colleagues from Canada, in their CanMEDS competency framework,⁷ framed the 7 domains of competence as roles rather than abstractions. For purposes of simplifying for my own faculty, we refer to the ACGME patient care competency as the “history of present illness” and the other 5 competencies as a “review of systems,” which ensures we don’t miss anything.

Perhaps more important, the milestones are concrete tasks that synthesize⁸ components of knowledge, skills, and attitudes. Studies will probably demonstrate more accurate use by faculty, with less training time. While observing a resident, the teacher is asked to judge whether a task is being performed correctly and with what degree of proficiency. Rather than using a menu of 6 competencies, the teacher observes 1 task, and relies on the behavioral anchors of the milestones to clarify areas for improvement and feedback.

The basis of most judgments—whether in diagnosing pneumonia or determining how a resident performs—is a side-to-side comparison between what we observe in front of us (the actual) and what we expect to see (the ideal). Education provides a rich description of the latter, and training develops our skills in the former. The milestones are designed to provide definitive expectations for residents that can be visualized by trained faculty; in this way, they are an advance on the abstractions of the 6 competencies. I emphasize concreteness of roles and tasks, since I believe that simplicity leads to acceptance, acceptance to consistency, consistency to fairness, and that fairness is a mark of professionalism.

The issue is how to elaborate a framework for assessment that can encompass or reflect a sharable concept of *competence* and also be simple, without being simplistic. In any case, my faculty will still need training in the milestones, but I anticipate less training than for the competencies themselves.

I expect that the developmental aspect of the milestone framework can be used by my faculty as they make judgments about whether residents are prepared to attempt a task, whether they can do it with direct supervision, whether they can do it with remote supervision, whether they can do it unsupervised, or—in the aspirational range—whether they can teach others. It remains to be seen whether most faculty can use this reliably (ie, with good

interrater agreement), but for years they have used an implicit, normative model when comparing one resident to another. I think that, with training, faculty can translate a progressive mastery model into a criterion-based framework. Still, there is a risk that the faculty will revert to using the framework as a numeric, 1 to 5 global rating scale without much attention to the behavioral anchors for each milestone. This is a problem that the milestones hope to avoid.⁶

This theoretic shift from the abstract to the concrete, from the competencies’ end-of-training checklist to the explicitly developmental model in the milestone framework, should be more congenial for teachers. In my institution, I have seen faculty quickly use our developmental model (the reporter-interpreter-manager-educator [RIME] framework) as a scaffold for milestones. Why is this facility important? All educational frameworks (the general competencies, CanMEDs, RIME, or others) propose to shape instructional methods and assessment, but some approaches come at a higher price.

The time it takes to learn and use a framework with consistency is a critical, though poorly studied, factor in our consideration of the utility of assessments. These practical matters in the use of assessment frameworks have been termed their *secondary effects*⁷ and cannot be ignored, any more than cost and risk can be ignored in judging clinical interventions. The milestones may have unintended consequences, and we have to be frank about discussing the cost and resources needed, especially for faculty development.

Williams and colleagues,⁹ among others, have asked whether milestones are actually a global rating form in disguise and, therefore, subject to cursory completion by faculty. These authors conjecture that this approach will increase the workload of faculty but will not provide new and useful information about residents’ competence. My argument is that since milestones are conceptually superior to global, 1 to 9 rating scales, and are synthetic and explicit, they will have an intuitive feel, which may well influence their acceptability to faculty.

The fact is that we don’t *know* yet that milestones will be easier to use, and more important, how much effort will have to go into calibrating the faculty. This is critical. It may even be that a major benefit of the milestone initiative is not in the vocabulary and theory of milestones, but in the effects of the efforts needed to train faculty in how to use them. Green and Holmboe¹⁰ remind us that the real advance needed is not another evaluation form, but more consistent use of a good form, by faculty. I ask myself: Could we have achieved the ACGME goal of public demonstration without something like the milestones? My answer is “probably not.” The competencies alone were

too abstract. The milestones are written for teachers, but we need a way to incentivize faculty, and we need sustainable mechanisms to calibrate their use.¹¹

The deeper question is not whether the milestones are adequate as an assessment framework, or even whether empirical evidence will show that they can be applied with some consistency. Rather, the key question is how will they fit into an assessment system? The milestones are simply one tool within the larger NAS, and should eventually be refined in that context.

As to a Third Cheer

What would it take to make it “three cheers” for the milestones? Initially, some accommodation in their use is needed for those who teach in the clinical workplace and who are subject to many social, economic, and professional pressures. The decrease in the number of Internal Medicine Milestones from well over 100 to approximately 22 reporting milestones was a good step. The ACGME is aware that educational research has shown that elaborate frameworks for assessment often collapse into something simpler.⁶ Studies using factor analysis show that we may be dealing with just 2 general domains—cognitive and noncognitive (or knowledge and interpersonal skills, or expertise and duty).

For internal medicine, the 22 subcompetencies of the 6 general competencies are now framed as 22 developmental milestones. This poses a “sampling problem”: Is each resident’s performance in each milestone to be documented every 6 months, or could there be an optimal time to expend the faculty’s time (eg, when some are critical for feedback, and others are critical for advancement)?

Finally, is there a commitment to the process of educational epidemiology to generate the generalizable knowledge about the milestones and how they work? Where is the commitment to secure funding for multisite studies that can generate an applicable evidence base? (I would be very happy to see a tiny percentage of the Medicare Indirect Medical Education supplements for academic health centers go to educational research.) Or, is this process actually occurring through the “donated” time

of individual program directors and faculty? To what degree is a manufacturing-like standardization of outcomes a cost-effective goal (ie, that a surgery resident in California is demonstrably the same as 1 from Georgia)?

To give the third cheer, I’d like a little more clarity about some of the assumptions underneath the effort. Is there any risk in the implementation of milestones? How do we factor in the time of residents, teachers, and program directors, and could the milestone effort turn out to be a major distraction from the real effort of faculty development and resident education? Do the usual “protection of human subjects” rules apply to such research?

In summary, we should be willing to stipulate, for now, that the goals of the milestone project are consistent with our professional duty, and that developing the expertise to improve their efficacy and minimize their cost must be a priority as well.

References

- 1 Accreditation Council for Graduate Medical Education. Milestones. <https://www.acgme.org/acgmeweb/tabid/430/ProgramandInstitutionalAccreditation/NextAccreditationSystem/Milestones.aspx>. Accessed November 10, 2014.
- 2 Wenger E, McDermott R, Snyder WM. *Cultivating Communities of Practice*. 1st ed. Cambridge, MA: Harvard Business Press; 2002.
- 3 Nasca TJ, Philibert I, Brigham T, Flynn TC. The next GME accreditation system—rationale and benefits. *N Engl J Med*. 2012;366(11):1051–1056.
- 4 Pellegrino ED. Toward a virtue-based normative ethics for the health professions. *Kennedy Inst Ethics J*. 1995;5(3):253–277.
- 5 Swing SR, Beeson MS, Carraccio C, Coburn M, Lobst W, Selden NR, et al. Educational Milestone development in the first 7 specialties to enter the next accreditation system. *J Grad Med Educ*. 2013;5(1):98–106.
- 6 Silber CG, Nasca TJ, Paskin DL, Eiger G, Robeson M, Veloski JJ. Do global rating forms enable program directors to assess the ACGME competencies? *Acad Med*. 2004;79(6):549–556.
- 7 Frank JR. *The CanMEDS 2005 Physician Competency Framework: Better Standards, Better Physicians, Better Care*. Ottawa, ON, Canada: Royal College of Physicians and Surgeons in Canada; 2005. http://fhs.mcmaster.ca/pathres/resident_resources/documents/CanMEDS2005_e.pdf. Accessed November 21, 2014.
- 8 Pangaro L, ten Cate O. Frameworks for learner assessment in medicine: AMEE Guide No. 78. *Med Teach*. 2013;35(6):e1197–e210.
- 9 Williams RG, Dunnington GL, Mellinger JD, Klamen DL. Placing constraints on the use of the ACGME Milestones: a commentary on the limitations of global performance ratings. *Acad Med*. 2014 Oct 7 [Epub ahead of print].
- 10 Green ML, Holmboe E. Perspective: the ACGME toolbox: half empty or half full? *Acad Med*. 2010;85(5):787–790.
- 11 Pangaro LN, Holmboe ES. Evaluation forms and formal rating scales. In: Holmboe ES, Hawkins RE. *Practical Guide to the Evaluation of Clinical Competence*. Amsterdam, the Netherlands: Elsevier; 2008:24–41.