

Reflections on the First 2 Years of Milestone Implementation

Eric S. Holmboe, MD, FACP, FRCP
Kenji Yamazaki, PhD
Laura Edgar, EdD, CAE
Lisa Conforti, MPH

Nicholas Yaghmour, MPP
Rebecca S. Miller, MS
Stanley J. Hamstra, PhD

Editor's Note: The ACGME News and Views section of JGME includes data reports, updates, and perspectives from the ACGME and its review committees. The decision to publish the article is made by the ACGME.

Background

The Accreditation Council for Graduate Medical Education (ACGME) and the American Board of Medical Specialties (ABMS) collectively constitute the foundation of professional self-regulation in the United States. In February 1999, the 2 organizations approved 6 general competencies broadly relevant for all medical practice, followed by the official launch of the Outcomes Project in 2001.¹ It was expected that the competencies would be an antidote to over-specification of accreditation standards, and that they would empower programs to create training programs grounded in meaningful outcomes in a developmental approach.¹

As many programs can attest, the implementation of outcomes-based (eg, competency-based) medical education has been challenging. One reason has been the difficulty in implementing the competencies in both curriculum and assessment. Program leaders lacked shared mental models within their own training programs, accompanied by a lack of shared understanding nationally within disciplines. It is important to remember that 1 of the thorny problems the milestones were intended to address was the sources of unwanted and unwarranted variability in educational and, by extension, clinical outcomes. In addition, the community cannot improve at scale what cannot be measured, and prior frames and approaches to measurement were insufficient and ineffective. A key goal for milestones thus is to help improve the state and quality of measurement through better assessment in graduate medical education to facilitate the improved outcomes everyone desires.

Approximately 10 years ago, conversations began on how to more effectively and meaningfully operationalize the competencies to help improve the design of residency and fellowship programs through the use of a developmental framework. In parallel, the ACGME began to explore mechanisms to move the accreditation system to a focus on outcomes using a continuous quality improvement philosophy.² Developmental milestones, using narratives to describe in more descriptive terms the professional trajectories of residents, were seen as a way to move the outcomes project forward.^{3,4} Starting in 2007, the disciplines of internal medicine, pediatrics, and surgery began to create developmental milestones for the 6 competencies.⁴⁻⁶

Surgery would subsequently delay the development of their milestones focusing first on the SCORE curriculum.⁷ The ACGME began to restructure its accreditation processes in 2009, and soon after, milestone groups were constituted for all specialties. Milestone writing groups were cosponsored by the ACGME and the ABMS member certification boards.⁴ Early groups had significant latitude in developing their subcompetencies and milestones; specialties that started the process after 2010 used a standard template. Each milestone set was subjected to review by the educational community in the specialty. BOX 1 provides an overview of the purposes of the milestones across key stakeholders, and FIGURE 1 provides an example of a key driver diagram of milestones as an educational and clinical intervention. As FIGURE 1 highlights, milestones can potentially trigger a number of drivers, or mechanisms, to help enable changes in residency and fellowship education.

In 2013, the milestones were officially launched in 7 core specialties (emergency medicine, internal medicine, neurological surgery, orthopaedic surgery, pediatrics, diagnostic radiology, and urology) as a formative, continuous quality improvement component of the new accreditation system.⁴ The remaining core disciplines and the majority of subspecialties implemented the milestones starting in July 2014. We have now reached an important “milestone” in the implementation process, and our commentary pro-

DOI: <http://dx.doi.org/10.4300/JGME-07-03-43>

vides a high-level overview of the first 2 years of the milestone experience, including information from the 2 most recent reporting cycles, and a description of what is next in the evaluation of the milestone initiative.

Milestones and Assessment in the NAS

FIGURE 2 provides an overview of how the milestones inform the graduate medical education system. At the program level, individual residents and fellows are assessed routinely through a combination of assessment tools, including direct observations; global evaluation; audits and review of clinical performance data; multisource feedback from team members, including peers, nurses, patients, and family; simulation; in-service training examination (ITE); self-assessment; and others. Assessment tools should be selected intentionally to allow routine, frequent, formative feedback to the resident or fellow to affirm areas of successful performance and to highlight competencies they need to improve.⁸ The clinical competency committee (CCC) should help to analyze and synthesize the assessment data, such as “quantitative” information from in-service examinations and clinical performance audits, as well as “qualitative” information from faculty, peers, and other raters through surveys and direct observation. Using the milestones, the CCC should reach a consensus judgment regarding each resident’s or fellow’s performance.⁹ The CCC provides those conclusions to the program director, who has the ultimate authority to determine residents’ or fellows’ milestone developmental level at least twice yearly. Milestones are used as a guiding framework and “blueprint” for individual learner performance and, aggregated to the program level, to assess the effectiveness of the curriculum and learning experiences.⁹

For the ACGME, the unit of analysis is the program, and this process uses the national data as a mechanism to help improve training overall. Collectively, the goal of this system is to help the entire medical education enterprise be accountable to the public for honest assessments of resident and fellow performance, and for truthful verification of their readiness to progress to unsupervised practice. As shown in FIGURE 2, while the ACGME is involved with the certification boards around research on the effectiveness of the milestones, milestone data are not used to determine eligibility for certification by the boards.

Box 1 Purposes of the Milestones

- 1) Training Programs
 - Guide curriculum development
 - Provide explicit expectations for learners
 - Support better assessment of learners and program
 - Provide framework for clinical competency committee deliberations
 - Enhance opportunities for early identification of underperforming learners so as to support early intervention
- 2) Residents and Fellows
 - Increase transparency of performance requirements in training
 - Encourage informed self-assessment and self-directed learning
 - Facilitate better feedback from program and faculty
 - Guide personal action plans for improvement
- 3) Accreditation Council for Graduate Medical Education
 - Support continuous monitoring and improvement of programs; lengthening of site visit cycles
 - Strengthen public accountability of national graduate medical education system through reporting at a national level on competency outcomes
 - Support community of practice for evaluation and research, with a focus on continuous improvement
- 4) Certification Boards
 - Support better assessment in residency and fellowship
 - Support research in graduate medical education innovation

Early Reporting Experience

Participation in milestone-based assessment and reporting obviously is critical to the long-term success of the milestone component of the NAS. Without robust reporting, meaningful feedback to the specialties and evaluation research is not possible, and lack of participation might send a negative signal to policy makers and the public about viability in graduate medical education self-regulation. The good news is that reporting has been very robust, with data capture across the 4 milestone cycles to date reaching 99% to 100%. For the 2014–2015 academic year, 7498 programs reported on 117 548 residents and fellows at midyear (99.9%) and 7628 programs reported on 118 360 residents and fellows at end-of-year reporting (99.9%). Between the 2 reporting periods, data were lacking for just 31 residents and fellows. For the first time, the US graduate medical education system has formative national data to guide assessment and curricular

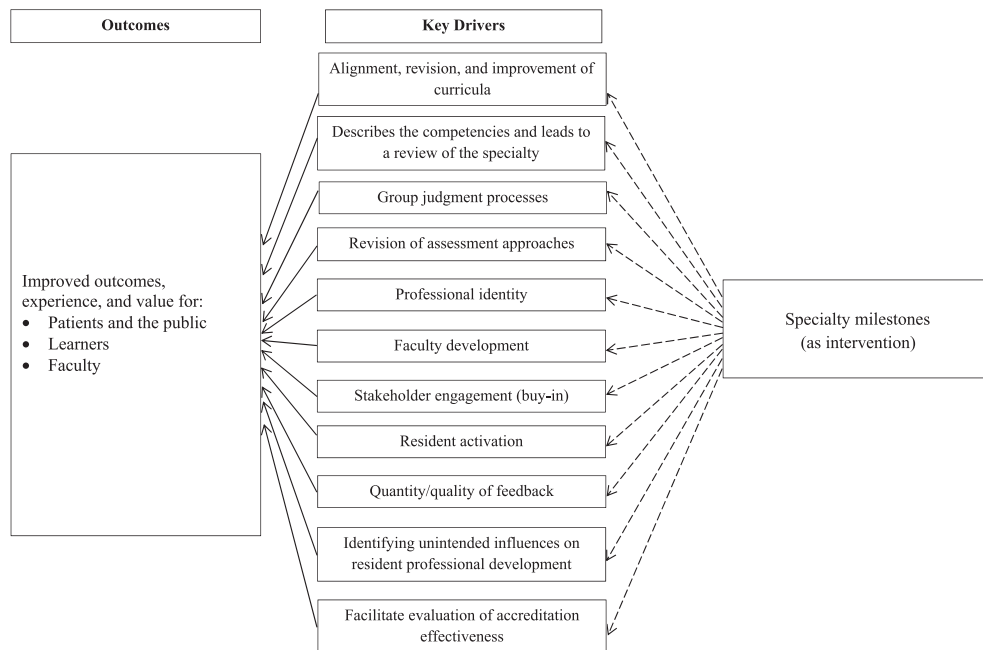


FIGURE 1
Key Drivers Diagram for Milestone Initiative

innovation and change, and as noted below, this is already happening in some specialties.

Early Signals From the Literature

While it is too early to perform a systematic review, several studies on the early experiences with milestones are worth noting as they provide a lens into needed ongoing evaluation research. One of the first national studies to find evidence of validity involved the first-year experience with the Emergency Medicine (EM) Milestones. This study examined reliability and milestone judgment distributions by training year across all emergency residency programs.¹⁰ An earlier mixed methods study involving program directors

from 17 internal medicine programs found the milestones to be useful for formative assessment, but faculty development was recognized as an important need to operationalize the milestones.¹¹ On the other hand, a group of internal medicine programs found only modest differences in perceived quality of feedback by residents after implementation of the milestone system.¹²

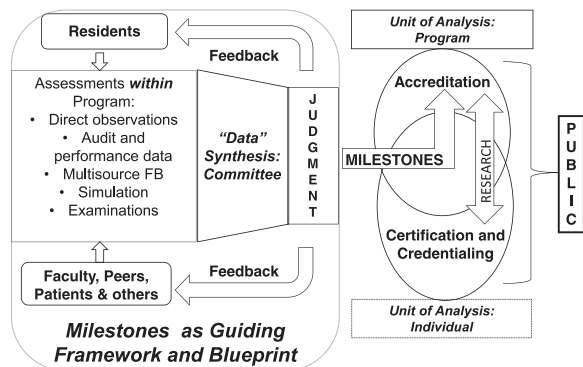


FIGURE 2
Milestones and the Assessment System

BOX 2 Key Issues for Milestone Validity

1. Content
 - Dependent on quality of milestone language developed by each specialty
2. Response Processes
 - Faculty rating process and understanding of the milestone language
 - How are faculty prepared to use the milestones and associated assessment instruments?
3. Internal Structure
 - What is the intrareliability of the clinical competency committee judgments?
4. Relations With Other Variables
 - Correlations with board scores, patient outcomes and experience surveys, registries, safety measures
5. Consequences
 - Understanding the needs of the various stakeholder groups and the manner in which milestone data might be interpreted by these different audiences

Regarding single institution studies, 1 program found the implementation of the first set of Internal Medicine Milestones improved faculty evaluations and feedback.¹³ Another study in a large internal medicine program found that transitioning to a milestone-based model produced a larger separation in the scores between postgraduate years (PGY) 1 to 3 and a wider use of a 5-point scale on an end-of-rotation evaluation form.¹⁴ Two studies determined that use of milestones was more effective than use of previous evaluation forms, and found better discrimination in ratings and a reduction in common rater errors.^{15,16} On the other hand, a study of a milestone “passport” intervention in an emergency medicine program found only modest increases in resident satisfaction with feedback.¹⁷ Another study reported that milestone-based assessments for end-of-shift evaluations led to grade inflation in an emergency medicine program.¹⁸ Using information technology is an additional growing theme of milestone research. For example, a surgery program is using a smartphone application to complete a Zwisch scale immediately after a procedure and linked this to milestones.¹⁹ The Foundation for Excellence in Women’s Health Care has also built mobile assessment tools for milestones, and the work is ongoing.²⁰ Collectively, these studies provide “early signals” and highlight the critical importance of ongoing, iterative, and rigorous research of the milestone initiative. We are truly only at the very beginning.

Next Steps for the Milestones

Now that the majority of specialties have completed their first year of implementation, the vital work of evaluating the milestones is picking up momentum. The milestones are not without their critics and concerns, and this early presumptive feedback will be important in framing the evaluation activities of the milestones.^{21–23} Evaluating the milestones will be a complex enterprise because in many respects the milestones represent a complex intervention. While there are a number of definitions of what constitutes a complex intervention, the milestones and the NAS meet a number of criteria for complexity:^{24,25}

- Consist of many interdependent and interactive component parts or activities that can behave nonlinearly (eg, small changes can lead to large effects)
 - For example, multiple assessment methods and tools are needed to inform a judgment on the milestones

- Display properties of emergence (the outcomes are not always predictable in advance; both positive and negative unintended consequences are probable)
- Depend on various mechanisms that act in context to produce an outcome (eg, how the milestones are implemented is critical to their impact on outcomes)
- Is highly sensitive to context (eg, availability and nature of patients and populations served by the training program, faculty expertise and motivation, infrastructural resources)

Any evaluation strategy will have to attend to these aspects of milestones and will require a mixed methods, comprehensive approach. From the quantitative (or psychometric) perspective, the use of a validity framework is crucial. Looking through the lens of the Messick validity framework as an example, there is a set of key issues for making inferences about the validity of the milestones (BOX 2).²⁶

These are the areas in the validity domain that are beginning to inform the research agenda, as exemplified by the national EM Milestones study. However, a purely psychometric point of view will be insufficient in evaluating and understanding the milestones. Furthermore, milestones are not “static”; as programs continue to work with them their understanding of them will change and, in turn, evaluative judgments and curriculum will also change and evolve. The ACGME signaled from the beginning that the current milestones are “version 1.0,” and with learning will come the need for revisions down the road.²⁷ Therefore, evaluation of the milestones will also utilize lessons and guidance from the program evaluation field on evaluating complex interventions.²⁵ Instead of just looking at milestones through an attribution lens, examining how milestones contribute to an outcome will be crucial.²⁸

Too often we treat educational interventions and innovations as “therapeutic interventions” (eg, pills) that if taken properly and in the appropriate dose will produce a desired outcome. This biomedical model, traditionally focused on attribution (cause and effect question: Did the milestones *cause* the resident or fellow to be better in X competency[ies]?), has dominated research discourse in medical education for decades. However, milestones represent an education intervention within a program (ie, embedded within the residency or fellowship curriculum) consisting of multiple interacting and interdependent components. Treating a complex programmatic intervention, such as a residency program and milestones,

as a medical procedure or pill will be insufficient to address the complexity of this intervention.^{25,28,29}

Newer programmatic evaluation models have increasingly moved away from purely cause-effect, linear models that heretofore were mostly concerned with making sure the inputs of an intervention were clear, standardized, and randomly assigned to subjects, and the outcomes were clearly defined, rigorously measurable, and meaningful. Much of the implementation activity in-between was essentially a “black box,” managed through randomization to ensure a determination of a mean effect. This model, however, has many shortcomings in evaluating complex interventions. First, the interactions, interdependencies, context, and quality of the implementation can and do have large effects on outcomes. Failure to understand these aspects of the intervention can lead to misguided conclusions about effect and generalizability. While it is beyond the scope of this article to cover all the program evaluation strategies available to assess complex interventions, a few concepts warrant mention.

Investigators studying the milestones should ask the fundamental questions: What works for whom, in what circumstances, and why? These questions form the core of realistic program evaluation strategy by Pawson and Tilley and other program evaluation strategies that emphasize the need to look deep into the “black box” of implementation to understand the mechanisms of a specific intervention and how context affects the success or failure of the intervention.^{25,30} The concept of “partial solutions” as a major aspect of program interventions is also important. As Pawson points out, no intervention, however complex and comprehensive, is ever a complete solution to a problem or need.

Some components will work better than others, and the key issue is to determine why so as to learn how to improve the next iteration of the intervention.²⁵ Without this understanding of what works, for whom, and in what circumstances, it will be very hard to generalize lessons from milestone and graduate medical education research in a single site or a small group of programs to a national cohort. Furthermore, “failures” can be rich sources of learning that can be fed forward into the iterative cycle of milestone and residency program improvement and development.

The second concept moves programmatic evaluation away from a sole focus on attribution (just cause and effect) to one of contribution. The fundamental question in a contribution analysis is “How much of a difference (or contribution) has the program made to the observed outcomes?”²⁸ Central to all complex program evaluation strategies is developing a theory

of change of how each component of the intervention *contributes* to the outcome, including interactions with the other components. The key driver diagram (FIGURE 1) provides some, but likely not all, of the possible ways milestones can effect change in programs. In this case, a theory of change really describes the hypothesized pathway to the desired outcome. The goal is to create a robust and credible contribution story. Building on the realist questions, for example, the story should describe how the interventions did or did not trigger the intended mechanisms, how well the interventions were implemented and functioned in specific contexts, and how, using the best evidence available, the intervention contributed to the outcomes measured. By now you have likely realized no single research method will likely be sufficient. Mixed method qualitative and quantitative methods will be needed. The specific methods will depend on the questions and outcomes.

Conclusions

The ACGME milestones are intended to describe the educational and professional trajectory of a resident or fellow from the beginning of their education and training through the achievement of competency and the ability to enter into the unsupervised practice of medicine. The milestones are also designed to help address the thorny problem of better addressing and identifying the sources of unwanted and unwarranted variability in educational, and by extension, clinical outcomes. Furthermore, prior frames and approaches to measurement have been ineffective and insufficient.

A key goal is to improve the quality of curricula and assessment to facilitate the improved outcomes everyone in the graduate medical education system desires. This year will mark the third year of milestone reporting for the first 7 core specialties and the beginning of reporting for the remaining subspecialties. Looking at the first 2 years of implementation, ongoing research, and new research being proposed, much more will be learned about the milestones, including how they should be used in programs, their effect on residents and fellows, and how they will improve graduate medical education.

References

1. 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.

2. 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.
3. Swing SR. The ACGME outcome project: retrospective and prospective. *Med Teach*. 2007;29(7):648–654.
4. Swing SR, Beeson MS, Carraccio C, Coburn M, Iobst 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.
5. Green ML, Aagaard EM, Caverzagie KJ, Chick DA, Holmboe ES, 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.
6. Schumacher DJ, Lewis KO, Burke AE, Smith ML, Schumacher JB, Pitman MA, et al. The pediatrics milestones: initial evidence for their use as learning road maps for residents. *Acad Pediatr*. 2013;13(1):40–47.
7. Surgical Council on Residency Education. SCORE Curriculum 2014–2015. <http://www.absurgery.org/xfer/curriculumoutline2014-15.pdf>. Accessed July 15, 2015.
8. Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR. The role of assessment in competency-based medical education. *Med Teach*. 2010;32(8):676–682.
9. Andolsek K, Padmore J, Hauer K, Holmboe ES. Clinical Competency Committees. A guidebook for programs. <http://www.acgme.org/acgmeweb/Portals/0/ACGMEClinicalCompetencyCommitteeGuidebook.pdf>. Accessed July 15, 2015.
10. Beeson M, Holmboe E, Korte R, Nasca T, Brigham T, Russ C, et al. Initial validity analysis of the emergency medicine milestones. *Acad Emerg Med*. 2015;22(7):838–844.
11. Aagaard E, Kane GC, Conforti L, Hood S, Caverzagie KJ, Smith C, et al. Early feedback on the use of the internal medicine reporting milestones in assessment of resident performance. *J Grad Med Educ*. 2013;5(3):433–438.
12. Angus S, Moriarty J, Nardino RJ, Chmielewski A, Rosenblum MJ. Internal medicine residents' perspectives on receiving feedback in milestone format. *J Grad Med Educ*. 2015;7(2):220–224.
13. Nabors C, Peterson SJ, Forman L, Stallings GW, Mumtaz A, Sule S, et al. Operationalizing the internal medicine milestones—an early status report. *J Grad Med Educ*. 2013;5(1):130–137.
14. Friedman KA, Balwan S, Cacace F, Katona K, Sunday S, Chaudhry S. Impact on house staff evaluation scores when changing from a Dreyfus- to a Milestone-based evaluation model: one internal medicine residency program's findings. *Med Educ Online*. 2014;19:25185.
15. Bartlett KW, Whicker SA, Bookman J, Narayan AP, Staples BB, Hering H, et al. Milestone-based ratings are superior to Likert-type assessments in illustrating trainee progression. *J Grad Med Educ*. 2015;7(1):75–80.
16. Raj JM, Thorn PM. A faculty development program to reduce rater error of milestones-based assessments. *J Grad Med Educ*. 2014;6(4):680–685.
17. Yarris LM, Jones D, Kornegay JG, Hansen M. The milestones passport: a learner centered application of the Milestone framework to prompt real-time feedback in the emergency department. *J Grad Med Educ*. 2014;6(3):555–560.
18. Dehon E, Jones J, Puskarich M, Sandifer JP, Sikes K. Use of emergency medicine milestones as items on end-of-shift evaluations results in overestimates of residents' proficiency level. *J Grad Med Educ*. 2015;7(2):192–196.
19. George BC, Teitelbaum EN, Meyerson SL, Schuller MC, DaRosa DA, Petrusa ER, et al. Reliability, validity, and feasibility of the Zwisch scale for the assessment of intraoperative performance. *J Surg Educ*. 2014;71(6):e90–e96.
20. Foundation for Excellence in Women's Healthcare. MyTIPReport. <https://mytipreport.org>. Accessed July 15, 2015.
21. Norman G, Norcini J, Bordage G. Competency-based education: milestones or millstones? *J Grad Med Educ*. 2014;6(1):1–6.
22. Pangaro LN. Two cheers for milestones. *J Grad Med Educ*. 2015;7(1):4–6.
23. Dewan M, Manning J, Satish U. The new milestones: do we need to take a step back to go a mile forward? *Acad Psychiatry*. 2015;39(2):147–150.
24. Rogers PJ. Implications of complicated and complex characteristics for key tasks in evaluation. In: Forss K, Marra M, Schwartz R, eds. *Evaluating the Complex: Attribution, Contribution, and Beyond*. New Brunswick, NJ: Transaction Publishers; 2011:33–53.
25. Pawson R. *The Science of Evaluation: A Realist Manifesto*. London, UK: Sage Publications; 2013.
26. Messick S. Validity. In: Linn RL, ed. *Educational Measurement*. 3rd ed. New York, NY: Macmillan Publishers; 1989:13–103.
27. Philibert I, Brigham T, Edgar L, Swing S. Organization of the educational milestones for use in the assessment of educational outcomes. *J Grad Med Educ*. 2014;6(1):177–182.
28. Mayne J. Contribution analysis. In: Forss K, Marra M, Schwartz R, eds. *Evaluating the Complex: Attribution, Contribution, and Beyond*. New Brunswick, NJ: Transaction Publishers; 2011:53–97.
29. Medical Research Council. Developing and evaluating complex interventions: new guidance. 2008. <http://www.mrc.ac.uk/documents/pdf/complex-interventions-guidance>. Accessed July 15, 2015.
30. Pawson R, Tilley N. *Realistic Evaluation*. London, UK: Sage Publications; 1997.



All authors are with the Accreditation Council for Graduate Medical Education. **Eric S. Holmboe, MD, FACP, FRCP**, is Senior Vice President, Milestone Development and Evaluation; **Kenji Yamazaki, PhD**, is Outcome Assessment Project Associate; **Laura Edgar, EdD, CAE**, is Executive Director, Milestones Development; **Lisa Conforti, MPH**, is Research Associate for Milestones Evaluation; **Nicholas Yaghmour, MPP**, is Research Associate for

Milestones Evaluation; **Rebecca S. Miller, MS**, is Senior Vice President, Applications and Data Analysis; and **Stanley J. Hamstra, PhD**, is Vice President, Milestone Research and Evaluation.

Corresponding author: Eric S. Holmboe, MD, FACP, FRCP, Accreditation Council for Graduate Medical Education, 515 N State Street, Suite 2000, Chicago, IL 60654, 312.755.5087, eholmboe@acgme.org