Early Feedback on the Use of the Internal Medicine Reporting Milestones in Assessment of Resident Performance

Eva Aagaard, MD
Gregory C. Kane, MD
Lisa Conforti, MPH
Sarah Hood, MS
Kelly J. Caverzagie, MD
Cynthia Smith, MD
Davoren A. Chick, MD
Eric S. Holmboe, MD
William F. Iobst, MD

Abstract

Background The educational milestones were designed as a criterion-based framework for assessing resident progression on the 6 Accreditation Council for Graduate Medical Education competencies.

Objective We obtained feedback on, and assessed the construct validity and perceived feasibility and utility of, draft Internal Medicine Milestones for Patient Care and Systems-Based Practice.

Methods All participants in our mixed-methods study were members of competency committees in internal medicine residency programs. An initial survey assessed participant and program demographics; focus groups obtained feedback on the draft milestones and explored their perceived utility in resident assessment, and an exit survey elicited input on the value of the draft milestones in resident assessment. Surveys were tabulated using descriptive statistics. Conventional content analysis method was used to assess the focus group data.

Results Thirty-four participants from 17 programs completed surveys and participated in 1 of 6 focus groups. Overall, the milestones were perceived as useful in formative and summative assessment of residents. Participants raised concerns about the length and complexity of some draft milestones and suggested specific changes. The focus groups also identified a need for faculty development. In the exit survey, most participants agreed that the Patient Care and Systems-Based Practice Milestones would help competency committees assess trainee progress toward independent practice.

Conclusions Draft reporting milestones for 2 competencies demonstrated significant construct validity in both the content and response process and the perceived utility for the assessment of resident performance. To ensure success, additional feedback from the internal medicine community and faculty development will be necessary.

Eva Aagaard, MD, is Associate Professor of Medicine, Assistant Dean for Lifelong Learning, Director of the Academy of Medical Educators, Director of the Center for Advancing Professional Excellence, and Director of Faculty Development in General Internal Medicine at the University of Colorado; **Gregory C. Kane, MD,** is Professor and Interim Chairman of Medicine at Jefferson Medical College; **Lisa Conforti, MPH,** is Research Associate for Academic Programs at the American Board of Internal Medicine; **Sarah Hood, MS,** is Director of Academic Affairs at the American Board of Internal Medicine; **Kelly J. Caverzagie, MD,** is Assistant Professor and Associate Vice-Chair for Quality and Physician Competence in Internal Medicine at the University of Nebraska Medical Center; **Cynthia Smith, MD,** is Adjunct Associate Professor of Medicine at the University of Pennsylvania School of Medicine and Senior Medical Associate for Content Development in Medical Education at the American College of Physicians; **Davoren A. Chick, MD,** is Clinical Assistant Professor in Internal Medicine at the University of Michigan Medical School; **Eric S. Holmboe, MD,** is Chief Medical Officer and Senior Vice President at the American Board of Internal Medicine; and **William F. Iobst, MD,** is Vice President for Academic Affairs at the American Board of Internal Medicine.

Funding: The authors report no external funding source for this study.

Corresponding author: Eva Aagaard, MD, Division of General Internal Medicine, University of Colorado, Mail Stop F523, Room 4102A, 13070 East 19th Avenue, Aurora, CO 80045, 303.724.5866, eva.aagaard@ucdenver.edu

Received December 30, 2012; revision received March 22, 2013; accepted April 1, 2013.

DOI: http://dx.doi.org/10.4300/JGME-D-13-00001.1

Editor's Note: The online version of this article contains an interview guide and participant comment form.

Introduction

In November 2007, the American Board of Internal Medicine (ABIM) and the Accreditation Council for Graduate Medical Education (ACGME) convened a 33-member task force consisting of internal medicine program directors, experts in evaluation, and representatives of internal medicine stakeholder organizations to identify the milestones of competency development for internal medicine residents. The result of this work was a list of 142 discreet, observable behaviors that were divided across the 6 ACGME competency domains.[1] A subgroup of the original task force has recently transformed those behaviors into synthesized reporting milestones, which will be the means by which internal medicine residency programs attest to the progress of their individual trainees in the Next

Accreditation System. The purpose of developing the reporting milestones (hereafter called *milestones*) was to provide learners, educators, programs, regulators, and the public with a criterion-based framework of trainee progression across the 6 ACGME competencies that could be used for assessment and training.[2,3] This article presents the results of a mixed-methods study assessing the construct validity,[4] perceived utility, and perceived feasibility of 2 of those draft milestones.

What was known

The educational milestones were designed as a criterion-based framework to assess resident progression on the 6 Accreditation Council for Graduate Medical Education competencies.

What is new

Internal medicine program leaders provided feedback on the construct validity, feasibility, and utility of the educational milestones for patient care and systems-based practice.

Limitations

Pilot study with limited sample size and potential for selection bias, reducing generalizability.

Bottom line

The current milestones for patient care and systems-based practice show construct validity and perceived utility for assessing resident performance.

Methods

The leadership from 50 internal medicine residency programs in 2 metropolitan areas (Philadelphia, PA, and Chicago, IL) of the United States were e-mailed and invited to identify up to 4 members of their program's competency or equivalent committee to participate in a focus group, with the goal of refining the milestones and providing feedback on their utility. A competency committee was defined as the committee charged with assessing the clinical competence of residents in that program. These individuals were felt to be uniquely suited to offer comment on the value of the milestones framework for attesting to resident competence. Participants were enrolled on a first-come, first-served basis and received a $250 per person honorarium for completion of the study. No preference was given to individuals or institutions with prior knowledge of, or positive predisposition toward, the milestones.

The study was approved by the American Institutes for Research Institutional Review Board.

Data Collection

Six focus groups (3 each [50%] in Philadelphia and Chicago) were completed in June 2011. Before the focus group, participants completed a brief online demographic survey detailing their age, sex, educational roles, academic rank, time spent teaching trainees, prior faculty-development experience, size and process of their institutions' competency committees, and their familiarity with the ACGME Milestones Project.

At the focus group, participants were given 15 minutes to review a sample of milestones in 2 of the ACGME competencies: (1) Patient Care and (2) Systems-Based Practice (BOXES 1 and 2). Note that these milestones are drafts (not the final milestones), and that the final milestones incorporated data from these focus groups as well as other input before publication.

One of 3 trained moderators facilitated each focus group using an interview guide. The guide was developed by the authors along with the ACGME and ABIM staff (provided as online supplemental material). Questions explored the structure of each participating institution's competency committee and posed a series of questions about the draft milestones. Each focus group lasted approximately 90 minutes, was audiotaped, and an independent note taker was present throughout.

At the end of each focus group, the moderator asked participants to complete a brief paper survey. This survey used a 5-point Likert scale (1 = strongly agree, 5 = strongly disagree) to assess the utility of the draft milestones for the assessments of trainees in the areas of Patient Care and Systems-Based Practice. Participants were also asked to make additional comments (provided as online supplemental material).

Data Analysis

We used a conventional content-analysis method to develop themes across the focus group data.[5] Following each focus group, the moderator and the note taker provided a summary of observed themes, as well as the notes and audio recording. The 3 focus group moderators and the note taker met halfway through the study to compare notes and refine the focus group guide. The moderators also conducted a final debrief at the conclusion of the study to share and confirm observations. A coordinator (L.C.) compiled the observations and key points and developed a summary document that identified major themes, referring back to notes and audio files when necessary. This summary document was shared with the moderators, and any discrepancies were resolved through consensus. Demographic survey data and post–focus group surveys were tabulated using descriptive statistics.

Results

Nineteen of 50 program directors (38%) responded to the initial request, with 39 potential individual participants identified. Of those 39 individuals, 34 (87%) participated in 1 of the focus groups, 1 (3%) participated in a one-on-one

Box 1 Draft From the Patient Care Milestones

Competency	Unacceptable	Early Learner	Advancing Performance	Ready for Unsupervised Practice	Aspirational
■ Gathering and synthesizing essential and accurate information to define each patient's clinical problem(s).	■ Displays no interest in collecting accurate and thorough historic data. ■ Does not use physical exam to confirm history. ■ Takes no personal responsibility for confirming accuracy of data.	■ Struggles to acquire accurate and thorough historic information in an organized way. ■ Does not always perform accurate physical exams and misses key history and physical exam findings. ■ Overly relies on secondary data.	■ Able to acquire accurate and relevant histories from patients in an efficient, prioritized, and hypothesis-driven fashion. ■ Able to seek and obtain data from secondary sources when needed. ■ Uses secondary data appropriately in the development of patient care plan. ■ Performs accurate physical exams that are targeted to the patient's complaints.	■ Able to obtain relevant historic subtleties that prioritize the differential diagnosis and include sensitive information that may not be volunteered by the patient. ■ Demonstrates and teaches historic data gathering and physical diagnosis to junior members of the team.	■ Able to be a role model in gathering subtle and difficult information from the patient. ■ Can identify subtle or unusual physical findings that may influence clinical decision making.

phone interview (data from that individual is not included here), and 4 (10%) dropped out (work-related and personal conflicts). Focus groups consisted of 4 to 9 participants, representing 17 separate training programs with each program represented by 1 (n = 7 [41%]), 2 (n = 5 [29%]), 3 (n = 3 [18%]), or 4 (n = 2 [12%]) individuals. Participants were between 29 and 55 years old, and 56% (19 of 34) were women. Most (56%, 19 of 34) were from university-based hospitals; 61% (21 of 34) were program directors or associate program directors, and 23% (8 of 34) were chief residents (TABLE).

Among the 17 programs represented, competency committees varied in size from 6 to 20 members. Although most members were physicians, other member types included nurse, nurse manager, nursing director, graduate medical education coordinator, pharmacist, and psychologist. Most committees met at least quarterly, but some met every month. Four of the 17 committees (24%) met only twice a year. Some of the larger programs used smaller committees with frequent meetings for routine deliberations, convening a larger, more formal committee twice annually or when more complex or serious decisions are necessary.

The focus groups identified several themes relative to assessment in internal medicine residency education and the use of the draft milestones.

Box 2 Draft From the Systems-Based Practice Milestones

Competency	Unacceptable	Early Learner to Advancing Performance	Ready for Unsupervised Practice	Aspirational
■ Works effectively within multiple health delivery systems.	■ Unable to identify the role for, or effectively use, various health delivery systems. ■ Communication with future caregivers is inefficient or inaccurate, requiring constant review by supervisor. ■ Frequent clarification and correction needed from other caregivers after transitions of care. ■ Is not entrusted with discharging a patient to another setting (ie, home care, skilled nursing) without direct supervision. ■ Does not respond to requests of caregivers in other delivery systems.	■ Understands basic roles of various health delivery systems. ■ Communication with future caregivers is generally acceptable, but sometimes requires clarification or correction by a supervisor or other team member. ■ Completes minimal responsibilities like discharging a patient from the hospital (ie, does just enough to get job done). ■ Delayed in responding to requests of caregivers in other delivery systems.	■ Appropriately uses other health delivery systems and personnel to provide safe and efficient care. ■ Communication with future caregivers is timely, complete, and purposeful with the goal of optimizing care. ■ Anticipates needs of future care providers and takes appropriate steps to mitigate those needs. ■ Entrusted with safely discharging a patient from the hospital without supervision. ■ Responds to requests of caregivers in other delivery systems.	■ Optimizes care across health delivery systems to reduce costs, increase patient safety, and reduce hospital readmissions. ■ Anticipates needs of patient, patient advocates, and future care providers and takes appropriate steps to mitigate those needs. ■ Consistently performs and models safe care transitions that are patient-centered. ■ Recommends appropriate use of resources to future care providers in other delivery systems. ■ Always available to caregivers in other delivery systems.

TABLE **Characteristics of the Focus Group Participants, n = 34**

Characteristics	Results
Mean age, y (range)	39 (29–55)
Sex, No. (%)	F: 19 (56); M: 15 (44)
Educational roles,[a] No. (%)	
Associate program director	14 (41)
Key/core faculty (as defined by ACGME)	11 (32)
Program director	7 (21)
Chief medical resident	8 (24)
Medical school rotation course director	6 (18)
Current academic rank, No. (%)	
Professor	2 (6)
Associate professor	6 (18)
Assistant professor	14 (41)
Instructor	5 (15)
Chief resident	4 (12)
Fellow	1 (3)
None, at a community hospital	1 (3)
No response	1 (3)
Affiliation, No. (%)	
University-based	19 (56)
Community-based	14 (41)
Both community- and university-based	1 (3)
Time on competency (or equivalently named) committee, y (%)	
<1	3 (9)
1	7 (21)
2–5	10 (29)
6–10	6 (18)
11–20	7 (21)
No response	1 (3)
Familiar with the Milestones Project, No. (%)	22 (65)

Abbreviations: F, female; M, male; ACGME, Accreditation Council for Graduate Medical Education.

[a] Some have multiple roles.

Theme 1: Value of the Milestones

In general, participants felt the milestones would be valuable in the assessment of trainees. Specifically, participants felt it would provide a shared language and set of expectations, help create objectivity in the assessment process, and be helpful to learners at multiple levels in both formative and summative feedback.

Provide a Shared Mental Model The focus group participants noted that they "struggle" with the ACGME general competency domains. Some voiced concern that some competency domains are "vague and open to interpretation." The milestones, on the other hand, provide "concrete language" and an appropriate representation of trainee "progression over time." There was consensus that

the milestones will help create a shared mental model that will help to clarify educational expectations for the learner as well as the faculty. The milestones were described as "a good guide" and a "good tool."

Behavioral Anchors Help With Discrimination and Create Objectivity There was broad consensus that the criteria provided in the milestones were more helpful than traditional, normative scales, ranging from unsatisfactory (1–3) to superior (7–9). Many agreed that faculty too often used the high end of such scales, limiting their ability to stratify residents. One attendee pointed out that the use of milestones by faculty would make the evaluation process "much clearer"; the milestones were also described as "more standardized" and "more analytic." Another believed that the milestones were more objective and less biased because of the anchoring to specific behaviors.

Helpful for All Learner Levels The focus groups provided insight into how the milestones might be used in the assessment of residents at all levels of achievement.

For the high performer, milestones could help identify the resident who would benefit from enhanced training through elective experiences or a "higher level of responsibility." In the third year of training, in particular, participants felt that the milestones could help high performers "move beyond the core" and graduate to research or advanced rotations within the training program. For the average resident, many felt the milestones would help them to identify things, such as "how can I be better" and "what to shoot for." For struggling residents, the milestones were viewed as having unique utility by making it "easier to identify a resident's gaps" (in skill or behaviors) and by making it easier to "verbalize" those deficiencies. One participant commented, "This would allow us to be on the same page for struggling residents," whereas another stated this would be "helpful to cut to the chase on what the problem is." There was consensus that struggling residents might be identified earlier.

Milestones Could Also Be Used for Formative Evaluation Several attendees speculated that by providing the milestones to the trainees, residents would gain a better understanding of the expectations of the training program. In addition, the milestones could help residents to develop individualized goals. Some participants noted that by providing faculty and trainees with specific examples of desired skills, the milestones would make it easier to "provide feedback that is specific," thereby allowing for individualized growth and development. The progressive nature of the milestones could also be helpful in monitoring individual trainees' progress year to year to demonstrate personal growth along the continuum toward unsupervised practice.

Theme 2: Challenges for Milestones

There were also concerns raised about the milestones.

Length and Complexity Some were concerned by the "size and complexity" of implementing such a new scheme. A few suggested that a more "truncated version" could be helpful in implementation. Many struggled with envisioning the feasibility of evaluating every resident on every milestone, citing time constraints and missing data as barriers.

Struggles With Aspects of the Milestones Themselves There were concerns raised with some of the language used and with the format of the milestones.

Some voiced concerns about the "unacceptable" label. Some trainees might be progressing but not meeting all milestones on schedule. These candidates would be different than those who show no growth or those that demonstrate clear and consistent unprofessional behavior. Only rarely does a resident perform in the unacceptable range, and participants felt that this delineation should be made clearer in future versions of the milestones.

It was also noted that the number of developmental stages for Patient Care and Systems-Based Practice differed across the spectrum, and some disliked that inconsistency.

Theme 3: Need for Faculty Development

There was strong consensus around the need for extensive faculty development to successfully introduce the use of milestones in assessment. Faculty who share a common understanding of the milestones and the assessment tools used to inform them will be crucial to ensuring that committees have the data they need to report to the ACGME in the Next Accreditation System.

Survey Data

Focus group themes were confirmed in the postgroup survey. Seventy-nine percent (27 of 34) of participants agreed or strongly agreed that the Patient Care, and 68% (23 of 34) agreed or strongly agreed that the Systems-Based Practice milestones, would help competency committees assess their trainees' progress toward independent practice.

Discussion

Our study demonstrated significant construct validity and general support for the utility of the draft milestones in resident assessment. The focus group results provide both content and response-process evidence of validity.[4] The milestones were felt to provide a common language of expectations for resident performance, which, along with the specific descriptions used in the milestones, was perceived to be helpful in both the formative and summative assessment of residents during their training. Our participants' perceptions were consistent with prior

studies demonstrating that criterion-based and narrative frameworks produce a broader range of score distributions, result in enhanced discrimination, identify an increased number of poor performers, result in higher interrater reliability, and increase the quality of evaluator comments when compared with traditional normative scales.[6–8]

Specific recommendations for changes to the milestones themselves were made, primarily related to improving the feasibility of implementation. These included concerns about the length and complexity of some of the draft milestones, a need to standardize the number of developmental stages for each assessed competency, and suggestions that we revise the unacceptable column within the milestones. Recommendations were subsequently incorporated into revisions of the milestones document, contributing to the final public format and content of the 22 milestones for internal medicine.[3,9]

Using milestones as a framework for competency assessment will require a paradigm shift. Faculty will need to embrace a new lexicon that includes sometimes confusing or variably defined terms, including *milestones*, *competencies*, and *entrustable professional activities*. Understanding and using key educational terminology is becoming essential to effective functioning as a program director. Faculty development will be critical to this process, not only to ensure appropriate understanding among program directors and their core faculty but also to create shared mental models that will subsequently help guide teaching, assessment, and feedback.[10] Such faculty development may need to embed concurrent competency training or retraining among the faculty.[11] Competency-based education will require faculty to provide increased frequency and quality of formative feedback for learners, with greater emphasis on direct observation of learner performance.[12] Faculty development, therefore, should aim to standardize criteria for rating competencies within rotations and programs and to clarify when specific behaviors must be achieved.[13] Additionally, faculty development should include strategies for how to do such assessments in a timely and efficient manner.

Despite faculty development efforts, the effect of these changes on faculty workload is likely to be significant; program directors and departments must consider workload in the development of any comprehensive assessment strategy. In addition, successful implementation will require the engagement and feedback of residents.[14] Early introduction of the milestones, clear descriptions of expectations for achievement at each developmental stage, and frequent feedback with mentored self-reflection are likely to enhance resident acceptance and enthusiasm for this change.

There are several limitations to our study. This was a small, pilot study involving programs in only 2 regions of the country. Program directors were somewhat more likely to be from university-based rather than community-based programs. Therefore, the results may not be generalizable to the larger population of program directors. Although our total number of participants was small, thematic analysis reached saturation quickly. Finally, this study assessed only construct validity and program director perceptions of utility and feasibility. We did not attempt to assess resident perception nor did we assess reliability or higher levels of validity.

Conclusion

Use of the milestones as an assessment framework holds substantial promise for advancing the current state of assessment in internal medicine training. Participants believed that the milestones were helpful in developing a shared mental model both among educators and between educators and residents. Feedback from our participants included suggestions for improved clarity and feasibility of drafted milestones, which were incorporated into the publicly released milestones document.[9] Feedback highlighted the need for faculty development to effectively implement milestones. Future studies should assess the true reliability and validity of this tool in the assessment of residents.

References

1 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.
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 Caverzagie KJ, Iobst WF, Aagaard EM, Hood S, Chick DA, Kane GC. The internal medicine reporting milestones and the next accreditation system. *Ann Intern Med*. 2013;158(7):557–559.
4 Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments. *Am J Med*. 2006;119(2):166e7–166e16. doi:10.1016/j.amjmed.2005.10.036/
5 Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. *Qual Health Res*. 2005;15(9):1277–1288.
6 Regher G, Regehr C, Bogo M, Power R. Can we build a better mousetrap? improving the measures of practice performance in the field practicum. *J Soc Work Educ*. 2007;43(2);327–343.
7 Speer, AJ, Solomon DJ, Ainsworth MA. An innovative evaluation method in an internal medicine clerkship. *Acad Med*. 1996;71(suppl 1):76–78.
8 Crossley J, Johnson G, Booth J, Wade W. Good questions, good answers: construct alignment improves the performance of workplace-based assessment scales. *Med Educ*. 2011;45(6):560–569.
9 Accreditation Council for Graduate Medical Education; American Board of Internal Medicine. The Internal Medicine Milestone Project. http://www.acgme-nas.org/assets/pdf/Milestones/InternalMedicineMilestones.pdf. Accessed December 29, 2012.
10 Dath D, Iobst W. The importance of faculty development in the transition to competency-based medical education. *Med Teach*. 2010;32(8):683–686.
11 Kogan JR, Conforti L, Bernabeo E, Iobst W, Holmboe E. Opening the black box of clinical skills assessment via observation: a conceptual model. *Med Educ*. 2011;45(10):1048–1060.
12 Carraccio C, Wolfsthal SD, Englander R, Ferentz K, Martin C. Shifting paradigms: from Flexner to competencies. *Acad Med*. 2002;77(5):361–367.
13 Holmboe ES. Assessment of the practicing physician: challenges and opportunities. *J Contin Educ Health Prof*. 2008;28(suppl 1):4–10.
14 Iobst WF, Sherbino J, Cate OT, Richardson DL, Dath D, Swing SR, et al. Competency-based medical education in postgraduate medical education. *Med Teach*. 2010;32(8):651–656.