

Educational Milestone Development in the First 7 Specialties to Enter the Next Accreditation System

SUSAN R. SWING, PhD
MICHAEL S. BEESON, MD, MBA
CAROL CARRACCIO, MD, MA
MICHAEL COBURN, MD
WILLIAM LOBST, MD
NATHAN R. SELDEN, MD, PhD
PETER J. STERN, MD
KAY VYDARENY, MD

Abstract

Background The Accreditation Council for Graduate Medical Education (ACGME) Outcome Project introduced 6 general competencies relevant to medical practice but fell short of its goal to create a robust assessment system that would allow program accreditation based on outcomes. In response, the ACGME, the specialty boards, and other stakeholders collaborated to develop educational milestones, observable steps in residents' professional development that describe progress from entry to graduation and beyond.

Objectives We summarize the development of the milestones, focusing on 7 specialties, moving to the next accreditation system in July 2013, and offer evidence of their validity.

Methods Specialty workgroups with broad representation used a 5-level developmental framework

and incorporated information from literature reviews, specialty curricula, dialogue with constituents, and pilot testing.

Results The workgroups produced richly diverse sets of milestones that reflect the community's consideration of attributes of competence relevant to practice in the given specialty. Both their development process and the milestones themselves establish a validity argument, when contemporary views of validity for complex performance assessment are used.

Conclusions Initial evidence for validity emerges from the development processes and the resulting milestones. Further advancing a validity argument will require research on the use of milestone data in resident assessment and program accreditation.

Introduction

The Accreditation Council for Graduate Medical Education's (ACGME) Next Accreditation System (NAS) will introduce a new approach to accreditation, beginning in July 2013. In the NAS, resident performance of educational milestones will be one of 9 indicators of programs' educational effectiveness¹ and will provide information on the extent to which programs meet educational outcome goals. Milestone data also will be used to inform improvements in resident and program performance and to assure the public that graduates of accredited residencies possess the knowledge, skills, and attitudes needed for the safe, effective, and compassionate practice of medicine.

Milestone development began in 2009 as the next step in the ACGME's Outcome Project and as a partnership between ACGME and the American Board of Medical Specialties' (ABMS) member boards. The aim of this initiative was for each specialty community to develop a consensus set of milestones in the 6 competencies.² It was envisioned that milestones would expand description of the

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

Susan R. Swing, PhD, is Vice-President, Outcome Assessment, Accreditation Council for Graduate Medical Education; **Michael S. Beeson, MD, MBA**, is Program Director of the Department of Emergency Medicine, Akron General Medical Center; **Carol Carraccio, MD, MA**, is Vice President, Competency-Based Assessment, American Board of Pediatrics; **Michael Coburn, MD**, is Department Chair, Urology, Baylor College of Medicine; **William Lobst, MD**, is Vice-President of Academic Affairs, American Board of Internal Medicine; **Nathan R. Selden, MD, PhD**, is Campagna Chair of Pediatric Neurosurgery, Oregon Health and Science University; **Peter J. Stern, MD**, is Norman S and Elizabeth CA Hill Professor and Chairman of Orthopaedic Surgery, University of Cincinnati College of Medicine; and **Kay Vydareny, MD**, is Professor Emeritus of Radiology, Emory University School of Medicine, and Associate Executive Director of Radiology, American Board of Radiology.

The authors thank the members of their respective milestone workgroups for their commitment to the milestone development work and Ingrid Philibert, PhD, MBA, for guidance in developing the manuscript.

Corresponding author: Susan Swing, PhD, Vice President of Outcome Assessment, Accreditation Council for Graduate Medical Education, 515 North State Street, Ste. 2000, Chicago, IL 60654, srs@acgme.org

Received: December 4, 2012; revision received: January 13, 2013; accepted: January 20, 2013.

DOI: <http://dx.doi.org/10.4300/JGME-05-01-33>

competencies into a progression of observable behaviors, making them more specific and easier to understand and assess. Each specialty's milestones would be organized into an assessment rubric that residency programs would use at least annually to report on residents' progress in reaching milestones.¹ This would extend the Outcome Project by collecting educational outcome data by using a standard reporting instrument.

This vision for the milestones and their use was a response to the reality that the Outcome Project fell short of its goals. While the Outcome Project stimulated enhancements in teaching and assessment at national and program levels,³ it did not result in the collection and use of outcome data in accreditation and public accountability. Factors that impeded achievement of these goals included a lack of clarity in the conceptual meaning of the general competencies, failure to develop and implement a uniform set of assessment approaches, and insufficient faculty development for their role in assessing the competencies.

Structuring of milestones into levels and the use of the milestone framework for specialty-wide assessment represents a significant change in current practice. To date, there are few reports in the published literature that document the use of milestone-like assessment frameworks in residency.^{4,5} A recent review of rubrics used in a variety of educational settings for student assessment reported positive findings related to their reliability and educational consequences but little other evidence for validity, possibly because of the narrow definition of validity used in the studies.⁶ Establishing the validity of milestone-based assessments is important in ensuring the education community's confidence in the milestones and the success of a new accreditation system based on educational outcomes.

Workgroups in all accredited core specialties are nearing completion of milestones. Here we describe milestone development for the 7 specialties that will be the first to launch the NAS in July 2013 and how the processes used by the specialties contributes to supporting a validity argument for the milestones.⁷

Methods

Validity Framework

The lens for examining initial evidence for milestone validity in this paper is a contemporary view that describes validity for complex performance assessments in terms of multiple qualities, including some that originated in a traditional psychometric perspective, construct validity, response processes, content quality and coverage, and reproducibility of decisions, as well as those identified more recently, including acceptability, meaningfulness, practicality, transparency, and consequences.⁸⁻⁹ It also proposes

that evidence for validity can be accumulated through various sources in addition to the traditional ones involving collection of performance data. Absent data from field tests, sources of evidence for validity include participation of content experts; use of content blueprints, models, curricula, and theory and logic; and existing relevant empirical studies.¹⁰⁻¹¹

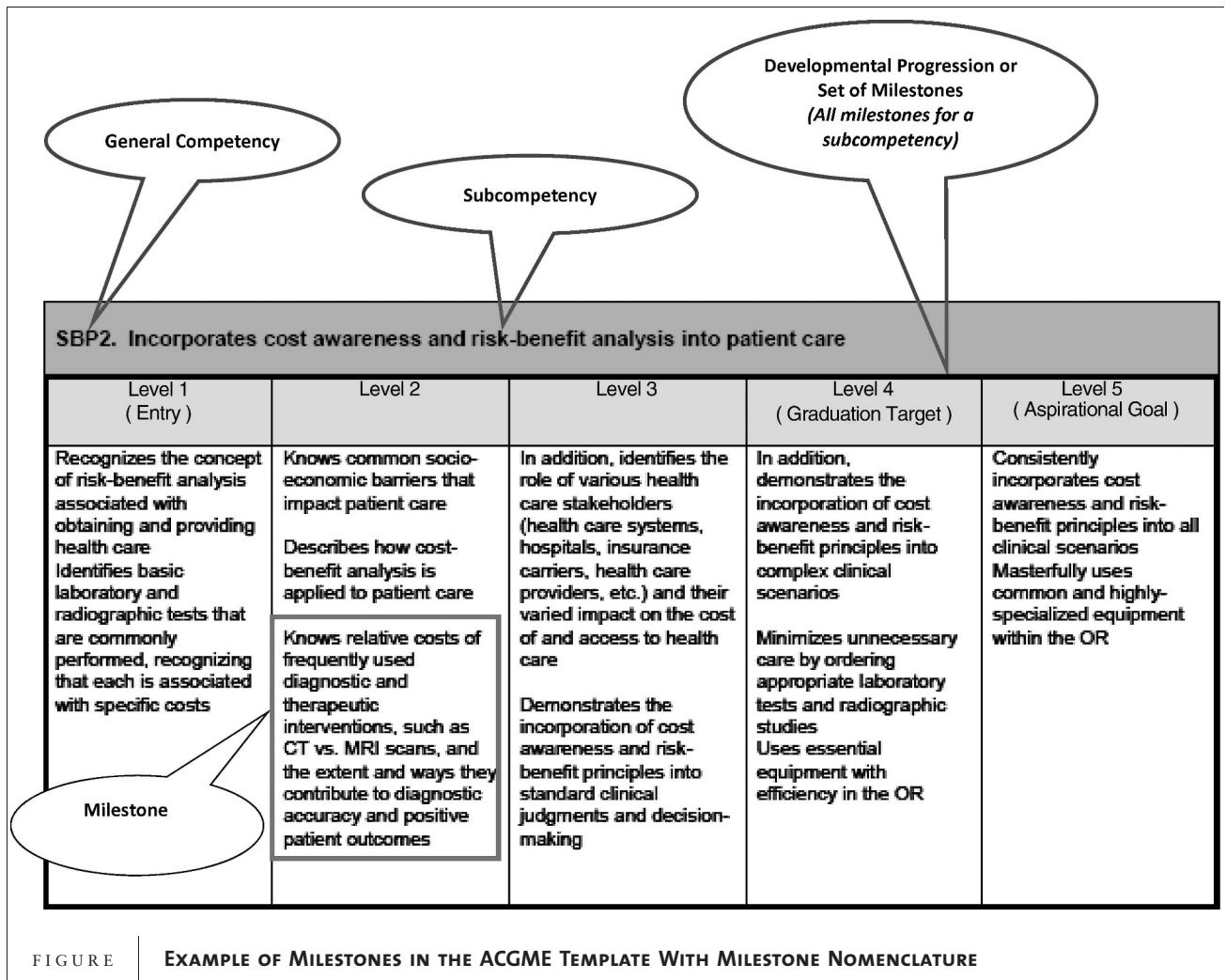
Milestone Development

The task of milestone development required specialty workgroups to first identify subcompetencies (components of competencies) in the 6 competencies by adopting those in the ACGME's common program requirements or reaching consensus on others drawn from specialty content or experiential knowledge. As the work progressed, the ACGME asked that the milestones be organized in accordance with a 5-level model beginning with milestones for matriculation into the specialty, progressing to a graduation target, and ending at an advanced level, with milestones that were "aspirational" goals that would be achieved only by exceptional residents. The 5-level framework was derived from the Dreyfus model of expertise.¹² As a general guideline, ACGME suggested that the workgroups develop milestones for approximately 36 subcompetencies distributed roughly equally across the 6 competencies. The *FIGURE* shows a milestone set for a subcompetency organized into the 5-level ACGME framework along with nomenclature for the milestone structure.

Milestone groups used numerous methods and processes to develop milestones. The sections below describe a subset of processes and related structural features selected because of their potential contributions to the validity argument for the milestones.

General Features A number of milestone development processes were common across milestone workgroups. Groups consisted of leading physician educators in the specialty, with representation from the specialty's certification board, the residency review committee (RRC), the program director community, and residents. Groups conducted their work through face-to-face meetings, conference calls, and individual and small group work.

Concurrently, a multidisciplinary panel of experts in practice-based learning and improvement, systems-based practice, interpersonal and communication skills, and professionalism created milestones in these 4 competency areas, informed by the literature and their experience as teachers and researchers. The intent of this process was to provide a model set of milestones that specialty milestone groups could use to inform and expedite their own milestone development. The milestone workgroups were presented with the opportunity to review and adopt or adapt these milestones but could decline to use them.



SPB2 is the notation for the second systems-based practice set of milestones.

Workgroups were encouraged to conduct pilots and/or gather feedback from their constituents about the milestones, and a pilot test prototype was made available. The prototype included instructions to programs for how to implement a realistic small-scale semiannual milestone assessment and a postimplementation survey for collection of faculty perspectives on milestone attributes and utility and recommendations for improvement (TABLE 1). Three specialty milestone groups implemented pilots that used the entire prototype (urology) or elements of it (internal medicine and orthopaedic surgery).

TABLE 2 summarizes other development processes exhibited by 2 or more milestone groups. These processes include, in addition to data-producing pilot studies, involvement of content specialists and use of existing specialty educational requirements and curricula, theoretical and empirical literature, and constituent feedback.

Results

Emergency Medicine The emergency medicine (EM) milestone workgroup included 10 individuals with significant experience in EM education and general competency scholarly activity and who represented 9 different organizations in EM. In addition, an advisory panel consisted of senior leadership from the American Board of Emergency Medicine (ABEM), the RRC-EM, and the ACGME.

The workgroup started its milestone development by selecting physician tasks from the Model of the Clinical Practice of Emergency Medicine.¹³ This model specifies the comprehensive core content of EM and includes conditions, an acuity assignment, and physician task definitions. Second, the group populated the physician task frameworks using knowledge, skills, and attitudes (KSAs) previously corroborated as physician task definitions during the ABEM's certification standards project that was

TABLE 1 ACGME PILOT TEST PROTOTYPE DESCRIPTION

Pilot Test Overview: The primary purpose of the test is to obtain information on the usability and usefulness of the milestones under conditions similar to those in which the milestones will be implemented for accreditation beginning in July 2013.	
Assessment of residents against milestones	
Instructions for performing the pilot were provided. The following instructions are abbreviated from the original pilot protocol.	
Preparing for the test	
1. Assemble a small group of faculty (eg, 3–5) to participate in the study as the program's competency committee. ^a The competency committee will assess residents against milestones.	
2. Select 2 residents at each year of training.	
- It is desirable to select residents who have different strengths and weaknesses. - It is desirable to select at least 1 or 2 residents who may not be progressing at the pace expected for a resident at that year of training.	
3. Assemble evaluations for these residents (eg, rotation evaluations, 360° evaluations, patient surveys, and others) for this academic year.	
4. Each competency committee member should review the milestones using the Milestone semiannual report form	
Performing the test by assessing residents on the milestones using the semi-annual evaluation form	
1. The competency committee (as a group) ^a will discuss each resident's performance in relation to each set of milestones, taking into account the resident's assessments for the academic year.	
2. The competency committee will complete the semiannual milestone report for each resident by (date).	
Postpilot Test Survey	
Questions	
1. Demographic	
2. Perspectives on milestones regarding:	
- Understanding of milestones - Adequacy of existing information about residents to enable accurate milestone evaluation - Milestone progression - Ease/difficulty of selecting a single level - Usefulness of existing resident evaluations for informing milestone reporting - Meaningfulness of milestone evaluations	
3. Time to perform milestone assessments and reporting	
4. Helpfulness of specific resident evaluation methods	
5. Modifications needed to milestones by general competency (open-ended)	

^a Variations in pilots occurred. Involvement of a competency committee was not required in orthopaedic surgery. Options are provided for how to conduct the assessments using a competency committee; for example, a key faculty member on the clinical competency committee (CCC), can complete the milestone form and bring it to the CCC for review, discussion, and finalization.

Abbreviations: ACGME, Accreditation Council for Graduate Medical Education; PGY, postgraduate year.

completed prior to the start of EM milestone development. During this project, KSAs within physician tasks were ordered hierarchically, from less to more difficult, so that the ability to perform a more difficult task indicated ability to perform a less difficult one. KSAs were selected as certification standards based on their position in the hierarchy.

The milestone workgroup used this information to organize the KSAs into its 5-level milestone structure. The

EM milestone advisory panel closely monitored milestone development and ensured that the ABEM certification standards were aligned with the level 4 milestone level (the graduate target) for each physician task, resulting in alignment of 14 of 23 subcompetencies. For the subcompetencies without ABEM standards, the workgroup modified milestones from the set developed by the ACGME expert panel.

The ABEM then conducted a large-scale study on behalf of the workgroup to investigate milestone level

TABLE 2 MILESTONE DEVELOPMENT PROCESSES CONDUCTED BY SPECIALTIES

Work Group	Specialties						
	EM	IM	NS	OS	PEDS	DR	URO
Included specialty content experts	*	*	**	**	*	**	*
Included specialists in the nonspecialty-specific competencies	*	*	*		**		*
Used review committee requirements, board content outlines, specialty models, or curricula	**	*	**	*	*		*
Used milestones created by ACGME expert panel	*		*	*			
Used theoretical and empirical literature		*			**		*
Used data to determine milestone levels	**						*
Obtained feedback from a broad base of specialty educators and changed milestones in response	**	**	**		*	**	*
Conducted implementation pilot test		*	*	*	*		*

* Fundamental process performed.

** Integral, robust process performed.

Abbreviations: DR, diagnostic radiology; EM, emergency medicine; IM, internal medicine; NS, neurological surgery; OS, orthopaedic surgery; PEDS, pediatrics; URO, urology.

validity. Program directors and faculty in each EM program received milestones for half of the subcompetencies and were asked to independently assign each to an appropriate level (ie, level 1, level 2, and so forth) in the framework (FIGURE), using the definitions provided in the study for each level. The results demonstrated a high level of agreement with the assignments made by the workgroup. Differences were reconciled predominantly by changing the levels of discordant milestones. The outcome was a final version of the EM milestones supported by several types of validity evidence.

Internal Medicine In November 2007, the American Board of Internal Medicine and the ACGME convened a 33-member milestone development task force composed of program directors, representatives of the major internal medicine (IM) stakeholder groups, and experts in evaluation and the ACGME-ABMS competencies. Over the course of multiple meetings and interim work, the workgroup reviewed the literature, accreditation standards, and “best practice” in graduate medical education curricula and developed 142 observable behaviors describing the development of competence in the 6 competencies.¹⁴ The internal medicine community was surveyed for feedback regarding the utility of the IM milestones, and more than 85% of the 253 faculty educators, residents, and fellows who responded believed the milestones would enhance residency curricula, improve

trainee evaluation and feedback, and facilitate faculty development.

While these milestones, now referred to as curricular milestones, had utility for assessing specific behaviors, a second group of IM educators was charged with developing a more synthetic set of milestones, using the original 142 milestones to write narrative-based developmental milestones describing individual learners. This organized the milestones into 22 subcompetencies that used the language of the ACGME common program requirements. In the spring of 2012, a “proof of concept” pilot involved 38 IM residencies in a field test of this version of the milestones. The pilot highlighted the need to standardize the language across the subcompetencies and eliminate redundancy between subcompetency milestones. It also highlighted that use of the milestones would increase the time required for resident evaluation. With this community feedback, the milestones were revised, and the current version was developed.

Neurological Surgery The neurological surgery RRC, in consultation with the Society of Neurological Surgeons (SNS), appointed a 17-member milestone workgroup representing 8 subspecialty disciplines, program directors, and department chairs, as well as an advisory group consisting of RRC and SNS leadership. Some members had expertise in the general competencies. The workgroup developed a priority list of key domains in each competency area based on an existing neurosurgical “matrix”

curriculum developed by the SNS and reflecting written board examination subject topics and RRC case categories. The priority list was further evaluated by pilot groups of program directors and faculty at a dozen programs.

The RRC appointed a milestones group of 8 neurological surgeons, 1 for each subspecialty area. The group first developed principles to guide milestone development, including the principle that milestone elements would represent samples of major domains of neurological surgery KSAs, would recur across developmental levels, and be clearly assessable.

The workgroup formulated milestones for 24 subcompetencies, with 16 subcompetencies covering medical knowledge and patient care in each subspecialty area, and 8 covering the other 4 general competencies. Draft milestones were approved by the advisory group and were presented to more than 100 program directors and coordinators in an interactive workshop. At this meeting, the milestones group demonstrated, and groups of 6–8 program directors subsequently performed, mock clinical competency committee semiannual assessments using summary information from fictional resident portfolios. Final milestones revisions were made based on suggestions from the interactive workshop and input from the advisory group.

Orthopaedic Surgery In 2011, the ACGME and American Board of Orthopaedic Surgery convened a workgroup of orthopaedic surgeons with expertise in the 8 major subspecialties to develop competency-based milestones. Each participant had expertise in at least one of the 8 major orthopaedic subspecialties.

The workgroup reviewed, selected, and modified milestones from the ACGME expert panel to create the orthopaedic milestones for the 4 nonspecialty-specific general competencies and spent the majority of its time developing milestones for patient care and medical knowledge. The workgroup chose 16 focus areas thought to represent a sample of the spectrum of orthopaedic surgery, including anterior cruciate ligament injury, shaft fracture of the tibia/femur, and metastatic disease. After selecting the milestones, the group organized them into the ACGME's 5-level template.

The milestones then underwent 2 rounds of field testing using an adapted version of the ACGME pilot test prototype. Selected and volunteer program directors, attending orthopaedic surgeons, and residents from 15 programs were asked to try the newly developed milestones. Eighty percent of the 75 participants who responded to the postpilot survey felt the milestones allowed for meaningful resident evaluation. There also was agreement that the milestones discriminated between levels of competency and represented a realistic progression of

knowledge, skills, and behaviors. There was some concern that it was difficult to accurately describe resident performance by selecting a single level of milestones.

Pediatrics In contrast to the other specialties, the pediatrics milestones span the continuum from the novice performance of early medical students to the expert performance expected of physician experts who were years into practice. The pediatrics milestone workgroup agreed on a set of guiding principles and conceptual considerations, and a national advisory board kept the working group focused on its desired outcome.¹⁵ Members of the Association of Pediatric Program Directors (AAPD) informed the process by providing feedback at an annual meeting to a working group query about subcompetencies that needed to be made more explicit. The feedback led to the addition of personal and professional development subcompetencies to those already specified in the pediatrics training requirements (eg, flexibility in adjusting to change with capacity to alter behavior, managing uncertainty, and demonstrating trustworthiness). The in-depth reviews of the medical, social science, and education literatures informed the selection and sequencing of behaviors described in brief narratives for each milestone, and the workgroup refined this work over 2 years.¹⁶ Content experts reviewed the milestones where the literature was not conclusive. The pediatrics workgroup views milestone development as an ongoing improvement project.¹⁷ It plans to study the milestones in vivo to determine when learners typically achieve the sequential levels of performance to set standards.¹⁸

Cognitive interviews with residents demonstrate these key stakeholders' understanding of the milestones and their ability to serve as learning roadmaps.¹⁹ The AAPD partnered with the National Board of Medical Examiners (NBME) in an ongoing pilot of a defined cluster of milestones related to 9 subcompetencies relevant to predicting readiness of subinterns and interns to care for inpatients. Twenty programs are observing and recording learner behaviors on smart phones, using assessment tools developed by AAPD and NBME, with AAPD Longitudinal Educational Assessment Research Network (LEARN) serving as a data repository. The American Board of Pediatrics' Foundation will support a scale up of this pilot to a national study of the 21 milestones that will be reported as part of the NAS.

Diagnostic Radiology The milestone workgroup was composed of 13 experts, covering the breadth of subspecialties in diagnostic radiology, and a resident. The workgroup recognized that it needed to focus on milestone utility for program directors to inspire program directors' acceptance and efforts to accurately assess milestones and

thereby make a difference. The workgroup involved program directors throughout the milestone development process. It presented the draft milestones at the spring 2012 meeting of the Association of Program Directors in Radiology (APDR) and urged the participants to share their thoughts, for example, were there too many Milestones? Did the model forget anything important? Were the benchmark anchors appropriate for each year of training? The milestone workgroup also regularly contacted the APDR committee on milestones to request feedback on versions of the draft milestones. The workgroup decreased the number of milestones from 30 to 12 based on this feedback, after program directors informed the group that evaluation of the original milestone set would be tedious at best and, at worst, impossible. The reduction was accomplished largely by combining multiple milestones into an overarching goal to allow program directors to easily see if residents were in “substantial compliance” with these larger concepts. When the final draft was circulated to the program directors, it met with support. A field test that uses the ACGME pilot prototype is being planned for the first half of 2013, and results will be used to decide whether changes to the milestones will be made.

Urology The Urology milestone workgroup included 8 physician educators, 1 resident, and 1 recent graduate now in practice. Members were selected who had special expertise in at least one of the 6 competencies. The workgroup reviewed published scholarly work related to the competencies in each of the nonspecialty specific domains, and worked with the American Urological Association’s Office of Education to explore how to use the new national urology core curriculum in milestone development.²⁰ Performance levels for medical knowledge were based on the national database of performance on the urology in-service examination.

The draft milestones underwent multiple iterations of review and revisions. After the first draft was completed, the workgroup presented the rationale, development considerations, and examples of the milestones to program directors at their annual meeting. The workgroup reviewed the milestones against those developed by ACGME’s expert panel and made modifications as needed. An 8-member advisory group made up of representatives from all major urology organizations with interest in graduate medical education reviewed the milestones, as did the RRC. Finally, residency programs associated with 8 workgroup and urology RRC members participated in an alpha pilot test using the ACGME pilot prototype. From 72%–88% of the 25 participants reported on the postmilestone test survey that they understood the milestones, thought they represented a realistic progression, and were meaningful for

assessment purposes. Overall, data showed an expected pattern of performance with senior residents assigned higher performance levels on the milestone than junior residents.

Specialty Milestones

The milestones developed by the workgroups in the 7 specialties are presented on the ACGME website²¹ and in a supplement to this issue. All milestones are organized based on the 6 competencies and in a developmental framework. TABLE 3 presents broad characteristics of the milestones: number of subcompetencies, typical number of milestones per level, and a characterization of milestones as continuous or discontinuous. The number of subcompetencies for which milestones have been developed ranges from 41 for orthopaedic surgery to 12 for radiology. The typical level is populated with 1–3 milestones; however, neurological surgery and orthopaedic surgery included 4 and 7 milestones, respectively, in their patient care and medical knowledge milestones. Milestones of most specialties are continuous, that is, the same KSA or work process is described across levels but becomes increasingly complex or more challenging to perform. The milestones within and across specialties also vary somewhat in their specificity and observability.

Validity and Utility

TABLE 2 reports development processes performed by each specialty. Processes are classified as fundamental or robust. This classification intends to capture the relative quality of the process; for example, fundamental use of the literature means that the workgroup consulted a small selection of articles for each competency, whereas robust indicates a thorough literature review. The milestone workgroups in most specialties performed at least 1 robust process that could enhance the validity and utility of the milestones, along with several fundamental processes. Four specialty workgroups (EM, IM, neurological surgery, and diagnostic radiology) engaged large numbers of constituents in milestone review and modified their milestones based on this feedback. Only pediatrics made extensive use of specialists in the nonspecialty specific competencies and the theoretical and empirical literature, and only EM used data extensively to verify milestone levels. Pilot test results affirmed the clarity and meaningfulness of milestones. For example, the urology pilot showed more senior residents were placed at higher milestone levels than junior residents.

Discussion

An important aim of the milestone initiative was for milestones to be developed through engagement of the specialty community. This occurred in all specialties through the specialty milestone groups and, in most specialties, through opportunities for resident educators to

TABLE 3 CHARACTERISTICS OF MILESTONES BY SPECIALTY

Characteristic	Specialty						
	EM	IM	NS	OS	PEDS	DR	URO
PC and MK subcompetencies							
Number of subcompetencies	15	7	16	32	6 ^d	4	10
Typical number of milestones per level	1–3	1–4	4–5	2–7	1 ^e	1–2	1–2
Progression Continuous [C] ^a Discontinuous [D] ^b Mixed [M] ^c	PC-M,MK-C	C	PC-M,MK-D	PC-M,MK-D	C	C	PC-C,MK-D
ICS, PROF, PBLI, SBP subcompetencies							
Number of subcompetencies	8	16	8	9	15 ^d	8	22
Typical number of milestones per level	1–2	2–3	2–3	1–3	1 ^e	1–2	1–2
Progression Continuous [C] ^a Discontinuous [D] ^b Mixed [M] ^c	C, D, M	C	M	C	C	C	C

^a Continuous, the same KSAs develop progressively across at least two to three levels.

^b Discontinuous, progression occurs as the acquisition of different KSAs across levels.

^c Mixed, some milestones are continuous and some are discontinuous.

^d The number of subcompetencies shown is for the subset of pediatrics milestones that will be used for accreditation. Pediatrics has 14 PC and MK subcompetencies and 35 ICS, PROF, PBLI, and SBP milestones in its full set, which is available for use by pediatrics residency programs.

^e Milestones are developed as single, integrated narratives that may or may not include more than one KSA or cognitive process.

Abbreviations: PC=Patient Care; MK=Medical Knowledge; ICS=Interpersonal and Communication Skills; PROF=Professionalism; PBLI=Practice-Based Learning and Improvement; SBP=Systems-Based Practice; EM=Emergency Medicine; IM=Internal Medicine; NS=Neurological Surgery; OS=Orthopaedic Surgery; PEDS= Pediatrics; DR=Diagnostic Radiology; URO=Urology.

provide feedback and participate in ongoing dialogue. In one specialty (radiology), program director and faculty feedback and dialogue resulted in significant reduction in milestones. Involvement of constituents in the development process and responsiveness to feedback contributes to acceptability of milestones.

Milestone frameworks for all specialties present a progression of KSAs from less to more advanced. Specific framework designs and interpretations, however, differ most notably in whether the levels are aligned with the ACGME-template (EM) or with theoretical and empirical concepts of cognitive-behavioral development (pediatrics). The implicit performance expectations of the ACGME frameworks could bias responses but also add to transparency.

The specialty workgroups used several approaches to ensure initial content-related evidence for validity, including participation of content experts and use of specialty curricula, existing certification and accreditation standards, blueprints and models, and existing empirical studies, beginning the process of establishing the basis for a validity argument.^{7–9} Milestones representing certification board expectations, for example, would likely increase meaningfulness of milestones for residents, as almost all residents

aspire to achieve specialty board certification. Content quality and coverage should be enhanced by using blueprints and models resulting from prior content experts' processing of the material.

The same processes that produce positive effects could have counterproductive ones. For example, responsiveness to feedback that results in significant reduction of to-be-assessed milestones may increase acceptability but reduces content coverage and content-related evidence for validity. Although content experts can ensure that milestone content is consistent with current best understanding of an area, they can set expectations too high, so that others deem the milestones unreasonable and unacceptable. Some of the milestones created by the ACGME expert panel were not accepted for this reason.

The milestones, although structured similarly as a progression of capabilities, are richly diverse across specialties. Some differences may affect the validity of milestone assessments, such as (1) the extent to which milestones are specific and observable; (2) the number of milestones clustered within a level; or (3) the overall number of subcompetencies and milestones. By default, assessors must bring their own standards and criteria to the assessment of a general milestone, such as “effectively communicates with

patients,” and these could vary with the assessor’s own levels of expertise,²² thereby affecting reproducibility of decisions. A large number of subcompetencies and milestones should increase content coverage while adversely effecting practicality for assessment by increasing resource requirements such as faculty time and cost.

Variations in milestones across specialties will make the task of developing assessment tools aligned with milestones and corresponding faculty training a more time consuming and resource-intensive undertaking. On the positive side, variations offer the possibility of investigating the relative utility, acceptability, and meaningfulness of the different milestones and evolving all milestones towards those that yield better overall results for ACGME, residency educators, residents, and the public.

Next steps for the NAS phase 1 milestone development and implementation include: (1) collection of resident milestone assessments performed by programs’ clinical competency committees; (2) RRC review of aggregate milestone data to ascertain whether residents are progressing; and (3) study of multi-year data collected from programs in the 7 specialties followed by modification of milestones as needed. Development of assessment approaches better aligned with milestones is planned and in some specialties is already in progress, as is development of faculty to perform milestone assessments.

Limitations

Other than initial evidence from pilot studies, the existing evidence of milestone validity results from the involvement of experts and analysis of the content and constructs of competence in the profession. Different conclusions about validity could be forthcoming based on more extensive use of the milestones in resident assessment and program accreditation.

Conclusions

The construction of specialty-specific frameworks of developmental milestones represents an innovative step toward ensuring the collection and use of educational outcome data, based on residents’ aggregate performance on milestones, in accreditation and public reporting. Milestone development processes provide initial evidence for the validity of milestones that will be used for these purposes. Additional investigations of milestone validity and utility as part of the ongoing pilot studies, and broader

field tests following the implementation of the milestones for the first 7 specialties, are needed to support a broader validity argument, which will be important to the success of the outcomes-based approach to accreditation envisioned under the NAS.

References

- 1 Nasca TJ, Philibert I, Brigham T, Flynn TC. The next accreditation system: rationale and benefits. *N Engl J Med*. 2012;366:1051–1056.
- 2 Nasca T. The next step in the outcomes-based accreditation project. *ACGME Bull*. 2008;2–4.
- 3 Swing SR. The ACGME Outcome Project: retrospective and prospective. *Med Teach*. 2007;29:648–654.
- 4 Varney A, Todd C, Hingle S, Clark M. Description of a developmental criterion-referenced assessment for promoting competence in internal medicine residents. *J Grad Med Educ*. 2009;1(1):73–81.
- 5 Meade LB, Borden SH, McArdle P, et al. From theory to actual practice: creation and application of milestones in an internal medicine residency program, 2004–2010. *Med Teach*. 2012;34(9):717–723.
- 6 Jonsson A, Svingby G. The use of scoring rubrics: reliability, validity, and educational consequences. *Educ Res Rev*. 2007;2:130–144.
- 7 Kane M. Current concerns in validity theory. *Journal of Educational Measurement*. 2001;38(4):319–342.
- 8 Linn RL, Baker EL, Dunbar SB. Complex, performance-based assessment: expectations and validation. *Educ Res*. 1991;20:15–21.
- 9 Bartman LKJ, Bastiaens TJ, Kischer PA, van der Vleuten, CPM. The wheel of competency assessment: presenting quality criteria for competency assessment programs. *Studies in Educational Evaluation*. 2006;32:153–170.
- 10 Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med*. 2006;119(2):166.e7–166.e16.
- 11 Downing S. Validity: on the meaningful interpretation of assessment data. *Med Educ*. 2003;37:830–837.
- 12 Dreyfus, HL. *On the Internet: Thinking in Action*. New York: Routledge; 2003.
- 13 Perina, DG, Beeson MS, Char DM, Counselman FL, et al. The 2007 model of the clinical practice of emergency medicine: the 2009 update. *Acad Emerg Med*. 2011;18(3):e8–e26.
- 14 Greene ML, Aagaard EM, Caverzagie KJ, et al. Charting the road to competence: developmental milestones for internal medicine residency training. *J Grad Med Educ*. 2009;1(1):5–20.
- 15 Hicks PJ, Schumacher DJ, Benson BJ, et al. The pediatrics milestones: conceptual framework, guiding principles, and approach to development. *J Grad Med Educ*. 2010;2(3):410–418.
- 16 American Board of Pediatrics and Accreditation Council for Graduate Medical Education. The pediatrics milestone project. <https://www.abp.org/abpwebsite/publicat/milestones.pdf>. Accessed January 26, 2013.
- 17 Englander R, Burke A, Guralnick S, et al. The pediatric milestones: a continuous quality improvement project is launched—now the hard work begins! *Acad Pediatr*. 2012;12(6):471–474.
- 18 Hicks PJ, Englander R, Schumacher DJ, et al. Pediatrics milestone project: next steps toward meaningful outcomes assessment. *J Grad Med Educ*. 2010;2(4):577–584.
- 19 Schumacher DJ, Lewis KO, Burke AE, et al. The pediatrics milestones: initial evidence for their use as learning roadmaps for residents. *Acad Pediatr*. 2013;13(1):40–47.
- 20 McDougall EM, Watters TJ, Clayman RV. 4-year curriculum in urology residency training. *J Urol*. 2007;178(6):2540–2544.
- 21 Accreditation Council for Graduate Medical Education. The next accreditation system: recent news: milestones-phase 1. ACGME website. <http://www.acgme-nas.org/>. Accessed January 26, 2013.
- 22 Kogan JR, Conforti LN, Bernabeo EC, et al. Faculty staff perceptions of feedback to residents after direct observation of clinical skills. *Med Educ*. 2012;46(2):201–215.