Pediatrics Milestone Project: Next Steps Toward Meaningful Outcomes Assessment

Patricia J. Hicks, MD
Robert Englander, MD, MPH
Daniel J. Schumacher, MD
Ann Burke, MD
Bradley J. Benson, MD
Susan Guralnick, MD
Stephen Ludwig, MD
Carol Carraccio, MD, MA

Summary

In the September 2010 issue of JGME, the Pediatric Milestones Working Group published "The Pediatrics Milestones: Conceptual Framework, Guiding Principles, and Approach to Development", a document that describes the construction of the first iteration of the Pediatric Milestones. These Milestones were developed by the Working Group as a group of practical behavioral expectations for each of the 52 sub-competencies. In constructing these Milestones, the authors were cognizant of the need to ground the Milestones themselves in evidence, theories or other conceptual frameworks that would provide the basis for the ontogeny of development for each sub-competency. During this next phase of the Milestones development, the process will continue with consultation with content experts and consideration of assessment of Milestones. We have described possible measurement tools, explored threats to validity, establishment of benchmarks, and possible approaches to reporting of performance. The vision of the Pediatrics Milestone Project is to understand the development of a pediatrician from entry into medical school through the twilight of a physician's career, and the work will require a collaborative effort of the undergraduate and graduate medical education communities, and the accrediting and certifying bodies.

Patricia J. Hicks, MD, is Director of the Pediatric Residency Program at The Children's Hospital of Philadelphia and Professor of Clinical Pediatrics in the Department of Pediatrics at University of Pennsylvania School of Medicine; **Robert Englander, MD, MPH,** is Senior Vice President of Quality and Patient Safety at Connecticut Children's Medical Center and Professor of Pediatrics at University of Connecticut School of Medicine; **Daniel J. Schumacher, MD,** is Clinical Fellow in Emergency Medicine at Cincinnati Children's Hospital Medical Center and in the Department of Pediatrics at the University of Cincinnati College of Medicine; **Ann Burke, MD,** is Director of the Pediatric Residency Program at Wright State University, Boonshoft School of Medicine, and in the Department of Pediatrics at the Dayton Children's Medical Center; **Bradley J. Benson, MD,** is Director of the Med-Peds Program at the University of Minnesota Amplatz Children's Hospital and Director of the Division of General Internal Medicine and Associate Professor of Internal Medicine and Pediatrics at the University of Minnesota School of Medicine; **Susan Guralnick, MD,** is Designated Institutional Official at Winthrop University Hospital and Director of Graduate Medical Education at Winthrop University Hospital, Associate Professor in the Department of Pediatrics at Winthrop University Hospital, and Associate Professor in the Department of Pediatrics at Stony Brook University School of Medicine; **Stephen Ludwig, MD,** is Designated Institutional Official I and Chairman of Graduate Medical Education at Children's Hospital of Philadelphia and Professor of Pediatrics and Professor of Emergency Medicine at the University of Pennsylvania School of Medicine; and **Carol Carraccio, MD, MA,** is Associate Chair for Education at the University of Maryland Hospital for Children and Professor in the Department of Pediatrics at the University of Maryland School of Medicine.

We would like to thank Lisa Johnson, MBA, for her helpful support of the Pediatric Milestones since its very beginnings. Her organizational assistance, formatting of the writing of the Milestones first iteration, and her work on translating our ideas to figures has been most appreciated. Ms Johnson developed FIGURE 2, which was also presented to the Milestones Chairs group.

Corresponding author: Patricia J. Hicks, MD, Children's Hospital of Philadelphia, Main Hospital 9NW–Room 64, 34th & Civic Center Boulevard, Philadelphia, PA 19104, 215.764.7973, hicksp@email.chop.edu

Received August 16, 2010; revision received September 20, 2010; accepted October 4, 2010.

DOI: 10.4300/JGME-D-10-00157.1

Introduction

In the September 2010 issue of the *Journal of Graduate Medical Education*, the Pediatrics Milestone Project Working Group published "The Pediatrics Milestones: Conceptual Framework, Guiding Principles, and Approach to Development."[1] The aim was to share the approach to constructing the first iteration of the Pediatrics Milestones, a compilation of documents (*Milestones*) for each of 52 subcompetencies. This work, which is grounded in the literature, attempts to bridge theoretical constructs about how competency develops with practical behavioral expectations for the developing pediatrician. Much work remains in transitioning from the current iteration of the Pediatrics Milestones to the realization of a dynamic, living document useful for formative and summative assessment of learners. Our purpose in this manuscript is 2-fold: (1) to describe the next steps in refining the Milestones, applying assessment principles to them, and setting performance standards; and (2) to explore the role the Milestones will play in advancing competency-based assessment.

Refining the Pediatrics Milestones

Engaging Content Experts

In developing the first iteration of the Pediatrics Milestones, the working group reviewed and built upon the literature on the ontogeny of development of the competencies. In some cases, the literature provides strong evidence for the details of this progression, as in the case of clinical reasoning.[2–7] For

Downloaded from https://prime-pdf-watermark.prime-prod.pubfactory.com/ at 2025-10-27 via free access

Note: This example uses the "Making informed diagnostic and therapeutic decisions that result in optimal clinical judgment" Milestone to provide select specific examples. To view this full Milestone document, with the references 2 and 3 cited in the developmental milestones below, please see Appendix C.

Competency: Patient Care

One of the six ACGME competency domains

Sub-competency: Making informed diagnostic and therapeutic decisions that result in optimal clinical judgment

One of the sub-competencies listed under the main competency domain as a current Requirement for Residency Training in Pediatrics, as created by the Pediatrics Review Committee of the ACGME[a]; each represents a complex task that integrates a number of knowledge, skill and attitude objectives

Background:
In this space, a brief summary of the literature is written to support the proposed developmental progression for the sub-competency addressed in the Milestone. Graphical illustrations and tables are used where deemed helpful in describing difficult concepts and organizing or clarifying the developmental progression of a sub-competency.

Developmental Milestones:

Recalls and presents clinical facts in the history and physical in the order they were elicited **without filtering, reorganization, or synthesis**,[2] resulting in a **list of all diagnoses considered**[3] rather than the development of working diagnostic considerations, making it **difficult to develop a therapeutic plan**.

Developmental milestone anchor
(4 anchors are included in this Milestone)

Focuses on features of the clinical presentation, making a unifying diagnosis elusive and leading to a **continual search for new diagnostic possibilities**.[2] Often reorganizes clinical facts in the history and physical exam to help decide on clarifying tests to order[2] rather than to develop and prioritize a differential diagnosis. This often results in **a myriad of tests and therapies** and **unclear management plans** since there is no unifying diagnosis[2]

In this developmental milestone anchor, each *element* is underlined (5 elements are included in this anchor). Note that not all anchors contain all 5 elements.

Abstracts and reorganizes elicited clinical findings in memory, **using semantic qualifiers** (paired opposites that are used to describe clinical information – e.g. acute and chronic) to compare and contrast the diagnoses being considered when presenting or discussing a case.[3] The result is often a **well synthesized and organized assessment of the focused differential diagnosis and management plan**.

Reorganized and stored clinical information (illness and instance scripts) leads to **early directed diagnostic hypothesis testing** with subsequent history, physical, and tests used to confirm this initial schema.[2] Able to **identify discriminating features** between similar patients and avoid premature closure. **Therapies are focused and based on a unifying diagnosis**, resulting in an effective and efficient diagnostic work-up and management plan tailored to address the individual patient.[2]

References:
List of references cited in the background and in the developmental milestones listed in Appendix C.

a. ACGME. The Common Program Requirements 2007; Available from: http://acgme.org/acWebsite/dutyHours/dh_dutyhoursCommonPR07012007.pdf

FIGURE 1 **Anatomy of a Milestone**

many others, the progression is not as well defined, and members of the working group had to use theories and constructs, frequently reaching beyond the medical literature, to create a hypothesis around the developmental progression of a subcompetency.

For the Milestones where the developmental progressions are not well defined (eg, role modeling or working in interprofessional teams), the next step will be to engage experts in the relevant fields to help review and refine those Milestones. These content experts will be asked to determine whether the conceptual framework chosen represents the best theory, evidence, or working model for each Milestone. In addition, they will be asked to identify any instruments or tools they believe can measure and report performance using the Milestones they are reviewing.

Moving from Generic to Content-Specific and Context-Specific Milestones

As currently written, many of the Pediatrics Milestones use generic behavioral descriptors or anchors that are not specific to a given specialty, clinical content area, or context. In their generic form, the Milestones do not enumerate specific criteria to allow rating at each developmental level. For the Milestones to be useful in assessing performance outcomes of residents, methodologies for achieving high interrater and intrarater reliability should be developed and employed. We will need to create vignettes that describe measureable behaviors aligned with the Milestone but specific to both the content (eg, pediatrics, or a particular subject within pediatrics) and the context (eg, inpatient setting) in which the learner is being assessed. These vignettes will serve to ground the Milestones in real-world experience. These standardized scenarios could be distributed by video recording and be used to train and calibrate raters through examples of learner performance at various stages.[8–20] This will facilitate high interrater and intrarater reliability[21] and contribute to the validity of the data produced by the Milestone assessments and to the assessment of the inferences based on them. There will optimally be multiple contexts in which each Milestone is studied, as each subcompetency applies to many clinical settings, both in training and in practice. Similar assessment

Downloaded from https://prime-pdf-watermark.prime-prod.pubfactory.com/ at 2025-10-27 via free access

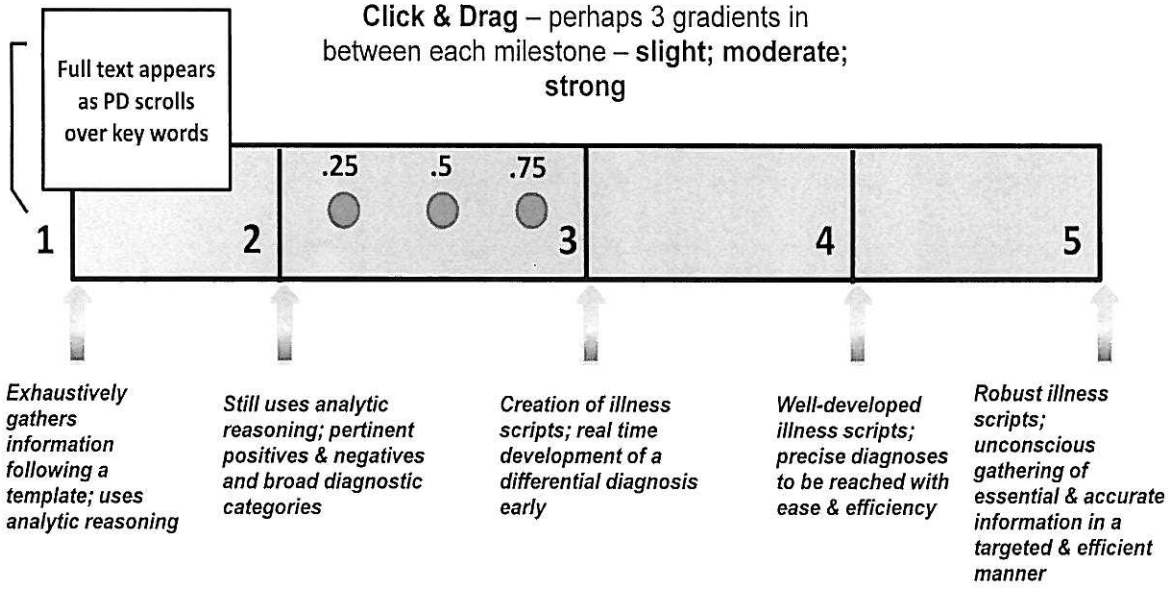

FIGURE 2 **The Slider Bar Method: a sliding scale spectrum of achievement, with clusters of elements pertaining to diagnostic reasoning for the subcompetency, "gathering essential and accurate information about the patient."**

data generated in multiple contexts would provide additional construct validity to the Milestones.

Application of Assessment Principles to the Pediatrics Milestones

Each developmental Milestone is constructed using one or more elements, shown in FIGURE 1—ANATOMY OF A MILESTONE—and discussed in our earlier article[1] (Appendix C and the references in the Figure can be found in the supplemental online materials for the September 2010 Hicks et al article). These elements range from simple and discrete variables, which are easy to measure, to complex and interrelated variables, which are challenging to measure and require measurements in clusters. An additional complicating factor is that some elements may develop synchronously and others asynchronously. For example, 2 elements of a developmental Milestone anchor in FIGURE 1 may develop well together, whereas another element may develop at a completely independent pace from the other elements. Given these complexities, the Pediatrics Milestone Project Working Group proposes 2 potential measurement methods and a reporting system that would accommodate the unique and varied nature of elements within the series of Milestones for a given sub-competency.

Proposed Measurement Tools

The "Slider-Bar" Method

The slider bar is illustrated in FIGURE 2, using the developmental Milestone for "Gathering essential and accurate information about the patient" as an example. In this method, the bar contains the elements of developmental Milestone anchors listed in clusters, with advancing mastery as one moves from left to right. The rater clicks on the display bar at the point that best represents the resident's performance, allowing for gradation between developmental levels to be reflected by where the bar is placed.

We propose the slider bar method for a number of reasons. First, it provides measurement along a true continuum, a key feature given that the developmental Milestones are not discrete variables. Second, it is technologically and conceptually easy for the rater to understand. Third, when the user clicks on a cluster of behaviors, the computer assigns a behind the scenes numerical value based on the location of the click along a predetermined numeric spectrum. This value is recorded on a back-end relational database organized to store data according to competency-specified categories assigned by the Accreditation Council for Graduate Medical Education (ACGME). While this method is user-friendly for the rater, it is also powerful in its measurement and storage of data, allowing for multiple queries that would be useful both for individual learners and for programmatic and accreditation purposes. This method is ideal for those subcompetencies in which the elements of the Milestones develop synchronously. Using the developmental Milestone in FIGURE 2 as an example, the slider bar method would be ideal if elements in the middle of the spectrum, such as "creation of illness scripts" and "real-time development of a differential diagnosis early in the information-gathering process," develop concurrently.

Downloaded from https://prime-pdf-watermark.prime-prod.pubfactory.com/ at 2025-10-27 via free access

Sub-domain	Element#	Level 0			Level χ
Ability overall - knowledge - skill	1	Global assessment score 25%	50th percentile	75th percentile	100%
	2	Scored test of knowledge25%	50th percentile	75th percentile	100%
	3	Assessment of skill (OSCE, simulation assessment	1 SD below mean	1 SD above mean	2 SD above mean
Discernment -expression of uncertainty	4	Lacks insight	Demonstrates some awareness and identification of gaps		Clear articulation and identification of gaps as determined through declared statements or through level and type of questioning [37]
	5	No spontaneous questions that indicate ambiguity or uncertainty	Some questioning is present indicating some awareness of uncertainty		Well framed questions that identify areas of uncertainty or ambiguity [38]
Conscientiousness - level of organization - level of consistency - prioritization and action aligned	6	No follow-through	Follow-through requires some level of external prompts		Follows through without any prompts (self-regulated, intrinsic)
	7	No organization	Occasional organized approach to completion of tasks		Highly organized and thorough in completion of tasks
	8	Poor follow-through on task completion	Sometimes has follow-through on task completion		Highly reliable and consistent in task completion and execution of processes
	9	Given specific instructions on what to do first and what is most important chooses to do other actions first	External prompts or reminders required		Prioritization of activities done without prompts or course correction
Truthfulness	10	Disconnects occur between reports/responses and facts/evidence	Evidence-based lack of straight-forward communication (subject to attribution failure); missing actions or incomplete tasks not discussed or if discussed were not initiated by learner		Clear communication with transparency and honesty; openly discusses missing data or lack of action in straightforward manner

FIGURE 3 **The Matrix Method: a 2-dimensional representation of achievement for complex Milestones such as Trustworthiness[36,37]**

In short, the slider bar essentially provides the opportunity for a global assessment of the elements clustered within a given series of Milestones. The disadvantage of this method is that it allows for only one value (or click) to select an entire cluster of elements rather than discrete elements of a developmental Milestone anchor. This single value is scored as a single measurement, regardless of potential for differential performance on various elements listed in that cluster. Thus, a learner who demonstrates performance of the behaviors of 3 elements of a lower developmental Milestone and 1 element of a more advanced developmental Milestone is unlikely to receive differential feedback and scoring for the more advanced element.

The "Matrix" Method

The matrix method is illustrated in FIGURE 3, using the Trustworthiness Milestone as an example.[1] A milestone matrix aims to display a table of the specific elements of the developmental Milestones in rows, with the columns representing behavioral criteria assigned to progressively advancing developmental outcomes. In this scoring rubric, the assessor is able to identify developmental progression at the individual element level. Therefore, this method is ideal for developmental Milestones with elements that may progress at asynchronous rates or at different interval lengths, allowing feedback and scoring to reliably reflect each element rather than a cluster of elements. Using the Milestone in FIGURE 3 as an example, the matrix method is more appropriate than a clustered-response system, such as the slider-bar, because elements of overall ability, discernment, conscientiousness, and truthfulness may develop at different paces.

Proposed Reporting Tool

We propose the wheel or target graphic, illustrated in FIGURE 4, for reporting when progress for Milestone development varies. This tool is effective for reporting a series of slider-bar assessments (eg, a given set of subcompetencies within a competency domain) or reporting

Downloaded from https://prime-pdf-watermark.prime-prod.pubfactory.com/ at 2025-10-27 via free access

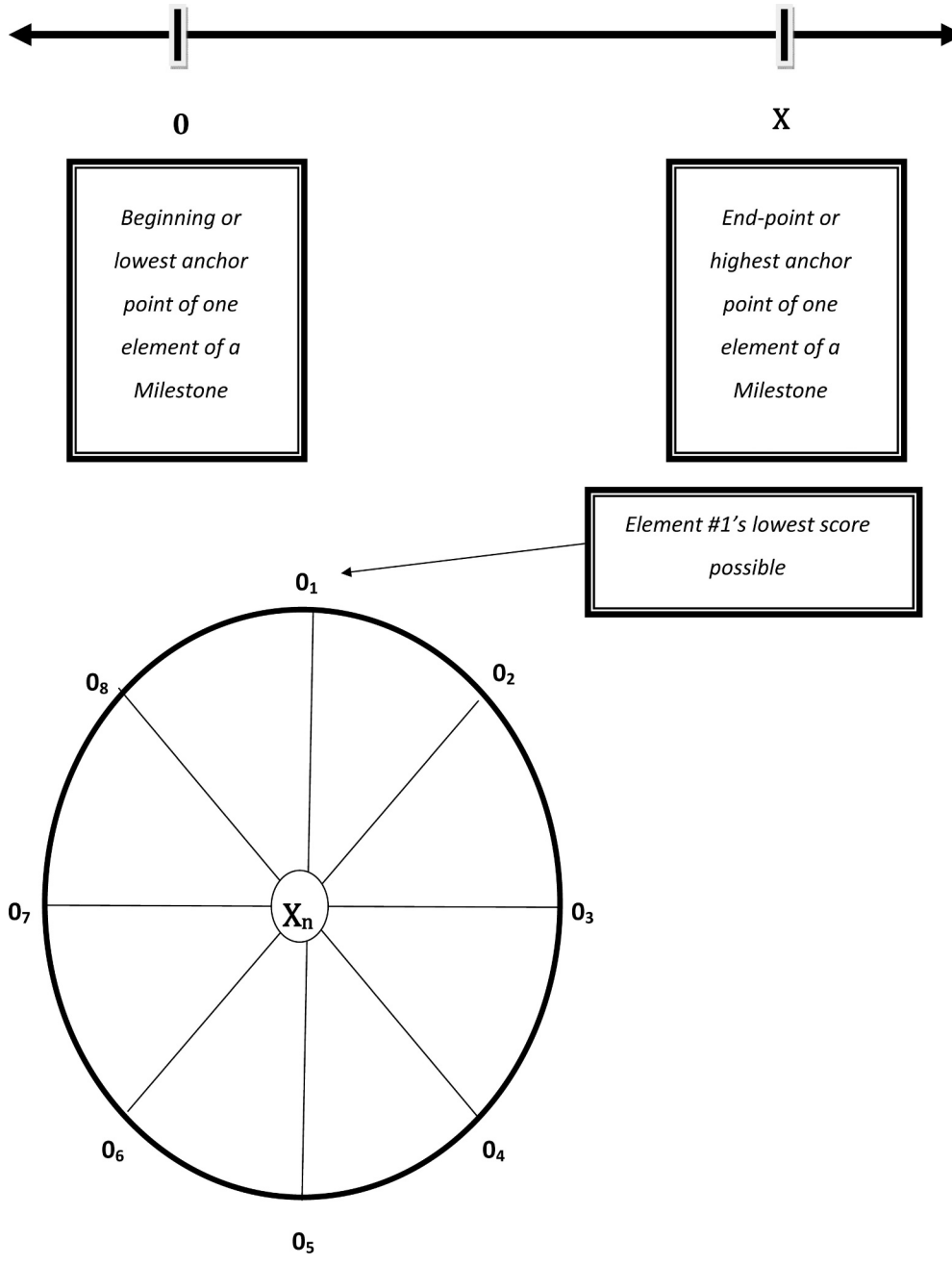

FIGURE 4 **Proposed Milestone Reporting: The Wheel or Target Graphic**

a single matrix displaying a series of elements within a given subcompetency. It plots the numerical score for each assessment on the spokes of the wheel, moving from the earliest learners along the periphery to a central target representing achievement of the most advanced learners. The advantages of this method for reporting performance are (1) the data for this report can be generated from the slider-bar or the matrix, (2) this method allows the scored elements to be reported in a simple visual that lets the learner know how close he or she is to the target, and (3) this report provides a method for visualizing either a learner's or a program's progress over time.

Validity and Reliability

Establishing validity evidence for data produced by the Milestones will require further study. The first step in establishing validity evidence is understanding possible threats to validity.[22,23] Messick[24] identifies 5 major threats to validity evidence: content and sampling errors, response

Downloaded from https://prime-pdf-watermark.prime-prod.pubfactory.com/ at 2025-10-27 via free access

process, internal structure and item performance, relationship to other variables, and consequences of the assessment.

Content and Sampling Errors

Sampling error occurs when the content chosen for the assessment does not represent the real-world setting or the intended construct of interest or does not reflect the relative distributive weight of the various content areas of the chosen construct. A sampling error would occur, for example, if we inadvertently did not include subcompetencies that are critical to training pediatric residents, such as gathering essential and accurate information about the patient, a subcompetency in the patient-care domain. In that case, relevant Milestones would be absent, and therefore, a gap would exist between important behaviors of practice and those of training. Blueprinting, a process of mapping the content of the real-world experience to an assessment tool,[25] is performed with careful attention to content representation and is a potential strategy to mitigate sampling errors.

Response Process

The second threat to the validity of an assessment tool is the effect of the assessment environment on the learner. Messick[24] calls this a *response process*. For example, a learner may perform differently in different contexts or situations. Assessing individual learners using the Milestones in various settings will help us understand the extent to which the Milestones are subject to this threat.

Internal Structure and Item Performance

The third threat to validity is the internal structure or the degree to which the Milestones represent the underlying constructs they are intended to measure. This includes internal consistency, a component of reliability, of specific items and how they perform in differentiating learners. Much future work is needed to establish interrater and intrarater reliability of the Milestones and to determine whether the developmental progression designed yields meaningful data regarding the real-world performance of learners.

Relationship to Other Variables

The relationship to other variables can be both a threat to and confirmation of the validity of an assessment. It will be important to study the relationship between the assessment using Milestones and the results from other established measures of performance. For example, evidence supporting the validity of the Milestones might include alignment of the outcome data for the medical knowledge Milestones with scores on American Board of Pediatrics In-Training Examinations; whereas, nonalignment of these data would threaten validity.

Consequences of the Assessment

The fifth threat to validity involves both the intended and unintended consequences of the assessment. These can include the implementation of the test itself, the reporting of results, the impact on curriculum and costs, and the other responses that are anticipated or unanticipated. An example of an unintended consequence threatening the validity of the Milestones is their use in broad and individual assessment and reporting too early in high-stakes assessment, such as accreditation or credentialing.

Utility

Another important concept to evaluate as we test and implement the Milestones is that of utility. Van der Vleuten[26] defines utility as a multiplicative function of reliability, validity, cost, practicality, and educational impact. If any one of these elements is absent or prohibitive, the overall utility is zero. In our efforts to provide assessment tools with high-validity evidence, we should not lose sight of the other critical variables in this model. The utility of an assessment looks beyond reliability and validity and considers the overall value or educational effect, weighed against the resources, costs, and acceptability required to achieve the assessment.[26]

Balancing Validity and Reliability: A Call for Assessment Across Multiple Contexts Using Multiple Methods

The consistency of resident performance across cases, or *intercase reliability*, is one of the most important aspects of performance assessment.[27] Because physicians do not perform consistently from task to task,[28–30] broad sampling across cases is essential to assess clinical competence reliably. This observation might not be surprising given the differences in individual experiences encountered during training and practice. However, it challenges the traditional approach to clinical competence testing, whereby the competence of individuals is assessed based on a single case, namely the case observed by the assessor.

It will also be important to compare the assessment data derived from the Pediatrics Milestone Project to that of other well-designed assessment methods. To be reliable, the Milestones should correlate with other methods designed to assess the same content (knowledge, attitudes, and skills). When considering this, it is important to note that the assessment method used for comparison should correspond to the nature of the content to be assessed. Kern and colleagues[31] provide guidance regarding strengths and limitations of a variety of assessment methods as those methods relate to the nature and type of content to be assessed. For example, if learner attitudes, feelings, descriptions of experiences, or perceived effects of experiences are to be assessed, the use of essays or narratives in a portfolio is rich in texture, provides unanticipated as well as anticipated information, and is respondent centered. However, essays and other learner-directed written responses are at risk of rater biases, are often subjective, and are often in low agreement with objective measurements. Consideration of the alignment of learning content, goals,

Downloaded from https://prime-pdf-watermark.prime-prod.pubfactory.com/ at 2025-10-27 via free access

and objectives with an assessment of the limitations and strengths of different assessment methods is important.

Standard Setting and Benchmarking

The term *standard setting* refers to a process that is used to create boundaries between categories that distinguish levels of performance. The work ahead in standard setting is 2-fold. First, the education community needs to identify sentinel Milestones, the accomplishment of which are requisite for assuming an advanced role, such as a supervisory resident or team leader. Second, the education community must collaborate to study when residents typically transition from one developmental Milestone to the next. Although development of competence is an individual learning curve, there is likely a range of time during which most learners will achieve a given developmental Milestone, much as there is an acceptable time during which most children achieve developmental Milestones in gross motor, fine motor, social, and language skills. Knowing these typical ranges will be helpful in identifying learners who could potentially accelerate through training or, conversely, require remediation.

The working group and the larger education community will also need to complete benchmarking of the Milestones. Benchmarks are descriptions of learners at each stage of development, as determined by the standard setting. Although the working group has developed the first iteration of the Milestones with developmental anchors that mirror benchmarks, refining the content of these developmental Milestones with data on the components represented through study with actual residents will be important to achieve true benchmarking.

The Pediatrics Milestones and Competency-Based Assessment

To promote meaningful learning, assessment should be educational and formative—residents should learn from assessments and receive feedback on gaps in their knowledge, skills, or attitudes so they can fill those gaps. The Pediatrics Milestones were constructed to be explicit, tangible, and meaningful descriptions of behaviors to provide a learning road map for physician development. Understanding how the Milestones fit in the overall context of assessment in competency-based medical education helps define how they contribute in a meaningful way to resident assessment.

The major challenge in implementing the ACGME Outcome Project to date has been assessment.[32] Lack of reliable and valid tools to measure these complex tasks, as well as a lack of faculty development in assessment, has led to a reductionist approach to assessment, whereby the competencies are broken down into discrete, observable behaviors which, at best, do not necessarily equal the whole when summed. A trainee may demonstrate all of the behaviors on a checklist for a given subcompetency and yet may not be able to integrate those behaviors to effectively care for a patient. In addition, the items on the checklist may not be attributes or skills that yield meaningful inference about performance but may be included because they are measurable. The Pediatrics Milestones move us from the realm of measuring what is easy and possibly meaningless to describing behaviors that are important to the professional formation of a physician. They clearly map to the subcompetencies and competencies from which they were derived but, instead of losing meaning through reduction, they embrace the complexity of the competencies and add meaning through explicit definitions of behaviors within their realm.

As an additional step to avoid the reductionist pitfall of the competencies, the Pediatrics Milestone Project Working Group will also take a step back and work to embed the Milestones in what has been described by ten Cate and Scheele[33] as *entrustable professional activities* (EPAs). These are the essential activities that define a given specialty. Framing the competencies in the context of these essential activities of a pediatrician puts them into the clinical realm in which we live, thereby adding meaning to them. Beginning with an EPA, such as caring for a healthy newborn, one can map it to the most relevant competencies and subcompetencies. One can then identify the Milestones within those subcompetencies at which a practitioner would be considered "entrustable"; that is, able to perform the professional activity without direct supervision. The aggregate of the Milestones at which one would be considered entrustable for a given professional activity, then, paint a behavioral picture of the learner ready to be entrusted.

Challenges Ahead

The public wants and deserves good medical care. A key element of good health outcomes is the proper training of physicians, with evidence of assessment-proven competency. We believe that to make the Pediatrics Milestone Project meaningful, the enormity of the work ahead, which we have outlined above, cannot be underestimated. Moving forward, human resources will need to include content experts for editing, setting standards, and striving toward test validity and reliability; preceptors for developing faculty; and experts in assessment for helping to develop the measurement tools that will be required to assess individuals and evaluate programs. Ensuring the availability of these resources has substantial financial implications. Absent the resources, the rate of Milestone testing will likely be so slow as to threaten their validity. Even with substantial support, the complexity of implementing the Milestones cannot be underestimated. Innovative approaches to assessment may be met with significant issues of feasibility, complex and varying challenges with implementation and program and user

Downloaded from https://prime-pdf-watermark.prime-prod.pubfactory.com/ at 2025-10-27 via free access

acceptability. Consideration for the further development and intervention of these innovative assessment methods will be informed by expert input.[34]

In addition to developing tools to assess Milestones, it will be critical for the education community to step back and evaluate whether the implementation of Milestones is moving along a trajectory to achieve desired outcomes. In light of the complexity and uncertainty of this unchartered territory, a developmental framework holds the most promise. This will require us to evaluate the overall process at each step along the way and make course corrections as we encounter both intended and unintended consequences.[35]

The vision of the Pediatrics Milestone Project is to understand the development of a pediatrician from entry into medical school through the twilight of a physician's career. For this vision to be realized, the work ahead will need to be owned by the undergraduate and graduate medical education communities, as well as by our accrediting and certifying bodies, in partnership with our working group. Only through this broad collaboration can we hope to realize the public's vision of improved health care outcomes.

References

1 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.
2 Bordage G. Prototypes and semantic qualifiers: from past to present. *Med Educ*. 2007;41(12):1117–1121.
3 Coderre S, Mandin H, Harasym PH, Fick GH. Diagnostic reasoning strategies and diagnostic success. *Med Educ*. 2003;37(8):695–703.
4 Bowen JL. Educational strategies to promote clinical diagnostic reasoning. *N Engl J Med*. 2006;355(21):2217–2225.
5 Schmidt HG, Rikers RM. How expertise develops in medicine: knowledge encapsulation and illness script formation. *Med Educ*. 2007;41(12):1133–1139.
6 Schmidt HG, Boshuizen HPA. On acquiring expertise in medicine. *Educ Psychol Rev*. 1993;5(3):205–221.
7 Schmidt HG, Norman GR, Boshulzen HPA. A cognitive perspective on medical expertise: theory and implications. *Acad Med*. 1990;65(10):611–621.
8 Livingston SA, Zieky MJ. *Passing Scores: A Manual for Setting Standards of Performance on Educational and Occupational Tests*. Princeton, NJ: Educational Testing Service; 1982.
9 Wood TJ, Humphrey-Murto SM, Norman GR. Standard setting in a small scale OSCE: a comparison of the Modified Borderline-Group Method and the Borderline Regression Method. *Adv Health Sci Educ Theory Pract*. 2006;11(2):115–122.
10 Wilkinson TJ, Newble DI, Frampton CM. Standard setting in an objective structured clinical examination: use of global ratings of borderline performance to determine the passing score. *Med Educ*. 2001;35(11):1043–1049.
11 van der Vleuten CPM, Swanson DB. Assessment of clinical skills with standardized patients: state of the art. *Teach Learn Med*. 1990;2(2):58–76.
12 Tamblyn R. Outcomes in medical education: what is the standard and outcome of care delivered by our graduates? *Adv Health Sci Educ Theory Pract*. 1999;4(1):9–25.
13 Talente GA, Haist SA, Wilson JF. A model for setting performance standards for standardized patient examinations. *Eval Health Prof*. 2003;26(4):427–446.
14 Plake BS, Hambleton RK. The analytic judgment method for setting standards on complex performance assessments. In: Cizek GJ, ed. *Setting Performance Standards: Concepts, Methods and Perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates; 2001:283–312.
15 Humphrey-Murro S, MacFadyen JC. Standard setting: a comparison of case-author and modified borderline-group methods in a small-scale OSCE. *Acad Med*. 2002;77(7):729–732.
16 Hulsman RL, Mollema ED, Oort FJ, Hoos AM, deHaes JC. Using standardized video cases for assessment of medical communication skills: reliability of an objective structured video examination by computer. *Patient Educ Couns*. 2006;60(1):24–31.
17 Holstein BL, Zangrilli BF, Taboas P. Standardized testing tools to support quality educational outcomes. *Qual Manag Health Care*. 2006;15(4):300–308.
18 Hambleton RK, Jaeger RM, Plake BS, Mills C. Setting performance standards on complex educational assessments. *Appl Psychol Meas*. 2000;24(4):355–366.
19 Cizek CJ, ed. *Setting Performance Standards: Concepts, Methods and Perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc; 2001.
20 Boursicot KA, Roberts TE, Pell G. Standard setting for clinical competence at graduation from medical school: a comparison of passing scores across five medical schools. *Adv Health Sci Educ Theory Pract*. 2006;11(2):173–183.
21 Downing SM. Reliability: on the reproducibility of assessment data. *Med Educ*. 2004;38(9):1006–1012.
22 Downing SM, Haladyna TM. Validity threats: overcoming interference with proposed interpretations of assessment data. *Med Educ*. 2004;38(3):327–333.
23 Downing SM, Haladyna TM. Validity and it's threats. In: Downing SM, Yudkowsky R, eds. *Assessment in Health Professions Education*. New York, NY: Routledge; 2009:21–56.
24 Messick S. Validity. In: Linn RL, ed. *Educational Measurement*. 3rd ed. New York, NY: American Council on Education & Macmillan; 1989:13–103.
25 Coderre S, Woloschuk W, McLaughlin K. Twelve tips for blueprinting. *Medical Teach*. 2009;31(4):359–361.
26 van der Vleuten CPM. The assessment of professional competence: developments, research, and practical implications. *Adv Health Sci Educ*. 1996;1(1):41–67.
27 Wass V, van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *Lancet*. 2001;357(9260):945–949.
28 Norman G, Bordage G, Page G, Keane D. How specific is case specificity? *Med Educ*. 2006;40(7):618–623.
29 Swanson DB, Norman GR, Linn RL. Performance-based assessment: lessons from the health professions. *Educ Res*. 1995;24(5):5–11.
30 Eva KW. On the generality of specificity. *Med Educ*. 2003;37(7):587–588.
31 Kern DE, Thomas PA, Howard DM, Bass EB. *Curriculum Development for Medical Education: A Six Step Approach*. Baltimore, MD: Johns Hopkins University Press; 1998.
32 Lurie SJ, Money CJ, Lyness JM. Measurement of the general competencies of the accreditation council for graduate medical education: a systematic review. *Acad Med*. 2009;84(3):301–309.
33 ten Cate O, Scheele F. Competency-based postgraduate training: can we bridge the gap between theory and clinical practice? *Acad Med*. 2007;82(6):542–547.
34 Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M; Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ*. 2008;337(a1655):979–983.
35 Patton MQ. *How to Use Qualitative Methods in Evaluation*. Newbury Park, CA: Sage; 1987.
36 Kennedy TJ, Rehehr G, Baker GR, Lingard L. Point-of-care assessment of medical trainee competence for independent clinical work. *Acad Med*. 2008;83(10 suppl):S89–S92.
37 Farnan JM, Johnson JK, Meltzer DO, Humphrey HJ, Arora VM. Resident uncertainty in clinical decision making and impact on patient care: a qualitative study. *Qual Saf Health Care*. 2008;17(2):122–126.

Downloaded from https://prime-pdf-watermark.prime-prod.pubfactory.com/ at 2025-10-27 via free access