

## EDUCATIONAL ADVANCE

# Programmatic Assessment of Level 1 Milestones in Incoming Interns

Samantha R. Hauff, MD, Laura R. Hopson, MD, Eve Losman, MD, Marcia A. Perry, MD, Monica L. Lypson, MD, MHPE, Jonathan Fischer, MPH, and Sally A. Santen, MD, PhD

## Abstract

**Objectives:** With the Accreditation Council for Graduate Medical Education (ACGME) Next Accreditation System, emergency medicine (EM) residency programs will be required to report residents' progress through the EM milestones. The milestones include five progressively advancing skill levels, with Level 1 defining the skill set of a medical school graduate and Level 5, that of an attending physician. The ACGME stresses that multiple forms of assessment should be used to ensure capture of the multifaceted competencies. The objective of this study was to determine the feasibility and results of programmatic assessment of Level 1 milestones using multisource assessments for incoming EM interns in July.

**Methods:** The study population was interns starting in 2012 and 2013. Interns' Level 1 milestone assessment was done with four distinct methods: 1) the postgraduate orientation assessment (POA) by the Graduate Medical Education Office for all incoming interns (this multistation examination covers nine of the EM milestones and includes standardized patient cases, task completion, and computer-based stations); 2) direct observation of patient encounters by core faculty using a milestones-based clinical skills competency checklist; 3) the global monthly assessment at the end of the intern orientation month that was updated to reflect the EM milestones; and 4) faculty assessment during procedural labs. These occurred during the July orientation month that included the POA, clinical shifts, didactic sessions, and procedure labs.

**Results:** In the POA, interns were competent in 48% to 93% of the milestones assessed. Overall, competency was 70% to 80%, with low scores noted in aseptic technique (patient care Milestone 13 [PC13]) and written and verbal hand-off (interpersonal communications skills [ICS]2). In overall communication, 70% of interns demonstrated competency. In excess of 80% demonstrated competency in critical values interpretation (PC3), informed consent (PC9), pain assessment (PC11), and geriatric functional assessment (PC3). On direct observation, almost all Level 1 milestones were achieved (93% to 100%); however, only 78% of interns achieved competency in pharmacotherapy (PC5). On global monthly evaluations, all interns met Level 1 milestones.

**Conclusions:** A multisource assessment of EM milestones is feasible and useful to determine Level 1 milestones achievement for incoming interns. A structured assessment program, used in conjunction with more traditional forms of evaluation such as global monthly evaluations and direct observation, is useful for identifying deficits in new trainees and may be able inform the creation of early intervention programs.

ACADEMIC EMERGENCY MEDICINE 2014;21:694-698 © 2014 by the Society for Academic Emergency Medicine

Emergency medicine (EM) residency programs are experiencing a historic time of change. The EM milestones were released in July 2012 as part of a joint effort by the Accreditation Council for

Graduate Medical Education (ACGME) and the American Board of Emergency Medicine and brought about a dramatic shift in how residents are assessed.<sup>1,2</sup> The milestones are intended to improve graduate medical

From the Department of Emergency Medicine (SRH, LRH, EL, MAP, SAS), the Department of Internal Medicine (ML), and the Department of Health Management and Policy School of Public Health (JF), University of Michigan, Ann Arbor, MI.

Received December 14, 2013; revision received January 16, 2014; accepted January 19, 2014.

Presented at the Council of Emergency Medicine Residency Directors Academic Assembly, Denver, CO, March 2013; and the Society for Academic Emergency Medicine Annual Meeting, Atlanta, GA, May 2013.

The authors have no relevant financial information or potential conflicts of interest to disclose.

Supervising Editor: John Burton, MD.

Address for correspondence and reprints: Sally A. Santen, MD, PhD; e-mail: ssanten@med.umich.edu.

education by enhancing competency-based education and assessment. They are “stepping stones” to competency, and include five progressively advancing skill levels, with Level 1 defining the skill level of a medical student graduate and Level 5 that of an attending physician.

Incoming interns are expected to have achieved Level 1 milestones by graduation from medical school, but a surprising number of EM interns struggle to meet this level of expertise.<sup>3</sup> Not all medical schools have mandatory EM rotations. This, in addition to the limited amount of time interns spend in the emergency department (ED) during their internship year, as they spend the majority of months on off-service rotations, makes assessment of Level 1 milestones especially difficult. Although preparation for Level 1 milestones takes place during medical school, it is the responsibility of EM residency directors to confirm that these milestones have been met. Thus, residency programs must accurately assess intern competency at arrival, so that necessary remediation of these deficiencies can occur as early as possible. The milestones can only have their intended effect if residency programs use high-quality assessment tools that accurately measure resident competency.<sup>4</sup>

The ACGME stresses that multiple forms of assessment should be used to ensure that the complexities of context are captured.<sup>5</sup> Further, the *Academic Emergency Medicine* consensus conference “Education Research in Emergency Medicine: Opportunities, Challenges, and Strategies for Success” highlighted the importance of educational research, especially in the area of assessment of competencies.<sup>6–8</sup> Direct observation, global assessment, and objective structured clinical examinations (OSCE) are common tools used for resident evaluation.<sup>5,9</sup> While medical educators struggle to determine the utility of individual assessment tools,<sup>7,8,10</sup> van der Vleuten and Schuwirth’s<sup>11</sup> work suggests that it is most important to develop a program of assessment with multiple low-stakes assessments. The tools used in multiple low-stakes assessments provide educators with a well-rounded view of learners’ performance. The objective of this study was to determine the feasibility and results of programmatic assessment of Level 1 milestones using multisource assessments for incoming EM interns in July.

## METHODS

### Study Design

This was a prospective, descriptive analysis of the use of multiple assessment tools for EM milestones that was determined by the institutional review board to be exempt from informed consent requirements.

### Study Setting and Population

The study was conducted at a large EM residency program with 28 interns (July 2012 and 2013), representing 19 Liaison Committee on Medical Education–accredited medical schools.

### Study Protocol

Assessment of milestones was done by four distinct methods: 1) the postgraduate orientation assessment

(POA);<sup>12,13</sup> 2) direct observation, using a milestone-based clinical skills assessment (CSA, Data Supplement S1, available as supporting information in the online version of this paper) performed by core education faculty; 3) global milestone assessment at the end of the intern orientation month complete by faculty members; and 4) faculty assessment during procedural labs. These occurred during the July intern orientation month that includes the POA, clinical shifts, didactic sessions, and procedure labs.

**POA.** Entering residents have variable medical school experiences and differing knowledge and skill levels. To structure curricula, enhance patient safety, and begin to meet competency-based accreditation requirements, baseline assessment of individual resident’s knowledge and skills is needed. To this end, in 2002 the University of Michigan created the POA, a 10-station OSCE for incoming residents. This assessment has built validity evidence through careful design and scholarship, and the scores have been shown to correlate with our residents’ performance.<sup>12,13</sup>

**Milestone-based CSA.** The milestone-based CSA uses direct observation performed by core education faculty. During the orientation month, each intern was observed twice by core faculty and the MS-CSA was completed (Data Supplement S1). The content validity of this instrument was derived from using the wording of the EM Level 1 milestones. Faculty scored each intern on whether the Level 1 milestone was achieved. Prior to the assessment, the core faculty had discussions about the Level 1 milestones, which formed the basis of faculty training on these new competencies.

**Global Milestones.** The monthly global evaluation form was based on the EM milestones and was completed by faculty working with the interns at the end of the month. The residency uses online software through MedHub (<http://medhub.com/>). On the form, all of the milestones are represented for faculty to assess residents. A composite score was created for each resident.

**Procedure Labs.** Finally, there were multiple procedure labs during the intern orientation month including airway, ultrasound, wound management, and simulation-based vascular access. Determination of the Level 1 milestones in these labs was assessed by a checklist and the successful completion of the procedures during the lab.

### Data Analysis

Simple descriptive statistics are reported.

## RESULTS

Most of the non-procedure-based EM Level 1 milestones were assessed by multiple modalities (Table 1). The monthly global evaluation and procedure labs found that 100% of interns met Level 1 milestones. However, the POA was more discriminating; EM interns were competent in 48% to 93% of milestones evaluated. Overall competency was 70% to 80%, with low scores noted in aseptic technique (48%) (patient

Table 1  
Intern Competency in Level 1 Milestones as Assessed by Different Modalities

Milestone	Postgraduate Orientation Assessment	Direct Observation, %	Monthly Global Evaluation, %	Procedure Lab, %
PC1 Emergency stabilization	X	96	100	X
PC2 Focused H&P	89% GFA	96	100	X
PC3 Diagnostic	93% Critical values 78% Imaging	96	100	X
PC4 Diagnosis	X	93	100	X
PC5 Pharmacotherapy	X	78	100	X
PC6 Observation, reassessment	63% Verbal hand-off 48% Written hand-off	X	X	X
PC7 Disposition	X	96	100	X
PC8 Multitasking	X	X	100	X
PC9 Procedures	48% Aseptic technique	X	100	X
PC10 Airway management	X	X	X	100
PC11 Anesthesia, pain	93% Pain assessment	X	X	X
PC12 Ultrasound	X	X	X	100
PC13 Wound management	48% Aseptic technique	X	X	100
PC14 Vascular access	X	X	X	100
MK Medical knowledge	93% Critical values	X	100	X
PROF1 Professional values	89% Consent 93% Pain assessment 89% GFA 74% Cultural competency	100	100	X
PROF2 Accountability	X	X	100	X
ICS1 Patient-centered communication	89% Consent 89% GFA 74% Cultural competency	100	100	X
ICS2 Team management	63% Verbal hand-off 48% Written hand-off	X	X	X
PBLI Performance improvement	X	96	100	X
SBP1 Patient safety	48% Aseptic technique 93% Hand washing 63% Verbal hand-off 48% Written hand-off	93	X	X
SBP2 Systems-based management	48% Aseptic technique	96	X	X
SBP3 Technology	X	100	X	X

GFA = geriatric functional assessment; H&P = history and physical; ICS = interpersonal communication skills; PBLI = practice-based learning and improvement; PC = patient care; PROF = professionalism; SBP = systems-based practice; X = not applicable.

care [PC]9, PC13, systems-based practice [SBP]1, SBP2), as well as written (48%) and verbal hand-off (63%) (PC6, interpersonal communications skills [ICS]2, SBP1). In overall communication, 70% of interns demonstrated competency. In excess of 80% achieved competency in informed consent (professionalism [PROF]1, ICS1), critical value interpretation (PC3, medical knowledge [MK]), pain assessment (PC11, PROF1), hand washing (SBP1), and geriatric functional assessment (PC2, PROF1, ICS1). On direct observation, almost all Level 1 milestones were achieved (93% to 100%); however, only 78% of interns achieved competency in pharmacotherapy (PC5).

## DISCUSSION

This pilot study shows that multisource assessment of incoming interns is both feasible and useful in determining Level 1 milestones. While global assessments and procedure labs found 100% Level 1 milestone competency, other methods in differing contexts were more discriminating. This study also demonstrates that many interns have not attained full Level 1 milestones competency by commencement of residency. These are expected to have been taught and assessed during

medical school. This gap poses a unique challenge for EM residency programs and emphasizes the importance of increased collaboration between medical schools and residency programs.

The path forward on milestone assessment of competency is undefined.<sup>7,8,10</sup> Our findings confirm the study by Santen and colleagues<sup>3</sup> showing that incoming EM interns were not taught or assessed on many of the Level 1 milestones. They noted that in the future, medical schools will need to be responsible for cross-cutting milestones that span across specialties, with individual clerkships needing to take responsibility for educating senior students in specialty-specific milestones. This will ensure that incoming interns are fully equipped to begin successful residency experiences. Nonetheless, it is critical for program directors to know where the residents are starting and where they may get into difficulty. This study was supported by the literature showing that the majority of problem residents are identified during internship and not during medical school.<sup>14</sup> Thus residencies should assess interns early in their training using a multisource assessment to determine their competency level. Early identification of struggling residents is associated with a higher likelihood of successful remediation.

Our study revealed interesting strengths and drawbacks of each method of assessment. Direct observation is a vital component of competency assessment. This workplace-based assessment provides a richness of information and provides the insight and close engagement of practicing clinicians.<sup>15-18</sup> Further, it is workplace-based and thus may have more content validity because it is measuring true work. However, direct observation provides a specific assessment of one encounter and as such is subject to pitfalls of context, lack of standardization, rater training, and bias. In this study, direct observation yielded high levels of competency.

The global monthly evaluation completed at the end of the rotation has the advantage of being authentic as it is workplace-based and is the type of resident evaluation that has been used by our program, as well as many others. On the other hand, it is at risk for hindsight bias and lack of specificity, as faculty may not remember the details weeks after the performance. This global “gestalt” may be a less sensitive metric and may not identify where additional skill and knowledge development are needed.<sup>16,17</sup> We observed that all interns were deemed competent at Level 1 for the global assessment, even those who did not achieve competency on other instruments.

In both workplace-based assessments, there is an element of social acceptability, meaning that there is a relationship with and commitment on the part of the faculty to the individual intern.<sup>16,17</sup> As a result, faculty may want to encourage interns to succeed and do well and thereby overrate performance. This may result in scoring interns as competent when they may not be. This is especially true if faculty assessments are not anonymous or the results are shared immediately with the learner. On the other hand, they have the advantage of providing immediate formative feedback for improvement.

The POA is a robust, comprehensive, standardized, and structured assessment and detected more areas of deficit.<sup>13</sup> The advantages are that the assessments are consistent and standardized and that there is a lack of social acceptability. Limitations of a standardized examination are that the artificial environment is less authentic and it requires attention to rater training and potential rater bias. While the POA is unique to our institution, assessment with an OSCE is widespread in medical student and some residency programs.<sup>19-22</sup> Further, generalizability of the OSCE is dependent on the number of stations.<sup>23</sup>

By combining all of the assessment tools, however, a sufficiently reasonable, comprehensive depiction of intern Level 1 milestones competency emerges. This finding supports the assertion of van der Vleuten and Schuwirth<sup>11</sup> that a single assessment tool cannot measure all aspects of a competency. Instead, multiple assessment methods that are well integrated into education curricula provide the most valid information. Many assessment methods do not have inherently reliable or valid characteristics, and different methods are better suited to certain contexts.<sup>11</sup> Thus, it may not be the characteristics of an assessment tool that determine its utility as much as how that tool is used. Van der Vleuten and Schuwirth note that multiple, context-specific, low-stakes observations are needed to create an overall assessment. In light of this, emphasis is best placed on

designing a multisource evaluation program as a whole and not on evaluation of individual tools.

These results also provide insight into the role of assessments for formative and summative purposes. Evaluation programs should be both formative (assessment for learning) and summative (assessment of learning).<sup>24</sup> The combination of many individual formative assessments as part of a comprehensive program, however, can be used effectively for summative purposes. Our program of assessment of intern milestones highlights this idea, as each individual tool provides unique formative information to students, while the assessments as a whole provide useful summative information to the program.

## LIMITATIONS

While the data for this study came from one residency program, that homogeneity is attenuated by the interns being observed across three different clinical sites by a diverse faculty body and that the interns come from a variety of medical schools. We do not know, however, if our population is representative of interns at other programs. The ACGME milestones were created recently and therefore time did not allow for a multi-institutional study. As such, generalizability to other institutions may be limited.

The study population was small enough that it was not possible to determine whether poor performance as measured by one assessment tool correlated across modalities. What constitutes an appropriate assessment of each milestone is a judgment and represents the best determination by authors.

## CONCLUSIONS

It is critical that emergency medicine program directors recognize early where their interns might need focused improvement. A structured assessment program, used in conjunction with more traditional forms of evaluation such as global monthly evaluations and direct observation, is useful for identifying deficits in new trainees and can inform the creation of early intervention programs, as well as tailor learning programs to both areas of strength and those that need developing. Many of the current resources used to determine the competency of our interns may be used, without any need to develop elaborate new programs. Most importantly, residencies need to design a program of multisource assessment for the intern Level 1 milestones.

The authors acknowledge James Mattimore, MD, and the faculty who contribute to the development of trainees

## References

1. Nasca TJ, Philibert I, Brigham T, Flynn TC. The Next GME Accreditation System—rationale and benefits. *NEJM* 2012;366:1051-6.
2. Accreditation Council for Graduate Medical Education. The Emergency Medicine Milestone Project. Available at: <http://www.acgme.org/acgmeweb/Portals/0/PDFs/Milestones/EmergencyMedicineMilestones.pdf>. Accessed Mar 17, 2014.

3. Santen SA, Rademacher N, Heron SL, Khandelwal S, Hauff S, Hopson L. How competent are emergency medicine interns for Level 1 milestones: who is responsible? *Acad Emerg Med* 2013;20:736–9.
4. Ankel F, Franzen D, Frank J. Milestones: Quo Vadis? *Acad Emerg Med* 2013;20:749–50.
5. Holt KD, Miller RS, Nasca TJ. Residency programs' evaluations of the competencies: data provided to the ACGME about types of assessments used by programs. *J Grad Med Educ* 2010;2:649–55.
6. Santen SA, Deiorio NM, Gruppen LD. Medical education research in the context of translational science. *Acad Emerg Med* 2012;19:1323–7.
7. Takayesu JK, Kulstad C, Wallenstein J, et al. Assessing patient care: summary of the breakout group on assessment of observable learner performance. *Acad Emerg Med* 2012;19:1379–89.
8. Chan TM, Wallner C, Swoboda TK, Leone KA, Kessler C. Assessing interpersonal and communication skills in emergency medicine. *Acad Emerg Med* 2012;19:1390–402.
9. Lurie SJ, Mooney CJ, Lyness JM. Measurement of the general competencies of the Accreditation Council for Graduate Medical Education: a systematic review. *Acad Med* 2009;84:301–9.
10. Kessler CS, Leone KA. The current state of core competency assessment in emergency medicine and a future research agenda: recommendations of the working group on assessment of observable learner performance. *Acad Emerg Med* 2012;19:1354–9.
11. van der Vleuten C, Schuwirth L. Assessing professional competence: from methods to programmes. *Med Educ* 2005;39:309–17.
12. Lypson ML, Downing SM, Gruppen LD, Yudkowsky R. Applying the bookmark method to medical education: standard setting for an aseptic technique station. *Med Teach* 2013;35:581–5.
13. Lypson ML, Frohna JG, Gruppen LD, Woolliscroft JO. Assessing residents' competencies at baseline: identifying the gaps. *Academic medicine special themes: educating for competencies. Acad Med* 2004;79:564–70.
14. Williams RG, Roberts NK, Schwind CJ, Dunnington GL. The nature of general surgery resident performance problems. *Surgery* 2009;145:651–8.
15. ten Cate OP, Scheele FP. Viewpoint: competency-based postgraduate training: can we bridge the gap between theory and clinical practice? *Acad Med* 2007;82:542–7.
16. Marks M, Humphrey-Murto S. Performance assessment. In: Dent JA, Harden RM (editors). *A Practical Guide for Medical Teachers*. Oxford, UK: Churchill Livingstone Elsevier, 2005:323–35.
17. Friedman Ben-David M. Principles of Assessment. In: Dent JA, Harden RM (editors). *A Practice Guide for Medical Teachers*. Oxford, UK: Churchill Livingstone Elsevier, 2005, pp 282–92.
18. Holmboe ES. Direct Observation by Faculty. In: Holmboe ES, Hawkins RE (editors). *Practical Guide to the Evaluation of Clinical Competence*: Mosby, 2008, pp 119–29.
19. Wagner D, Hoppe R, Lee CP. A patient safety OSCE for PGY-1 residents: a centralized response to the challenge of culture change. *Teach Learn Med* 2009;20:8–14.
20. Yudkowsky R, Alseidi A, Cintron J. Beyond fulfilling the core competencies: an objective structured clinical examination to assess communication and interpersonal skills in a surgical residency. *Curr Surg* 2004;6:499–503.
21. Ben-Isaac E, Keefer M, Thompson M. Modified Objective Structured Clinical Examination for Pediatric Interns. Available at: <https://www.meded-portal.org/publication/9488>. Accessed Mar 17, 2014.
22. Short MW, Jorgensen JE, Edwards JA, Blankenship RB, Roth BJ. Assessing intern core competencies with an objective structured clinical examination. *J Grad Med Educ* 2009;1:30–6.
23. Hawkins RE, Boulet JR. Direct Observation: Standardized Patients. In: Holmboe ES, Hawkins RE, (editors). *Practical Guide to Evaluation of Clinical Competence*. Philadelphia, PA: Mosby Elsevier, 2008:102–18.
24. Bok HG, Teunissen PW, Favier RP, et al. Programmatic assessment of competency-based workplace learning: when theory meets practice. *BMC Med Educ* 2013;13:123.

### Supporting Information

The following supporting information is available in the online version of this paper:

**Data Supplement S1.** Assessment of Level 1 milestones of incoming interns direct observation of clinical skills by faculty (formerly CSA).