

## BREAKOUT SESSION

# 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

Chad S. Kessler, MD, MHPE, and Katrina A. Leone, MD

### Abstract

In 2012, the Accreditation Council for Graduate Medical Education (ACGME) introduced the Next Accreditation System (NAS) for residency program accreditation. With implementation of the NAS, residents are assessed according to a series of new emergency medicine (EM)-specific performance milestones, and the frequency of assessment reporting is increased. These changes are driving the development of new assessment tools for the NAS that can be feasibly implemented by EM residency programs and that produce valid and reliable assessment data. This article summarizes the recommendations of the writing group on assessment of observable learner performance at the 2012 *Academic Emergency Medicine* consensus conference on education research in EM that took place on May 9, 2012, in Chicago, Illinois. The authors define an agenda for future assessment tool research and development that was arrived at by consensus during the conference.

ACADEMIC EMERGENCY MEDICINE 2012; 19:1354–1359 © 2012 by the Society for Academic Emergency Medicine

In 1999 the Accreditation Council for Graduate Medical Education (ACGME) introduced the Outcomes Project,<sup>1</sup> a multiyear process to accredit residency programs based on the assessment of individual resident performance within a framework of six core competency domains: 1) patient care, 2) medical knowledge, 3) practice-based learning and improvement (PBLI), 4) interpersonal and communication skills (ICS), 5)

professionalism, and 6) systems-based practice (SBP). Since 2001, the medical education community has passed through the implementation phases of the Outcomes Project and now routinely assesses learners according to this framework.<sup>2</sup>

In 2012 the ACGME introduced the Next Accreditation System<sup>3</sup> (NAS), which builds on the principles of the Outcomes Project by defining a continuum of performance milestones that culminate in full achievement of competency in each domain (Table 1). Emergency medicine (EM) is an early adopter of the NAS and will begin program accreditation according to this framework in 2013. The NAS differs from the previous accreditation system by requiring more frequent collection and biannual submission of resident assessment data, while reducing the frequency of formal site visits. Because both the assessment standards (milestones) and the frequency of reporting of resident assessment will be changing with the implementation of the NAS, there exists an imperative to develop assessment tools that can feasibly be implemented by multiple residency programs and produce valid and reliable assessment data.

This article summarizes the recommendations of the breakout group on assessment of observable learner performance at the 2012 *Academic Emergency Medicine*

From the Department of Emergency Medicine, Jesse Brown VA Medical Center (CSK), Chicago, IL; and the Department of Emergency Medicine, Oregon Health and Science University (KAL), Portland, OR.

Received July 1, 2012; accepted July 2, 2012.

A full list of breakout session participants is available in Appendix A.

This paper reports on a workshop session of the 2012 *Academic Emergency Medicine* consensus conference, "Education Research in Emergency Medicine: Opportunities, Challenges, and Strategies for Success," May 9, 2012, Chicago, IL.

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

Supervising Editor: John Burton, MD.

Address for correspondence and reprints: Katrina Leone, MD; e-mail: katrina.leone@gmail.com.

**Table 1**  
Comparison of ACGME Core Competency Domains and NAS Milestones for EM

ACGME Core Competency Domain	NAS Milestones for EM*
Patient care	PC-1 Emergency stabilization PC-2 Performance of focused history and physical exam PC-3 Diagnostic studies PC-4 Diagnosis PC-5 Pharmacotherapy PC-6 Observation and reassessment PC-7 Disposition PC-8 Multitasking (task-switching) PC-9 General approach to procedures PC-10 Airway management PC-11 Anesthesia and acute pain management PC-12 Other diagnostic and therapeutic procedures: ultrasound PC-13 Other diagnostic and therapeutic procedures: wound management PC-14 Other diagnostic and therapeutic procedures: vascular access
Medical knowledge PBLI	MK Medical knowledge PBLI-1 Teaching PBLI-2 Practice-based performance improvement
ICS	ICS-1 Patient-centered communication ICS-2 Team management
Professionalism	P-1 Professional values P-2 Accountability
SBP	SBP-1 Patient safety SBP-2 Systems-based management

ACGME = Accreditation Council for Graduate Medical Education; ICS = Interpersonal and communication skills; NAS = New Accreditation System; PC = patient care; MK = medical knowledge; P = professionalism; PBLI = practice-based learning and improvement; SBP = systems-based practice.  
\*Council of Emergency Medicine Residency Directors. Academic Assembly 2012. Atlanta, GA, April 2012.

consensus conference “Education Research in Emergency Medicine: Opportunities, Challenges, and Strategies for Success,” that took place on May 9, 2012, in Chicago, Illinois. We define an agenda for future assessment tool research and development that was arrived at by consensus during the conference.

**THE CONSENSUS BUILDING PROCESS**

In preparation for the consensus conference, our group carried out an extensive review of the literature on tools for evaluation of the six ACGME core competency domains. Writing group members focused literature searches on assessment tools testing the six ACGME core competency domains, so the majority of identified studies pertain to resident learners. Increasingly, undergraduate medical education and maintenance of certification for practicing physicians also utilize aspects of the core competency framework, so future research on assessment tools may ultimately have applicability to these learner populations as well, but these areas were not a primary focus for this article. Similarly, because the ACGME has jurisdiction to accredit residency and

fellowship programs in the United States, much of the literature reviewed for this article comes from this country. We acknowledge that important advances in assessment tools are also taking place in other countries, so we specifically sought out literature on assessment tools developed, tested, and implemented in other English-speaking countries in the hopes of identifying novel tools with applicability in the United States. Of particular note, The Royal College of Physicians and Surgeons of Canada has organized the core knowledge, skills, and abilities of physicians into the CanMEDs Physician Competency Framework.<sup>4</sup> Learners in Canadian medical schools and residency training programs, as well as physicians performing maintenance of certification, are assessed according to this framework. CanMEDs roles have similarity with several ACGME core competencies.

The writing group’s opinions on the strengths and weaknesses of the assessment tool literature for each competency were presented to participants of a breakout session during the consensus conference. The primary aim of the breakout session was to review the current state of assessment tool literature for the ACGME core competencies and create an agenda for future research on assessment tools relevant to the transition to the NAS. An open discussion on each research agenda item was conducted with the 77 participants of the breakout session (Appendix A). A summary of the proposed research agenda for each core competency domain is described in the next six articles in this issue.<sup>5-10</sup>

At the conclusion of the conference day, nearly 80 participants used an audience response system to submit answers to two key questions developed during our breakout session (Table 2). The goal of the first question was to determine which of the core competency domains are most difficult to assess with our current

**Table 2**  
Key Questions Presented for Consensus Building

<i>Question 1: Which ACGME core competency has the most significant assessment tool gap?</i>	
Core competency domain	Percentage of respondents
Medical knowledge	0
Patient care	16
Professionalism	15
ICS	11
PBLI	27
SBP	31

  

<i>Question 2: Can we use assessment tools from other specialties, or unstudied tools, on EM learners in the emergency department setting?</i>	
Response	Percentage of respondents
Yes	71
No	29

N for both approximately 80 respondents.  
ICS = interpersonal and communication skills;  
PBLI = practice-based learning and improvement; SBP = systems-based practice.

tools. The audience indicated that SBP and PBLI are the two core competency domains where quality assessment tools are lacking. The medical knowledge competency domain was placed at last priority, with no conference attendees indicating that assessment tools for this competency are lacking. Our interpretation of this finding is that the conference attendees feel comfortable with our current assessment tools for medical knowledge. Despite this finding, future research on medical knowledge assessment tools is suggested in our research agenda, but this may take a lower priority than the development and implementation of assessment tools for other domains.

The goal of the second question was to determine how strongly the conference participants believe that assessment tools must be studied specifically on EM residents and in an ED setting prior to implementation. Seventy-one percent of respondents indicated that assessment tools developed for the assessment of residents in other specialties, or untested tools, could be used for the assessment of EM residents in an ED setting. We interpret this response to mean that the conference participants feel pressure to utilize untested and poorly validated tools to meet ACGME assessment and reporting mandates. With time and increased collection of validity evidence supporting the use of multiple tools for the assessment of EM residents, we expect that the response to this question will change to indicate that only well-studied tools should be utilized.

## **GENERAL RECOMMENDATIONS**

The ACGME core competencies and associated NAS milestones can be evaluated with multiple assessment tools. Similarly, some competencies with multiple milestone criteria (patient care, for example) may ultimately require the use of several assessment tools for complete evaluation. To address this issue, a component of the ACGME Outcomes Project was the creation of the ACGME/American Board of Medical Specialties Toolbox of Assessment Methods,<sup>11</sup> which provided examples of assessment tools applicable for each competency. Successfully translating this long list of potential tools into a short, well-studied list of EM-specific assessments hinges on several issues. Assessment tools must be shown to have sufficient validity evidence to produce valuable and reliable conclusions and be feasibly implemented into EM residency programs without burdening residency program leadership or faculty with the evaluation process or causing negative patient care effects.

Validity is defined as the degree to which the result of an assessment reflects the construct (or reality) it is meant to measure.<sup>12</sup> There is no single test of validity or a point in time when an assessment tool is said to change from invalid to valid. Instead, validity evidence is a collection of data about the design and performance characteristics of an assessment tool that supports the interpretations drawn from data collected using the tool.<sup>12,13</sup> There are five sources of validity evidence that should be sought for assessment tools used for high-stakes, summative assessment: content, response process, internal structure, relationship to other variables, and consequences<sup>14,15</sup> (Table 3). The different types of

validity evidence complement each other and demonstrate a rigorous approach to assessment tool development, implementation, and ongoing use. Greater degrees of validity evidence are recommended for summative assessments than for formative assessments, since judgments from summative evaluations used for resident promotion can significantly alter the future of trainees. We recommend that specific attention be paid to collecting these different types of validity evidence for new assessment tools developed for the NAS.

An element of validity evidence that warrants particular attention is reliability. Collection of evidence of reliability of an assessment falls under the heading of internal structure validity evidence. Reliability is the degree to which the result of a single assessment reflects all other assessments of the same learner at a given point in time.<sup>16</sup> Two examples of reliable tests include 1) a situation where multiple raters give similar scores after direct observation of a single encounter or 2) a single rater gives similar scores to a learner after observing multiple encounters where the performance was similar. Reliability of assessment data is of critical importance because we must rely on a small sample of assessments and assume that behaviors witnessed in these settings are generalizable to the resident's clinical practice as a whole. We suggest that all newly developed assessment tools be specifically studied to determine data reliability characteristics, with particular focus paid to the number of raters or samples required to yield reproducible scores. Training of evaluators will also be integral to the collection of reliable data using new assessment tools.

Issues of feasibility are important for the successful integration of newly created assessments. Information about the amount of time and resources required to perform assessments of each core competency domain and the associated milestones should be collected. Other feasibility issues requiring consideration include the physical resources required to complete an assessment and the availability of these resources at programs across the country, the amount of faculty time required for training and completion of assessments, and the acceptance of the assessments by trainees. A balance must be struck between feasibility concerns and obtaining the appropriate degrees of reliability and validity to draw meaningful conclusions from assessments.

## **RECOMMENDATIONS FOR FUTURE RESEARCH BY CORE COMPETENCY DOMAIN**

Summarized below are the recommendations for future research on assessment tools for each core competency domain proposed by the writing group.

### **Patient Care**

1. Determine the number of direct observation assessments and types of patient encounters (e.g., critical diagnoses, chief complaints, diagnostic complexity) that are needed to provide a valid reflection of patient care competence for an individual resident.
2. Design and codify a process to create reliable and valid simulation, objective structured clinical exam (OSCE), and oral examination assessments that use

Table 3  
Sources of Validity Evidence<sup>8</sup> Applied to EM Assessment Tools

**Content**

Evidence of direct correlation between learning objectives and assessment items.

*Example:* For a multiple-choice question exam of EM knowledge, an exam blueprint should be created to demonstrate that the distribution of questions on the exam represents the spectrum of illnesses and acuity of presentations in the Model of the Clinical Practice of Emergency Medicine.<sup>9</sup>

**Response process**

Evidence of data integrity, demonstrating that possible sources of error in an assessment are identified and controlled.

*Example:* Documentation of training and ongoing quality control for faculty raters utilizing a global rating scale for direct observation of residents' ICS.

**Internal structure**

Data on the statistical characteristics and reliability of an assessment tool once put into practice.

*Example:* Determining how well items on a performance checklist to assess a procedural skill correlate with high scoring and low scoring (item discrimination).

**Relationship to other variables**

Measurement of the correlation of assessment scores to an existing or older standard.

*Example:* Comparing the results of a newer assessment of medical knowledge, like a SCT, with a well-known test of medical knowledge, the in-training exam.

**Consequences**

Collecting information on how an assessment affects trainees, faculty, patients, and society.

*Example:* Collecting and publishing information used to determine the cut scores for passing or failing a high-stakes exam, like the ABEM qualifying examination.

Adapted from Downing SM. Validity: on the meaningful interpretation of assessment data. *Med Educ.* 2003;37:830-7.<sup>14</sup>

ABEM = American Board of Emergency Medicine; ICS = interpersonal and communication skills; SCT = script concordance test.

checklists (time to event or critical action) and global ratings to assess competence in ways that reflect expert clinical practice (which may use shortcuts) rather than simply the accomplishment of basic task lists.

3. Determine the number of global assessments needed to compose a valid assessment of a resident's patient care competence accounting for the known biases of this method.
4. Assess the validity and relevance of nonclinician evaluations in patient care competence given the influence of potential confounders and greater relevance to other core competencies such as ICS and professionalism.
5. Determine the validity of clinical metrics relative to other more-studied forms of assessment with good reliability and validity, such as direct observation, OSCE, and simulation.
6. Develop training programs and assessments for procedural skill acquisition starting with no-risk methods such as simulated, cadaveric, or OSCE experiences and concluding with direct observation assessment during actual patient care and correlation to complications and patient outcomes.

**Medical Knowledge**

1. Compare performance on American Board of Emergency Medicine (ABEM) certification examinations and maintenance of certification participation with patient outcomes.
2. Compare script concordance Test (SCT) performance with current criterion-standard tests such as the ABEM qualifying and oral examinations. This would require development of a national database of scripts relevant to EM. At least 20 EM experts should be recruited to develop a scoring metric for each script, to be used with the SCT.

The consensus group also felt that the SCT should be studied compared to other standards such as patient-centered outcomes.

3. Multicenter data collection to compare performance on the Council of EM Residency Directors Question Bank to the in-training exam or to the ABEM qualifying examination.
4. Further investigate the role of mini-clinical evaluation examination (CEX) in EM and the value of the standard direct observation tool (SDOT) for medical knowledge assessment.

**PBLI**

1. Characterize and disseminate the methods currently utilized by EM residency programs to assess the PBLI competency.
2. Determine the applicability of and generate validity evidence for the use of existing tools from other specialties for the assessment of PBLI in EM residents.
3. Develop and generate validity evidence for other reliable processes for assessing this competency in EM.
4. Develop methods for capturing the effort on existing activities such as patient follow-up logs, morbidity and mortality conferences, and quality improvement projects into a formal assessment process.
5. Develop automated methods that would integrate with medical records to gather patient outcome information and provide real-time reporting to individual providers.
6. Investigate the benefits, limitations, validity, reliability, liability, and patient confidentiality issues specific to the use of portfolios for assessment of PBLI within EM.
7. Develop reliable ways to assess the evidence-based medicine subcomponents of PBLI and the correlation between improved evidence-based medicine skills and patient outcomes.

## ICS

1. Identify specific, observable, and desirable ICS best practices and support primary research on physician communication skills specific to EM.
2. Specifically design and evaluate a formative assessment tool to provide feedback to residents on their acquisition of ICS and application of skills in clinical practice.
3. Evaluate the appropriateness of tools for summative assessment of specific ICS, like team resource management and delivering bad news, which are included in the EM milestones.
4. Encourage collaborative development and collection of multi-institutional validity evidence for ICS assessment tools.

## Professionalism

1. Because multiple experts advocate for a multimodal approach to professionalism assessment, develop and evaluate triangulation strategies that combine data accuracy with efficiency and determine if tools have sufficient psychometric rigor to allow for use in summative assessment.
2. Evaluate whether non-EM tools (e.g., DIT2, P-MEX) can produce data that are a valid and reliable reflection of professionalism in EM residents.
3. Evaluate qualitative tools, such as portfolios, for EM professionalism assessment by conducting a needs assessment of current practices among EM residencies, determining the essential elements of a portfolio, and developing portfolio assessment rubrics.
4. Evaluate existing multisource feedback instruments by determining the number of evaluations needed to achieve adequate reliability, especially for patient evaluations, and evaluating strategies, such as rater training, to minimize biases that may compromise the validity of multisource feedback assessments.
5. Investigate implementation strategies that improve the feasibility of direct observation for the assessment of professionalism. This may include generation of validity evidence for use of the SDOT specifically for professionalism or creation of a new direct observation tool incorporating the NAS professionalism milestones.
6. Explore multiple formats for reporting critical professionalism lapses and determine the role of these reports in formative and summative assessment.

## SBP

1. Collect validity evidence for the various teamwork and situational awareness rating scales using simulation in EM education and investigate the association between performance in simulation and patient care.
2. Refine the current direct observation assessment tools (e.g., mini-CEX, SDOT) utilizing scales with progressively developing observable behaviors or milestones. Determine evidence of validity for these assessments, including the optimal number of observations, association with other outcomes, and the consequences of performance of the assessment.
3. Develop forms to be used by multiple evaluators (multisource feedback) to measure EM resident performance in the SBP domains using scales with

- progressively developing observable behaviors or milestones. Determine the effect of multisource feedback on resident performance (consequential validity).
4. Develop structured self-assessment tools that describe residents with different engagement, abilities, and skills in quality improvement and develop tools for assessing quality improvement projects (e.g., checklist for chart review) using objective quality measures for the SBP competency.
5. Develop self-assessment tools that encourage informed self-assessment and reflection on specific SBP domains and behaviors and develop structured scoring rubrics for faculty to use in assessing portfolio reflections.

## CONCLUSIONS

We have presented a summary of recommendations from the writing group on assessment of observable learner performance at the 2012 *Academic Emergency Medicine* consensus conference on education research in EM. The primary aim of the writing group was to review the current state of assessment tool literature for the ACGME core competencies and create an agenda for future research on assessment tools relevant to the transition to the Next Accreditation System. General recommendations to meet this aim include the purposeful collection of five types of validity evidence, as well as reliability and feasibility data for all tools utilized for the assessment of EM residents.

We acknowledge the following writing group members for work that contributed to the creation of the research agenda presented here: Doug Franzen, Nikhil Goyal, Amer Aldeen, Thomas Swoboda, Christine Kulstad, Esther Chen, James Kimo Takayesu, Elliot Rodriguez, David Salzman, Teresa Chan, Jeffery Siegelman, Joshua Wallenstein, Camiron Pfennig, Clare Wallner, Jeremy Branzetti, Fiona Gallahue, David Gordon, Jonathan Ilgen, and Patricia O'Sullivan. We also recognize the conference co-chairs Nicole Deiorio, Lalena Yarris, and Joseph LaMantia; student volunteers Nathan Haas and Robert Furlong; and academic assistant Megan McCullough and thank David Sklar and William McGaghie for feedback on the manuscript.

## References

1. Accreditation Council for Graduate Medical Education. Common Program Requirements. Available at: [http://www.acgme-2010standards.org/pdf/Common\\_Program\\_Requirements\\_07012011.pdf](http://www.acgme-2010standards.org/pdf/Common_Program_Requirements_07012011.pdf). Accessed Dec 9, 2012.
2. Holt KD, Miller RS, Nasca TJ. Residency program's evaluations of the competencies: data provided to the ACGME about types of assessments used by programs. *J Grad Med Educ*. 2010; 2:649–55.
3. Nasca TJ, Philibert I, Brigham T, Flynn TC. The next GME accreditation system—rationale and benefits. *N Engl J Med*. 2012; 366:1051–6.
4. Royal College of Physicians and Surgeons of Canada. CanMEDS 2005 Framework. Available at: <http://www.med.uottawa.ca/sim/data/CanMEDS%202005%20Framework.pdf>. Accessed Sep 20, 2012.
5. Chan TM, Wallner C, Swoboda TK, Leone KA, Kessler CS. Assessing interpersonal communication

- skills in emergency medicine. *Acad Emerg Med.* 2012; 19:1390–1402.
6. Goyal N, Aldeen A, Leone K, Ilgen JS, Branzetti J, Kessler CS. Assessing medical knowledge of emergency medicine residents. *Acad Emerg Med.* 2012; 19:1360–65.
  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. Salzman DH, Franzen DS, Leone KA, Kessler CS. Assessing practice-based learning and improvement. *Acad Emerg Med.* 2012; 19:1403–10.
  9. Rodriguez E, Siegelman J, Leone KA, Kessler CS. Assessing professionalism: summary of working group on assessment of observable learner performance. *Acad Emerg Med.* 2012; 19:1372–78.
  10. Chen EH, O'Sullivan PS, Pfennig CL, Leone KA, Kessler CS. Evaluating systems-based practice. *Acad Emerg Med.* 2012; 19:1366–71.
  11. Accreditation Council for Graduate Medical Education. Toolbox of Assessment Methods. Available at: [http://sklad.cumc.columbia.edu/acgme/toolbox/toolbox46/acgme\\_toolbox.pdf](http://sklad.cumc.columbia.edu/acgme/toolbox/toolbox46/acgme_toolbox.pdf). Accessed Sep 20, 2012.
  12. Downing SM, Haladyna TM. Validity and its threats. In: Dowing SM, Yudkowsky R (eds). *Assessment in Health Professions Education*. New York, NY: Routledge, 2009, pp 21–55.
  13. Downing SM, Haladyna TM. Validity threats: overcoming interference with proposed interpretations of assessment data. *Med Educ.* 2004; 38:327–33.
  14. Downing SM. Validity: on the meaningful interpretation of assessment data. *Med Educ.* 2003; 37:830–7.
  15. 2011 EM Model Review Task Force. The 2011 model of the clinical practice of emergency medicine. *Acad Emerg Med.* 2012; 19:e19–40.
  16. Axelson RD, Kreiter CD. Reliability. In: Dowing SM, Yudkowsky R (eds). *Assessment in Health Professions Education*. New York, NY: Routledge, 2009, pp 57–73.

## APPENDIX A

Participants in the Breakout Session on Assessment of Observable Learner Performance at the 2012 AEM Consensus Conference on Education Research in Emergency Medicine:

Mike Beeson, Steven L. Bernstein, Kevin Biese, William Bond, Jeremy Branzetti, John Burton, Esther Chen, Rob Cloutier, Lauren W. Conlon, David Cook, Suzanne Dooley-Hash, Lillian Emler, Michael T. Fitch, Doug Franzen, Robert Furlong, Gary Gaddis, Fiona Gallahue, Maureen Gang, David Gordon, Jim Gordon, Nikhil Goyal, Richard Gray, Marna Greenberg, Nathan Haas, Danielle Hart, Nick Hartman, Cullen Hegarty, Corey Heitz, Sheryl Heron, Talmage Holmes, Hans House, Butch Humbert, Roger Humphries, Jonathan Ilgen, Lisa Jacobson, Julianna Jung, Sharhabeel Jwayyed, Colleen Kalynych, Chad Kessler, Gloria Kuhn, Christine Kulstad, James Kwan, Daniel Lakoff, Richard Lammers, Katrina Leone, Judy Linden, Elise Lovell, John Marshall, Kerry McCabe, Megan McCullough, Chris McDowell, William C. McGaghie, Brian Nelson, Jason Nomura, Susan Promes, Elliot Rodriguez, Nestor Rodriguez, David Salzman, Kaushal Shah, Peter Shearer, Jeff Siegelman, Alison Southern, Alison Suarez, Tom Swoboda, Lindsey Tilt, Vicken Totten, Seth Trueger, Danielle Turner-Lawrence, Phyllis Vallee, Salvator J. Vicario, Joshua Wallenstein, Clare Wallner, Susan Watts, Brian Weitzman, John Wightman, and Mildred Willy.