Assessing Resident Performance: Do We Know What We Are Evaluating?

Peter C. Wyer, MD*

*Corresponding Author. E-mail: pwyer1@gmail.com.

0196-0644/$-see front matter

https://doi.org/10.1016/j.annemergmed.2019.09.019

A **podcast** for this article is available at www.annemergmed.com.

SEE RELATED ARTICLES, P. 647, 660, AND 670.

[Ann Emerg Med. 2019;74:679-681.]

Residency training in emergency medicine has been under way for half a century. During the first 2 decades of that history, evaluation of residents by clinical supervisors and program faculty reflected a holistic approach based on the experience of working clinically with them during emergency department (ED) and other rotations. At the time, program faculty would likely have expressed confidence that they were able to accurately identify both outstanding and deficient residents, despite that they might not be able to explain precisely how they arrived at their determinations. However, on the level of accreditation of graduate medical education, there was concern that predetermined educational outcomes needed to be better specified, as well as the alignment of these outcomes with desired attributes of the 21st-century physician.[1] What was being evaluated and were the right things included? Competency-based medical education was born of these concerns.

In the late 1990s, general competency descriptions were advanced by the Royal College of Physicians and Surgeons of Canada and the Accreditation Council for Graduate Medical Education (ACGME) (Outcomes Project).[1] These descriptions included characteristics such as interpersonal and communication skills and professionalism that were intended to embody humanistic values and a commitment to patient-centered care. However, the use of those descriptions directly as the basis for resident evaluation within various specialties proved problematic. Ten years after the ACGME Outcomes Project competency descriptions were announced, Lurie et al[2] reviewed published reports of attempts to validate instruments derived from the individual competencies and concluded that efforts to use the descriptions as a primary assessment platform should be abandoned. Extensive overlap between the competency descriptions and poor consistency of definitions, with resulting ambiguity, contributed to the dilemma.

In the United States, the failure of the ACGME general competency descriptions to translate successfully to assessment tools resulted in a return to specialty-specific frameworks. The New Accreditation System, announced by the ACGME in 2012,[3] introduced a template for defining developmental milestones that promises to serve as a model for the accreditors of graduate medical education in Canada as well. The American Board of Emergency Medicine, and the other specialty boards, enjoyed substantial license to develop a set of timelined longitudinal milestone descriptions satisfying both the New Accreditation System criteria and the perceived needs of emergency medicine. A full set of milestones for emergency medicine residency training soon followed.[4]

The Milestone framework reflects a movement away from the general competency framework. The majority of the ACGME Milestone data points correspond to technical skills specific to the practice of emergency medicine. Milestone categories corresponding to the original ACGME general competency descriptions are clustered at the end of the grid. An analysis of a large US data set of milestone ratings of emergency medicine residents in 2015 revealed the presence of 5 independent factors within the content of the instrument, 4 of which were composed of technical skills. The remaining factor included the original ACGME competency categories other than medical knowledge, combined with technical items such as computer skills. In the analysis, it accounted for only 4% to 7% of the total variance, suggesting that the former received little attention in the course of the evaluation process[5]; hence, we seem to be facing the very dilemma we started with 2 decades ago. Overriding questions remain: what is being evaluated by the new frameworks, and how confident should we be that they are helping us produce better physicians?

Three reports in this issue of *Annals*, from 2 national health systems outside of the United States, illuminate some of the challenges, strengths, and opportunities posed by the current status of assessment of graduate medical education.[6-8] Of these, Castonguay et al[6] elicited evaluations of emergency residents during an ED rotation from supervising attending physicians, nurses, and patients within a single French Canadian residency program. Although all of the survey forms included questions reflecting the collaborator, professional, and communicator roles within the CanMEDS framework of the Royal College of Physicians and Surgeons of Canada, only the patient and nurse participants commented on these qualities. Supervising physicians focused on resident medical expertise. These observations are consistent with the findings of the previously cited factor analysis of the emergency medicine milestones.[5]

The remaining 2 studies, by Townend et al[8] and Szulewski et al,[7] entail the use of evaluation frameworks that suggest interesting alternatives to those we have been struggling with in North America. Crisis resource management and nontechnical skills constitute closely related clusters of skills that pertain to learner performance of roles within specific organized settings and contexts such as team-based resuscitation or management and decisionmaking within an ED acute care unit. The skills included within these constructs pertain to team building, communication, and leadership and embody elements of many, if not most, of the US and Canadian general competency categories.[8]

Of the 2 reports, that of Townend et al[8] is particularly impressive. The investigators previously derived and preliminarily validated a tool for nontechnical skills evaluation. The Royal College of Emergency Medicine of the United Kingdom used this as the basis of the Extended Supervised Learning Event, a standardized system of evaluative observation of emergency medicine residents that is used throughout the country. To assess the reliability of the Extended Supervised Learning Event, Townend et al[8] used a statistical model that allows the simultaneous analysis of multiple determinants of resident scores and also the determination of the resource requirements for an acceptable level of reproducibility. The authors relied on the data generated in the course of routine periodic evaluations by faculty members in the different programs who were not part of the research effort. They determined that 2 assessors, evaluating residents 3 times a year, could achieve such a level.

Keeping in mind that the evaluation frameworks used by Szulewski et al[7] and Townend et al[8] are limited, there are nonetheless several potential lessons that more comprehensive frameworks such as those of the ACGME and its Canadian counterpart might consider. First, the resident behaviors and skills targeted by the general competencies of the ACGME and the Royal College of Physicians and Surgeons of Canada systems are built in to these more limited frameworks in such a way that they cannot be given short shrift by physician evaluators. Second, the limited skill assessment frameworks are characteristically coupled to specific and well-defined assessment contexts that are themselves standardized. The crisis resource management framework is used in simulation settings involving critically ill patients,[9] whereas the Extended Supervised Learning Event protocol involves structured observation of senior residents practicing in acute care ED units. Defining the context within which trainee evaluation is to take place should serve to limit the extent to which specific criteria are subject to variation in rater interpretation. This was borne out by the findings of Townend et al.[8]

The third potential lesson to be learned from the Extended Supervised Learning Event experience has to do with practicality of implementation from the perspective of the working emergency clinical educator and program leader.

The Extended Supervised Learning Event system is composed of 12 unique items, uses a standardized rating scale that applies to all items, and requires only a small number of raters per resident to be reliable. In contrast, the 24 emergency medicine milestones[10] use a rating scale that is unique to each item and mandates a set of greater than 8,200 data points per year for a 3-year training program with 12 residents per class. Even considering that the Extended Supervised Learning Event does not provide assessment of the emergency medicine–specific technical skills, this difference is striking.

In summary, in the United States, the ACGME Milestone system has resulted in increased demands on residency programs and has generated skepticism among medical educators.[11] Elsewhere, alternative standardized approaches to emergency medicine resident evaluation are being demonstrated to be practical, reliable, and faithful to the tenets of competency-based medical education. I suggest that the agencies that oversee the content and conduct of graduate medical education in North America remain open to the need for and the possibility of ongoing innovation in a fashion that is mindful of the needs of learners, educators, and society at large. Eight thousand data points per year is too many, particularly if we are not in control of what we are measuring.

Supervising editor: David L. Schriger, MD, MPH. Specific detailed information about possible conflict of interest for individual editors is available at https://www.annemergmed.com/editors.

Author affiliations: From the Department of Emergency Medicine, Columbia University Medical Center, New York, NY.

Authorship: All authors attest to meeting the four ICMJE.org authorship criteria: (1) Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND (2) Drafting the work or revising it critically for important intellectual content; AND (3) Final approval of the version to be published; AND (4) Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

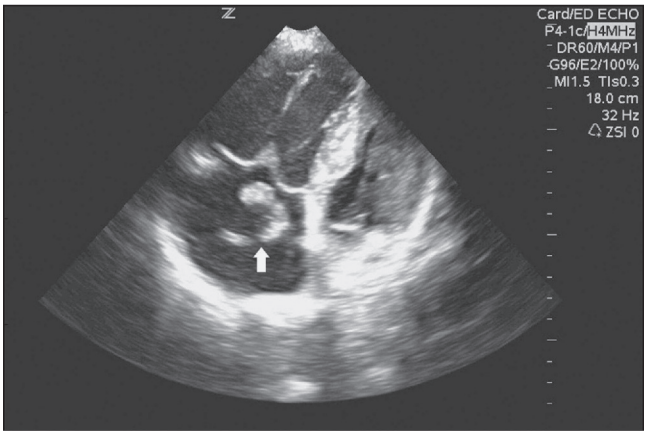
Funding and support: By *Annals* policy, all authors are required to disclose any and all commercial, financial, and other relationships in any way related to the subject of this article as per ICMJE conflict of interest guidelines (see www.icmje.org). The author has stated that no such relationships exist.

REFERENCES

1. Batalden P, Leach D, Swing S, et al. General competencies and accreditation in graduate medical education. *Health Aff (Millwood)*. 2002;21:103-111.
2. 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;84:301-309.
3. Nasca TJ, Philibert I, Brigham T, et al. The next GME accreditation system—rationale and benefits. *N Engl J Med*. 2012;366:1051-1056.
4. Beeson MS, Carter WA, Christopher TA, et al. The development of the emergency medicine milestones. *Acad Emerg Med*. 2013;20:724-729.
5. Beeson MS, Holmboe ES, Korte RC, et al. Initial validity analysis of the emergency medicine milestones. *Acad Emerg Med*. 2015;22:838-844.
6. Castonguay V, Lavoie P, Karazivan P, et al. Perceptions of emergency medicine residents of multisource feedback: different, relevant, and useful information. *Ann Emerg Med*. 2019;74:660-669.
7. Szulewski A, Braund H, Egan R, et al. Starting to think like an expert: an analysis of resident cognitive processes during simulation-based resuscitation examinations. *Acad Emerg Med*. 2019;74:647-659.
8. Townend W, Gopal A, Flowerdew L, et al. The Extended Supervised Learning Event (ESLE): assessing non technical skills in emergency medicine trainees in the workplace. *Ann Emerg Med*. 2019;74:670-679.
9. Flowerdew L, Brown R, Vincent C, et al. Identifying nontechnical skills associated with safety in the emergency department: a scoping review of the literature. *Ann Emerg Med*. 2012;59:386-394.
10. Accreditation Council for Graduate Medical Education; American Board of Emergency Medicine. The emergency medicine milestone project. Available at: https://www.acgme.org/Portals/0/PDFs/Milestones/EmergencyMedicineMilestones.pdf. Accessed June 13, 2019.
11. Nabors C, Peterson SJ, Forman L, et al. Operationalizing the internal medicine milestones—an early status report. *J Grad Med Educ*. 2013;5:130-137.

Images in Emergency Medicine

The *Annals* Web site *www.annemergmed.com* contains a collection of hundreds of emergency medicine-related images, complete with brief discussion and diagnosis, in 18 categories. Go to the Images pull-down menu and test your diagnostic skill today. Below is a selection from the Ultrasound Images.



"Elderly Female With Syncope" by Byrne, Czuczman, and Hwang, July 2011, Volume 58, #1, pp. 105, 115.