

Competence: The Link between Education and Quality Patient Care

Matthew D. McEvoy, M.D., Arna Banerjee, M.B.B.S., F.C.C.M., David J. Murray, M.D.

THE processes in training and assessment that ensure that a physician will be capable and able to perform their job throughout years of medical practice are essential for maintaining the public's trust and confidence in the social contract that exists with medical education.¹ The opening statement in the article by Blum *et al.* in this edition of *ANESTHESIOLOGY* articulates this concept by stating that "evaluating whether graduates of anesthesiology residency programs are competent is an essential goal."² The major aim of the study was to help "improve resident proficiency by further evaluation of a methodology to assess a resident's critical performance behaviors that are not typically captured in a standardized way over the course of residency training." The findings by Blum *et al.* advance the literature on performance measurement in two significant ways. First, the authors describe a valid and reliable set of tools for assessing anesthesiology resident performance in domains of practice that are not typically evaluated in a standardized fashion during clinical training. Specifically, the research team used validated simulation scenarios to assess critical behaviors for managing a wide range of perioperative events that occur infrequently in clinical practice. Second, the authors advance upon previous studies by testing these tools across three residency training programs, thus adding to the generalizability of the findings. Simply put, the study demonstrates that standardized and validated simulation scenarios paired with valid and reliable performance assessment scores can discriminate between senior and junior trainees. This is of profound importance because implementation of such a system of evaluation should allow anesthesiology program



"...[S]tandardized and validated simulation scenarios paired with valid and reliable performance assessment scores can discriminate between senior and junior trainees....[T]he decision about whether a resident's performance meets a threshold of competence must be the next step in educational research."

directors to assess resident skills in an objective manner with confidence and, if necessary, guide educational interventions with the goal of improving the quality of their graduates' performance.

However, the study does not provide a standard of competence to which the scores could be compared. As such, the decision about whether a resident's performance meets a threshold of competence must be the next step in educational research. In fact, defining competence is important not only for trainees but also in continuing professional development for all practicing physicians. Moreover, assessment of competence is far from standard and often may not be valid across many anesthesiology training programs in the United States and globally. Accordingly, a major frontier in graduate medical education and continuing professional development is defining the science of competency-based training and assessment with greater rigor, which will have direct benefit for our patients and society.

A simple method to make a decision about competence is to ask an expert who has in-depth information about the skill and ability of the physician. In an often-quoted study by Slogoff *et al.*,³ anesthesiology program directors were asked to indicate whether they would let their graduates perform anesthesia for three surgeries *on them*: elective cholecystectomy, laparotomy for acute bowel obstruction, and sitting posterior fossa craniotomy. The program directors reported fewer deficits in the domains of knowledge, character, response to stress, clinical performance, and work habits in those graduates who would be allowed to perform all three anesthetics as compared to those who would be allowed to perform fewer cases

Image: STRATUS Center for Medical Simulation, Brigham and Women's Hospital.

Corresponding article on page 821.

Accepted for publication November 29, 2017. From the Department of Anesthesiology, Vanderbilt University School of Medicine, Nashville, Tennessee (M.D.M., A.B.); the Center for Innovation in Perioperative Health, Education, and Research (M.D.M.), and the Center for Experiential Learning and Assessment (A.B.) Vanderbilt University Medical Center, Nashville, Tennessee (M.D.M., A.B.); and the Department of Anesthesiology, Washington University Clinical Simulation Center, Washington University in St. Louis, St. Louis, Missouri (D.J.M.).

Copyright © 2018, the American Society of Anesthesiologists, Inc. Wolters Kluwer Health, Inc. All Rights Reserved. *Anesthesiology* 2018; 128:707-9

or none. A concerning finding in the study was that in the ratings of the 1,310 graduates who took the American Board of Anesthesiology (ABA) Certification Exam in 1992, only 63% of graduates would be permitted to perform anesthesia for all three cases on their program director, but all of the residents being rated had graduated from residency and were being allowed to practice *any* case on the public. Even more surprising, 7% of graduates would not be allowed to perform *any* anesthetic for their program director, and they still graduated! The strong correlation between passing the certification exam and the number of cases that the graduate would be able to perform for the program director was cited as evidence of the validity of the decisions as well as the board certification process. However, few would accept an approach to competence decisions that use a single, potentially biased expert as a reasonable approach to a valid and reliable competence assessment.

When we examine the concept of clinical competence in medicine, confusion still exists as to what constitutes a rating of “satisfactory” performance that is simply more than a gut feeling.⁴ We need to strive to make the judgment of competence more objective with research focusing on reliable and valid tools that will help with this assessment. Researchers and educators are now using the term *entrustable professional activities* as a way to determine what a trainee can be trusted to do alone. There still remains an element of subjectivity in the judgment, but having a large number of entrustable professional activities (*e.g.*, more than three case types!) and many raters besides the program director may increase the validity and reliability of such ratings.

Into this ongoing area of investigation, the study by Blum *et al.* makes a significant contribution by giving program directors and educators a validated set of simulation scenarios as well as an assessment tool that produces valid and moderately reliable scores.² As noted above, the goal of the study was to identify important gaps in anesthesiology resident performance that are not typically evaluated in a standardized fashion. Accordingly, the investigators evaluated 67 residents who each participated as the team leader in seven perioperative scenarios. Each scenario performance was evaluated across five domains similar to those used in the study by Slogoff *et al.*³: synthesizes information to formulate a clear anesthetic plan, implements a plan based on changing conditions (*i.e.*, patient state), demonstrates effective interpersonal and communication skills with patient and staff, identifies ways to improve performance, and recognizes [their] own limits. In addition to using validated scenarios and assessment tools, the study was performed across three residency training programs, which adds some degree of generalizability to the findings. The researchers found that their tools could discriminate performance between senior and junior residents and also between the performance of residents at the same level. Assessment tools that can perform this function are rare and needed.

However, it should be noted that further research is needed on these tools. For instance, it would be of great utility

to know if the evaluations of the simulated performances cohere with the clinical competency committee ratings that residents receive every 3 to 6 months during training, and if so, to what extent? Additionally, do the ratings from the simulation assessments confirm what is already known or add new information to faculty evaluations from clinical rotations and in-training exam scores? Furthermore, trainees are now required to be evaluated with Milestones, which are defined as competency-based outcomes of knowledge, skills, behaviors, attitudes, and clinical/professional performance that are described across five levels of development spanning the range from novice to expert.⁵ The goal of this system is to provide more objective data for how a trainee is developing such that the faculty can give them progressive and graded increases in clinical and professional responsibility as they move toward graduation and then to unsupervised practice. Thus, an important question would be whether the evaluation system studied by Blum *et al.* can incorporate minimum passing scores or score ranges that indicate where the trainee's performance falls within the Milestones system. All of these questions require a research road map and collaboration just as any other area of investigation in perioperative research.

But for now, what should program directors and educators do with the evidence that is presented by Blum *et al.*? The assessment tools, if paired with the simulation scenarios used by the authors, could be used for formative assessment, but not summative assessment. Formative assessments are used to monitor the learning of trainees and give them feedback about their strengths and weakness as they are *progressing through* a training program. Alternatively, summative assessments (*e.g.*, ABA BASIC, ADVANCED, and oral exams) are used to evaluate the learning of trainees at the end of a training paradigm and are used to make high-stakes decisions about graduation and board certification status. This distinction is of profound importance from a practical perspective if these performance ratings are to be used by program directors and clinical competency committees in making decisions of remediation for a resident who does not have acceptable scores. Along these lines, the authors note that “identification of performance concerns early enables remediation and a higher likelihood that interventions will be effective.”² We agree with this statement. However, while the work presented by Blum *et al.* is significant for the education community, from a practical perspective it should be used only as one piece of information among many to make decisions about resident progression through anesthesiology training.

Finally, this study is also of importance from the national perspective because the scenarios and approach used by Blum *et al.* to measure these domains of competence are similar to the domains that the ABA Objective Structured Clinical Examination seeks to assess as part of the certification standard starting in spring 2018. In preparation for this high-stakes exam, the methodology employed by Blum *et al.* could be used for longitudinal assessment of resident performance throughout training in order to produce more competent graduates and also

address the social contract of safer patient care through higher practice standards. Furthermore, if previous research bears out in the future, we may learn that adding another domain of assessment will increase the validity and reliability of high-stakes decisions concerning board certification for residency graduates,⁶ and possibly guide decisions about maintenance of certification in anesthesiology in the years to come.

Anesthesiologists have led advances in patient safety through improvements in systems of care and perioperative therapeutic interventions. The next frontier in improving patient safety is ensuring competence throughout the duration of a specialist's practice. This is especially important given the recent report showing that approximately 25% of practicing anesthesiologists received ratings of poor performance during simulated perioperative scenarios that are similar to those presented in this article.⁷ The report by Weinger *et al.* adds credibility to the need for the maintenance of certification in an anesthesiology program, especially since *more* years in practice was associated with lower performance.⁷ While the current methods of assessment are certainly not perfect, they continue to improve and are currently quite valid and reliable by known metrics. As our toolbox of tests and tools for assessment is far from full for both trainees and those in practice, rigorous educational research like that reported by Blum *et al.* needs to continue so that valid and reliable testing paradigms and rating tools can continue to be developed and shared across training and maintenance of certification programs. Additionally, future versions of these ratings systems need to move beyond evaluating residents by training year (*i.e.*, clinical anesthesia 1, 2, or 3), as the intent of the Milestones system is to evaluate each resident in *their* progression through a training program within a specialty, and more work needs to be done in this regard.^{5,8}

While adding complexity and rigor to any evaluation process should be done with care, physicians and the public should welcome well-validated performance assessments as one more way of ensuring that we are fulfilling the social contract that we have with society as a whole and with each

new patient encounter. The future product of research in this area of inquiry should provide educators with the tools necessary to define and measure competence such that they truly know it when they see it.

Competing Interests

The authors are not supported by, nor maintain any financial interest in, any commercial activity that may be associated with the topic of this article.

Correspondence

Address correspondence to Dr. McEvoy: matthew.d.mcevoy@vanderbilt.edu

References

1. Lee TH: Certifying the good physician: a work in progress. JAMA 2014; 312:2340–2
2. Blum RH, Muret-Wagstaff SL, Boulet JR, Cooper JB, Petrusa ER; for the Harvard Assessment of Anesthesia Resident Performance Research Group: Simulation-based assessment to reliably identify key resident performance attributes. ANESTHESIOLOGY 2018; 128:821–31
3. Slogoff S, Hughes FP, Hug CC Jr, Longnecker DE, Saidman LJ: A demonstration of validity for certification by the American Board of Anesthesiology. Acad Med 1994; 69:740–6
4. Albanese MA, Mejicano G, Mullan P, Kokotailo P, Gruppen L: Defining characteristics of educational competencies. Med Educ 2008; 42:248–55
5. Ebert TJ, Fox CA: Competency-based education in anesthesiology: History and challenges. ANESTHESIOLOGY 2014; 120:24–31
6. Zhou Y, Sun H, Culley DJ, Young A, Harman AE, Warner DO: Effectiveness of written and oral specialty certification examinations to predict actions against the medical licenses of anesthesiologists. ANESTHESIOLOGY 2017; 126:1171–9
7. Weinger MB, Banerjee A, Burden AR, McIvor WR, Boulet J, Cooper JB, Steadman R, Shotwell MS, Slagle JM, DeMaria S Jr, Torsher L, Sinz E, Levine AI, Rask J, Davis F, Park C, Gaba DM: Simulation-based assessment of the management of critical events by board-certified anesthesiologists. ANESTHESIOLOGY 2017; 127:475–89
8. Nasca TJ, Philibert I, Brigham T, Flynn TC: The next GME accreditation system—Rationale and benefits. N Engl J Med 2012; 366:1051–6