The American Journal of Surgery (2008) 196, 609–615

The American Journal of Surgery®

Surgical Education

Competency assessment in simulation-based procedural education

James D. Michelson, M.D.[a,*], Lance Manning, M.B.A.[b]

[a]*The University of Vermont College of Medicine, Stafford Hall 436A, 95 Carrigan Dr., Burlington, VT 05405, USA;* [b]*The George Washington University School of Business, Washington, DC, USA*

KEYWORDS:
Medical education;
Simulation;
Minimally invasive surgery

Abstract. The field of simulation-based procedural education in medicine is undergoing rapid development, with significant improvements in both graphic and tactile fidelity. As a consequence, the use of simulation for competency assessment and credentialing is evolving rapidly. Establishing the conceptual framework for developing such assessments using simulation is becoming critical to the expansion of simulation-based education and assessment in medicine. Although medical literature explores the educational theories in other aspects of medical education, the applicability of those concepts to the simulation of time-critical and clinically dangerous procedures has not been addressed explicitly. In particular, the issue of how one establishes standards for simulation-based procedures is central to how simulation would be used for competency assessment. This article briefly reviews the current medical education theories and discusses their applicability to simulation-based education. An alternative methodology of standards setting involving the use of benchmarking may be more appropriate to assessing critical procedural skills. Although much of the existing simulation literature (and practice) implicitly uses benchmarking methods, the conceptual framework that justifies its use has not been discussed explicitly. Finally, the development of clinical benchmarks as the standards-setting mechanism for procedural simulation-based learning, feedback, and assessment will be critical to establishing the clinical relevance of simulation.

The past several years has witnessed the emergence of the use of high-fidelity simulation to teach procedural-based medicine. This includes computerized mannequins used to teach emergency resuscitation and airway management, as well as a growing number of surgical simulators. Although these can be used for clinical skills training, the focus of this article is on their use in invasive and critical care procedures in which there are performance time constraints in the clinical environment. The educational concept underpinning the use of simulators has been to recognize that successful clinical outcomes require the learning of not only simple task training goals but also clinical decision making, teamwork, and deliberate practice.[1,2] To date, the majority of research into simulation-based procedural medical education has focused on different aspects of its educational validity. In contrast to the cognitive medical education literature, the issue of standards setting in simulation-based procedural education has received relatively little explicit debate.[2]

Notwithstanding the absence of such a discussion within the simulation literature, several investigators have taken a common-sense approach to the issue, and have used benchmarking to establish the competency standard.[3,4] Although what might be the most appropriate group to use as a benchmark reference is not agreed upon, this technique of

* Corresponding author. Tel.: +1-802-656-2250; fax: +1-802-656-4247.

E-mail address: james.michelson@uvm.edu

Manuscript received June 1, 2007; revised manuscript September 17, 2007

0002-9610/$ - see front matter © 2008 Excerpta Medica Inc. All rights reserved.
doi:10.1016/j.amjsurg.2007.09.050

standard setting is distinct from that used previously in medical education. In light of this recognition, it would be useful to review how the educational process in simulation lends itself to the use of benchmarking standards-setting for assessment.

The context of simulation-based procedural education in medicine

Among the many other facets of medicine, an unwarranted intellectual distinction has sometimes been drawn between cognitive knowledge and procedural skills (eg, consider that several medical schools are colleges of Physicians and Surgeons). Although this has been fodder for an endless stream of social commentary, there is, in fact, some justification for this distinction on a purely educational basis. Based on the educational concepts of Fitts and Posner[5] and the assessment pyramid of Miller,[6] the lower levels of attainment ("knows," "knows how") can be learned effectively from the use of books, journals, lectures, and so forth, whereas adequate education in procedural skills requires an additional component of practice of the procedure to be learned ("shows how"). In the context of this discussion, the term *procedure* will not include clinical skills training or assessment because there is already an extensive amount of literature reporting on the use of simulation in such settings.

The attainment of procedural competency can be viewed from 2 perspectives (Fig. 1). The most frequently discussed is that of Fitts and Posner,[5] in which skill competency acquisition progresses from the cognitive level (can explain and demonstrate the task but cannot perform skills consistently) to the integrative level (improved consistency of performance with fuller understanding of the skill, which permits constructive feedback) to the autonomous level (basic performance without active cognition, accompanied by cognitive focus on continual improvement). This has been elaborated on by Ericsson,[1] who identified deliberate practice as critical to the progression to the integrative phase of motor skill acquisition, as well as being central to the transition from competency to expertise. In deliberate practice, specific components of the skill are identified and practiced in an environment that provides immediate feedback to promote improvement.

In parallel with this model of development, the nature of the skills being learned also has a spectrum. In order of increasing complexity the skills progress from single task training to integrated skills performance to performance of the integrated skills in a team-oriented clinical environment.[2,7,8] Within each of these stages, the student may have to go through each of the psychomotor skills acquisition levels before being able to advance to the next stage.[9]

Simulation-based medical training offers the opportunity to gain and assesses these procedural skills through deliberate practice[1,2,8,10] in an environment where no patient harm can occur. It is gaining increasing prominence because it also addresses the constraints on time and resources needed to train medical personnel with actual patients. As has been observed by several investigators,[2,8,10] the provision of immediate feedback in simulation is a critical component to its educational effectiveness. The feedback, which is essentially formative assessment, is based on the comparison of the student performance to an established stan-

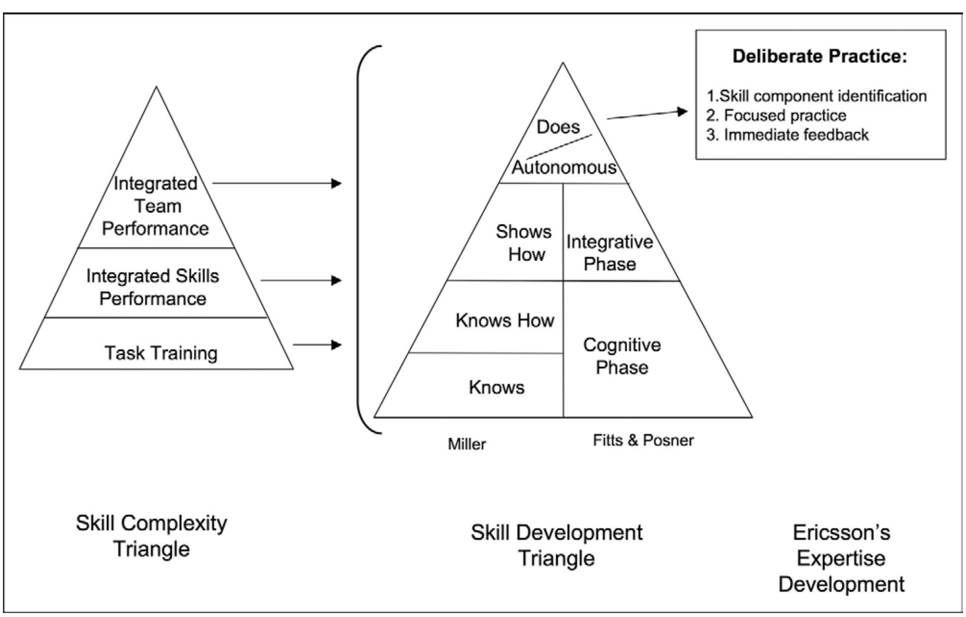

Figure 1 Diagram of the relationship between the level of skills that are to be learned (left triangle) and the process of skills acquisition that occurs within each skill level (right triangle). The Skills Development Triangle represents the original conceptualization of Miller[6] (with his terms listed on the left) and the modifications discussed by Fitts and Posner[5] (with their terms listed on the right). Once the top level of skills development is entered, further improvement is achieved by the use of Deliberate Practice, as described by Ericsson.[1]

dard. The standard used for the summative assessment is either the same or similar. With the multitude of validated measures that are possible under a simulation environment,[11] the question remains as to which are clinically relevant and how to use them in setting the performance standard.[2]

The potential roles for procedural simulation in medicine

Simulation-based education in critical, time-constrained procedures is, or will be, increasingly integrated into the educational curriculum (undergraduate, postgraduate, and continuing education). It is likely that simulation assessments (such as the Toronto Objective Structured Assessment of Technical Skills, McGill Inanimate System for Training and Evaluation of Laparoscopic Skills, and the Imperial College Surgical Assessment Device) also will be a component of high-stakes examinations at the local or national level.[2] Recently, it was proposed by the Food and Drug Administration and participating physician specialty societies that credentialing for carotid stent placement should include satisfactory completion of a simulation-based course for stent placement.[12] The Institute of Medicine also has advocated that the introduction of new medical technologies should be accompanied by simulation-based education and certification requirements.[13]

Standards-setting and the determination of competency in simulation-based procedural education

Given the potential patient harm that can occur if potentially dangerous procedures are performed by inexperienced personnel in the clinical setting, it is important to establish a methodology to effectively teach, and learn, these complex procedures before clinical exposure. There is an extensive amount of literature showing the various components of validity for laparoscopic, endoscopic, and mannequin-based simulators.[4,14–20] There also is emerging evidence that the skills learned in the simulation environment carry over to improved clinical performance (sometimes termed *transfer validity*).[4,21–31] Ultimately, the goal is to establish such a strong linkage between simulation performance and clinical performance that competency in the latter can be predicted by fulfilling a competency standard in the former.

Because such competency-based education using simulation is the objective, the question becomes one of how competency is declared in the simulation environment. Previous investigators have divided the performance indicators into 2 groups: the objective (modular) measures that can be collected as a by-product of the simulation (eg, task duration, error rates, economy of instrument manipulation, and so forth), and the subjective (global) assessments made by the preceptor.[32] There is generally good correlation between these 2 types of measures,[21,33–35] but the specific modular parameters that are predictive of overall performance may be only a limited subset of what can be measured at a granular level.[3]

In both the clinical and simulation environments, the intersection between measuring performance and determining competency lays the issue of standard setting. Given that the goal of medical education, and the assessment of learning, is to produce competent physicians, how does one go about setting the standard against which the performance of students, residents, and established physicians is compared during their licensing and credentialing?

As reviewed elsewhere,[36] the classic methods of standards-setting involve the determination of what is minimally acceptable by a panel of one or more experts. One version, called the *Angoff method*, is the most well known of the absolute type of standard. In this process, a panel of experts estimates the performance of borderline students for each item in a test. This establishes the minimal passing score. Another type of standards-setting explicitly incorporates more subjectivity. This is typified by the Hofstee method, in which expert judges define the acceptable upper and lower limits for both passing scores and failing rates, then apply these to the known testing data to determine the passing score. This involves the absolute application of the relative judgments of what is an acceptable minimal passing score to determine the standard. It also is recognized that the very act of setting the standards for competency by using these techniques involves subjective judgments[36–40] and noneducational influences.[37] Given the inherent subjectivity of these methods, it is not surprising that these various standard-setting methodologies can yield widely divergent passing standards when used on the same test data.[39–41] Lastly, an underlying policy consideration in the establishment of these standards is that setting them too high would lead to an unacceptable rate of failure—thereby limiting the availability of trained physicians, or overwhelming the ability of the educational system to remediate the deficiencies.[37,39]

One approach to addressing this difficulty in determining competency is to anchor it to clinical patient outcomes.[42–46] However, there are very little data to determine the linkage between medical education and clinical outcomes for the majority of clinical situations. In large part this is owing to the complex and poorly understood relationship between the clinical skills taught and how they subsequently are used clinically. This has sparked an interest in discovering those linkages through techniques that have evolved for evidenced-based medicine.[8,42–45,47]

Historically, the procedural training standards have always been implicitly based on the performance of the precepting surgeon. In recent years, this use of benchmarking to establish procedural standards has been more formalized by the advent of simulation. The simulation curricula standards are based on the performance of experienced expert surgeons[3,4,8,23] using the simulation systems, thereby linking the assessment standards to clinical benchmarks of acceptable performance. The definition of what is expected is determined by the design of the simulation curriculum, specifically, the standards used for feedback during the training.

There are several consequences of this method of standards-setting. First, there is no allowance provided for nonclinical performance criteria in the standard (eg, how many people are desired to be trained, or how easy or difficult it is to reach the standard). Allowing insufficiently skillful physicians to perform such procedures is not an option. Second, because the simulation curriculum is based completely on such clinically relevant standards, the entire educational process is focused explicitly on obtaining optimal patient outcomes. Last, the distinction between formative and summative assessment in simulation becomes an artificial boundary. The key educational process (deliberate practice[1]) consists of simulating the clinical situation, with a graduated degree of difficulty until the desired performance is achieved, in conjunction with immediate feedback from clinical experts on the performance.[8,31,48,49] This is typified in the study by Mayo et al,[50] in which "The interns were debriefed extensively and given hands-on training by the attending using the simulator until they achieved perfect performance." By the time the formative stage of assessment is completed, the student has essentially completed the tasks required to achieve competency. The only difference is that for the summative evaluation the student runs through the simulation without immediate feedback.

In simulation-based medical education, then, the standard-setting method is based explicitly on clinical benchmarks, rather than a convocation of experts to determine what constitutes a passing score. Although this helps to establish a closer relationship between medical education and patient outcomes, it only serves to move the question of how to set the standard (in this case, the clinical benchmark) to a different place—the clinical environment. However, this is occurring at a time when there is a great deal of interest in, and study of, clinical care processes and how they can be measured and optimized. Consequently, there is a rapidly expanding body of knowledge of clinical care measurements that is available for use in establishing benchmarks at both the local and national levels. Current work in this field has been exploring various methods of setting benchmarks, from modeling it after the best practitioners in a medical center[4,51] to sophisticated calculations of benchmarks of the highest practical and achievable levels of care.[52–55]

In the context of resident education, the benchmark standards can be adjusted easily for each postgraduate year level by requiring an increasing subset of the full competency benchmarks for higher residency levels. Currently, there will be some subjectivity involved in determining what components of each procedural skill-set are appropriate at each level of residency training. It is likely that some consensus on these training benchmarks will be developed in the near future within the surgical and procedural specialties. The basis of such milestone benchmarks may include contemporary performance abilities of the residents (which could be program-based or derived from specialty society resident data). However, because the graduated acquisition of skills through the residency years is designed to ultimately result in complete competency based on faculty-level performance benchmarks, each year-level benchmark milestone also must be established from this broader perspective. Viewing the achievement of competency as a continuum in simulation performance, it can be appreciated that the process of attaining the required mastery to move to the next level of difficulty or complexity in a procedural simulation will vary among residents, but their quantifiable progress can be used to tailor their educational experience appropriately.

Current uses of benchmarking in simulation-based assessment

The perspective has been expressed that medical education should be seen to exist in the context of the overall quality improvement framework in medicine.[44] Consequently, there should be a focus on defining the parameters to be measured, measuring them, and benchmarking them to assess educational outcomes. This is the underlying concept behind the Best Evidence Medical Education Collaboration,[44] which is attempting to bring the practices of evidence-based medicine to the study of medical education. In a review of the literature on the use of simulation on medical education sponsored by the Best Evidence Medical Education Collaboration, it was noted that explicit reference to educational outcome benchmarks was a critical element in the effective use of simulation.[8]

Aside from the inherent linkage to clinical practice that is provided by benchmarking, there are other advantages of this assessment methodology with respect to standards-setting procedures such as the Angoff or Hofstee methods. In contrast with the consistency of assessment when benchmarking is used,[24,56,57] the difficulty in determining the performance level of the borderline student (required in the latter type of procedures) leads to poor consistency in standards-setting.[58] Furthermore, the passing grades established by such methods vary significantly depending on which specific procedure is used.[10,59] The lack of any explicit or definable linkage to clinical practice has led to comments that such standards-setting can appear to be arbitrary and drawn from "thin air."[10]

An implicit acceptance of this philosophy of performance assessment, as opposed to the test-based methodologies typified by the Angoff and Hofstee techniques, is the way in which simulation has been used in Quality Assessment/Quality Improvement (QA/QI) activities. Primarily using computerized mannequins and endoscopic simulators, simulation has been used for ongoing educational assessment of residents to guide future training.[50,60–62] In this setting, failure to reach the clinical benchmark for performance is used to determine the subsequent educational program. At the team level, it has been similarly used to identify educational targets for improvement[63–66] or in formal QA activities such as Morbidity and Mortality conferences.[67] Following the approach of the Institute of Medicine,[13] several investigators have used simulators to develop and assess new clinical techniques and equipment (including a systems approach to avoid design errors that predispose to mistakes).[68–72] All of these studies were based on a pre-established clinical benchmark for performance or outcome. The actual work environment itself also can be subject to evaluation using simulation in an effort to minimize potentially distracting influences on clinical performance[73] or to evaluate the preparedness to deal with equipment failure.[74]

Of course, the ultimate clinical QA/QI activity is that of clinical certification or credentialing. Although not yet thoroughly validated in this role,[75] simulation-based assessment is beginning to be incorporated into some Board examinations[76] because of its superior capacity to expose what the examinee can "show how" to perform in addition to showing that she/he "knows how" what to do in a given clinical situation.[10,56]

The future of simulation and competency assessment

As the intrinsic validity of procedural simulation has become established, attention is beginning to be focused on using procedural simulation as a tool in competency assessment (eg, credentialing) and clinical process improvement (eg, retrospective and prospective QA/QI activities). The very nature of procedural simulation, however, leads to a different model of standards-setting than is currently practiced in other educational assessments in medicine. The explicit relationship of simulation-based procedural education to established clinical practice benchmarks leads to an outcomes-based competency benchmarked assessment strategy for simulation. In such a setting, using non–outcomes-based standards-setting methodologies (eg, Angoff, Hofstee, and so forth) would only serve to weaken the link between competency assessment and clinical performance.

Although the use of clinical benchmarking as the standards-setting mechanism for procedural simulation-based learning, feedback, and assessment is critical to establishing the clinical relevance of simulation, it should be appreciated that this only heightens the need for more extensive and better-delineated evidenced-based clinical benchmarks.[2] Although this is not trivial, there is emerging evidence that procedural simulation can be used effectively in the development and testing of such benchmarks.[68–72]

References

1. Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. Acad Med 2004;79(Suppl):S70–81.
2. Reznick RK, MacRae H. Teaching surgical skills—changes in the wind. N Engl J Med 2006;355:2664–9.
3. Cotin S, Stylopoulos N, Ottensmeyer M, et al. Metrics for laparoscopic skills trainers: the weakest link! In: Dohi T, Kikinis R, eds. Conference on Medical Image Computing and Computer Assisted Intervention, 2002. Berlin: Springer-Verlag; 2002:35–43.
4. Seymour NE, Gallagher AG, Roman SA, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. Ann Surg 2002;236:458–63.
5. Fitts PM, Posner MI. Human Performance. Reprint of the ed. published by Brooks/Cole Pub. Co. Westport, CT: Greenwood Press; 1979.
6. Miller GE. The assessment of clinical skills/competence/performance. Acad Med 1990;65(Suppl):S63–7.
7. Epstein RM, Hundert EM. Defining and assessing professional competence. JAMA 2002;287:226–35.
8. Issenberg SB, McGaghie WC, Petrusa ER, et al. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. Med Teach 2005;27:10–28.
9. Aggarwal R, Grantcharov T, Moorthy K, et al. A competency-based virtual reality training curriculum for the acquisition of laparoscopic psychomotor skill. Am J Surg 2006;191:128–33.
10. Wayne DB, Fudala MJ, Butter J, et al. Comparison of two standard-setting methods for advanced cardiac life support training. Acad Med 2005;80(Suppl):S63–6.
11. Kneebone R. Simulation in surgical training: educational issues and practical implications. Med Educ 2003;37:267–77.
12. Gallagher AG, Cates CU. Approval of virtual reality training for carotid stenting: what this means for procedural-based medicine. JAMA 2004;292:3024–6.
13. Kohn LTCJ, ed. To Err is Human: Building a Safer Health System. Washington: National Academy Press; 1999.
14. Gaba DM, DeAnda A. The response of anesthesia trainees to simulated critical incidents. Anesth Analg 1989;68:444–51.
15. Ritter EM, McClusky DA III, Lederman AB, et al. Objective psychomotor skills assessment of experienced and novice flexible endoscopists with a virtual reality simulator. J Gastrointest Surg 2003;7: 871–7.
16. Fried MP, Satava R, Weghorst S, et al. Identifying and reducing errors with surgical simulation. Qual Saf Health Care 2004;13(Suppl 1): i19–26.
17. Haluck RS, Gallagher AG, Satava RM, et al. Reliability and validity of Endotower, a virtual reality trainer for angled endoscope navigation. Stud Health Technol Inform 2002;85:179–84.
18. Holcomb JB, Dumire RD, Crommett JW, et al. Evaluation of trauma team performance using an advanced human patient simulator for resuscitation training. J Trauma 2002;52:1078–85.
19. Gaba DM. The future vision of simulation in health care. Qual Saf Health Care 2004;13(Suppl 1):i2–10.
20. Lee SK, Pardo M, Gaba D, et al. Trauma assessment training with a patient simulator: a prospective, randomized study. J Trauma 2003; 55:651–7.
21. Edmond CV Jr. Impact of the endoscopic sinus surgical simulator on operating room performance. Laryngoscope 2002;112:1148–58.
22. Di Giulio E, Fregonese D, Casetti T, et al. Training with a computer-based simulator achieves basic manual skills required for upper endoscopy: a randomized controlled trial. Gastrointest Endosc 2004;60: 196–200.
23. Sedlack RE, Kolars JC. Computer simulator training enhances the competency of gastroenterology fellows at colonoscopy: results of a pilot study. Am J Gastroenterol 2004;99:33–7.

24. Fried GM, Feldman LS, Vassiliou MC, et al. Proving the value of simulation in laparoscopic surgery. Ann Surg 2004;240:518–25.
25. Grantcharov TP, Kristiansen VB, Bendix J, et al. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. Br J Surg 2004;91:146–50.
26. Ost D, DeRosiers A, Britt E, et al. Assessment of a bronchoscopy simulator. Am J Respir Crit Care Med 2001;164:2248–65.
27. Hyltander A, Liljegren E, Rhodin PH, et al. The transfer of basic skills learned in a laparoscopic simulator to the operating room. Surg Endosc 2002;16:1324–8.
28. Hamilton EC, Scott DJ, Kapoor A, et al. Improving operative performance using a laparoscopic hernia simulator. Am J Surg 2001;182: 725–8.
29. Hamilton EC, Scott DJ, Fleming JB, et al. Comparison of video trainer and virtual reality training systems on acquisition of laparoscopic skills. Surg Endosc 2002;16:406–11.
30. Scott DJ, Bergen PC, Rege RV, et al. Laparoscopic training on bench models: better and more cost effective than operating room experience? J Am Coll Surg 2000;191:272–83.
31. Grober ED, Hamstra SJ, Wanzel KR, et al. The educational impact of bench model fidelity on the acquisition of technical skill: the use of clinically relevant outcome measures. Ann Surg 2004;240:374–81.
32. Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg 1997;84:273–8.
33. Adrales GL, Chu UB, Witzke DB, et al. Evaluating minimally invasive surgery training using low-cost mechanical simulations. Surg Endosc 2003;17:580–5.
34. Adrales GL, Park AE, Chu UB, et al. A valid method of laparoscopic simulation training and competence assessment. J Surg Res 2003;114: 156–62.
35. Wilhelm DM, Ogan K, Roehrborn CG, et al. Assessment of basic endoscopic performance using a virtual reality simulator. J Am Coll Surg 2002;195:675–81.
36. Norcini JJ. Setting standards on educational tests. Med Educ 2003;37: 464–9.
37. Kane MT, Crooks TJ, Cohen AS. Designing and evaluating standard-setting procedures for licensure and certification tests. Adv Health Sci Educ Theory Pract 1999;4:195–207.
38. Searle J. Defining competency—the role of standard setting. Med Educ 2000;34:363–6.
39. Cusimano MD, Rothman AI. The effect of incorporating normative data into a criterion-referenced standard setting in medical education. Acad Med 2003;78(Suppl):S88–90.
40. Downing SM, Lieska NG, Raible MD. Establishing passing standards for classroom achievement tests in medical education: a comparative study of four methods. Acad Med 2003;78(Suppl):S85–7.
41. Kaufman DM, Mann KV, Muijtjens AM, et al. A comparison of standard-setting procedures for an OSCE in undergraduate medical education. Acad Med 2000;75:267–71.
42. Shea JA, Arnold L, Mann KV. A RIME perspective on the quality and relevance of current and future medical education research. Acad Med 2004;79:931–8.
43. Colliver JA. Educational theory and medical education practice: a cautionary note for medical school faculty. Acad Med 2002;77: 1217–20.
44. Dauphinee WD, Wood-Dauphinee S. The need for evidence in medical education: the development of best evidence medical education as an opportunity to inform, guide, and sustain medical education research. Acad Med 2004;79:925–30.
45. Glick TH. Evidence-guided education: patients' outcome data should influence our teaching priorities. Acad Med 2005;80:147–51.
46. Carraccio C, Englander R, Wolfsthal S, et al. Educating the pediatrician of the 21st century: defining and implementing a competency-based system. Pediatrics 2004;113:252–8.
47. Reed D, Price EG, Windish DM, et al. Challenges in systematic reviews of educational intervention studies. Ann Intern Med 2005;142: 1080–9.
48. Rogers DA, Regehr G, Yeh KA, et al. Computer-assisted learning versus a lecture and feedback seminar for teaching a basic surgical technical skill. Am J Surg 1998;175:508–10.
49. Rogers DA, Regehr G, Howdieshell TR, et al. The impact of external feedback on computer-assisted learning for surgical technical skill training. Am J Surg 2000;179:341–3.
50. Mayo PH, Hackney JE, Mueck JT, et al. Achieving house staff competence in emergency airway management: results of a teaching program using a computerized patient simulator. Crit Care Med 2004;32: 2422–7.
51. Sedlack RE, Kolars JC, Alexander JA. Computer simulation training enhances patient comfort during endoscopy. Clin Gastroenterol Hepatol 2004;2:348–52.
52. Allison J, Kiefe CI, Weissman NW. Can data-driven benchmarks be used to set the goals of healthy people 2010? Am J Public Health 1999;89:61–5.
53. McGlynn EA. Selecting common measures of quality and system performance. Med Care 2003;41(Suppl):I39–47.
54. Healthy People 2010 Toolkit. A Field Guide to Health Planning. 2nd ed. Washington, DC: Public Health Foundation; 2002.
55. Kiefe CI, Weissman NW, Allison JJ, et al. Identifying achievable benchmarks of care: concepts and methodology. Int J Qual Health Care 1998;10:443–7.
56. Savoldelli GL, Naik VN, Joo HS, et al. Evaluation of patient simulator performance as an adjunct to the oral examination for senior anesthesia residents. Anesthesiology 2006;104:475–81.
57. Taffinder NJ, McManus IC, Gul Y, et al. Effect of sleep deprivation on surgeons' dexterity on laparoscopy simulator. Lancet 1998;352:1191.
58. Boursicot K, Roberts T. Setting standards in a professional higher education course: defining the concept of the minimally competent student in performance-based assessment at the level of graduation from medical school. Higher Educ Quart 2006;60:74–90.
59. Downing SM, Tekian A, Yudkowsky R. Procedures for establishing defensible absolute passing scores on performance examinations in health professions education. Teach Learn Med 2006;18:50–7.
60. Murray DJ, Boulet JR, Kras JF, et al. A simulation-based acute skills performance assessment for anesthesia training. Anesth Analg 2005; 101:1127–34.
61. Morgan PJ, Cleave-Hogg D, DeSousa S, et al. Identification of gaps in the achievement of undergraduate anesthesia educational objectives using high-fidelity patient simulation. Anesth Analg 2003;97:1690–4.
62. Schwid HA, Rooke GA, Carline J, et al. Evaluation of anesthesia residents using mannequin-based simulation: a multiinstitutional study. Anesthesiology 2002;97:1434–44.
63. Marsch SC, Tschan F, Semmer N, et al. Unnecessary interruptions of cardiac massage during simulated cardiac arrests. Eur J Anaesthesiol 2005;22:831–3.
64. Barsuk D, Ziv A, Lin G, et al. Using advanced simulation for recognition and correction of gaps in airway and breathing management skills in prehospital trauma care. Anesth Analg 2005;100:803–9.
65. Hammond J, Bermann M, Chen B, et al. Incorporation of a computerized human patient simulator in critical care training: a preliminary report. J Trauma 2002;53:1064–7.
66. Marsch SC, Tschan F, Semmer N, et al. Performance of first responders in simulated cardiac arrests. Crit Care Med 2005;33:963–7.
67. Vozenilek J, Wang E, Kharasch M, et al. Simulation-based morbidity and mortality conference: new technologies augmenting traditional case-based presentations. Acad Emerg Med 2006;13:48–53.
68. Nouraei SA, McPartlin DW, Nouraei SM, et al. Objective sizing of upper airway stenosis: a quantitative endoscopic approach. Laryngoscope 2006;116:12–7.
69. Perkins GD, Augre C, Rogers H, et al. CPREzy: an evaluation during simulated cardiac arrest on a hospital bed. Resuscitation 2005;64: 103–8.
70. Wright MC, Taekman JM, Barber L, et al. The use of high-fidelity human patient simulation as an evaluative tool in the development of clinical research protocols and procedures. Contemp Clin Trials 2005; 26:646–59.

71. Dalley P, Robinson B, Weller J, et al. The use of high-fidelity human patient simulation and the introduction of new anesthesia delivery systems. Anesth Analg 2004;99:1737–41.
72. Lim TJ, Lim Y, Liu EH. Evaluation of ease of intubation with the GlideScope or Macintosh laryngoscope by anaesthetists in simulated easy and difficult laryngoscopy. Anaesthesia 2005;60:180–3.
73. Sanderson PM, Tosh N, Philp S, et al. The effects of ambient music on simulated anaesthesia monitoring. Anaesthesia 2005;60:1073–8.
74. Lorraway PG, Savoldelli GL, Joo HS, et al. Management of simulated oxygen supply failure: is there a gap in the curriculum? Anesth Analg 2006;102:865–7.
75. Wong AK. Full scale computer simulators in anesthesia training and evaluation. Can J Anaesth 2004;51:455–64.
76. Berkenstadt H, Ziv A, Gafni N, et al. Incorporating simulation-based objective structured clinical examination into the Israeli National Board Examination in Anesthesiology. Anesth Analg 2006;102:853–8.