

# Challenges and Opportunities in Simulation and Assessment

Ronnie J. Glavin, MB, M.Phil  
David M. Gaba, MD

“Ideals, **standards**, aspirations,- those are chameleon words, and take color from their speakers.”-Carolyn Wells

What do you call the person who graduates **first** in their medical school class?  
“*The valedictorian*”

What do you call the person who graduates **last** in their medical school class?  
“*Doctor*” Old Joke

In this issue of the journal we feature an article by Boulet et al<sup>1</sup> on standard setting for examinations using simulation. Standard setting is an extremely important component of the assessment process. It is often thought of in the context of high-stakes assessments that have a tangible outcome riding on them (such as successful graduation, or specialty board certification) but both assessment and standard setting are common in many other settings. This is a good opportunity to review some challenges we face in simulation-based assessment. Most readers use “assessment” in various forms of simulation-based teaching but the assessments are predominantly of the “formative” variety, where the emphasis is on helping participants reflect upon their strengths and limitations. Even today simulation in its broad sense has a well established role in “summative” examinations (where a definitive pass/fail decision is made). Such simulation can take the form of part task trainers used in Cardiac Life Support style courses, simulated patients (“standardized patients” in US parlance) in clinical skills examinations, and medium capability simulators as used in certain Objective Structured Clinical Examinations (OSCE) stations in the primary examination of the Royal College of Anaesthetists in the UK and in the National Anesthesiology Board exams in Israel.<sup>2</sup>

There are several factors that place pressure on the health care simulation community to become involved increasingly in summative assessment. First, the development of the competency-based approach to training in the healthcare professions has moved the focus of assessment away from knowledge and simple practical skills (more suited to tests with multiple choice questions (MCQs) and simple OSCE type assessments) to more complex clinical activities where non-technical skills such as information gathering, prioritisation and effective working with other health care professionals are factors critical to successful management. Second, the momentum of the patient safety movement (to which many from the health care simulation community have made significant contributions) is also focussing on the range of countermeasures to human error with expectations that clinicians may be expected to demonstrate their abilities to a recognised standard.

To-date, many of us have shied away from involvement in summative assessments because i) we were concerned that fear of assessment would deter clinicians from participating in simulation activities; ii) we thought that the technologies and techniques of simulation were too poor to be used for this purpose. These thoughts were perfectly understandable when simulation-based teaching was new and there were many sceptical colleagues who had to be won over. On the other hand, many are now realizing that the traditional systems of assessment—based on MCQs, oral examinations, and (occasional) direct observation is itself extremely limited and cannot fairly distinguish between clinicians with adequate versus inadequate skills

From the Victoria Infirmary Glasgow, Scotland (R.J.G.) and the Department of Anesthesia, Stanford University School of Medicine, Stanford, CA (D.M.G.).

The authors have indicated that they have no conflicts of interest to disclose.

Copyright © 2008 Society for Simulation in Healthcare  
DOI: 10.1097/SIH.0b013e31817bb8f6

or performance. Moreover, simulation has now been around long enough that its characteristics are better known, and there is sufficient familiarity with it as a teaching tool that clinicians can readily understand the difference between learning simulations and examination simulations. Therefore, members of the simulation-based teaching community will now more often find themselves in a position to offer help and support to those who have the responsibility of arranging summative assessment in a variety of clinical arenas.

The first challenge in assessment is to identify those competencies or abilities that need to be assessed. There are parallels here with the process of identifying and articulating learning objectives. Those involved in simulation-based teaching do not try to cover the whole of a particular curriculum but select areas that more traditional methods of teaching may not deliver so effectively. For example, they can help trainees prepare for clinical challenges by giving them an opportunity to apply theoretical knowledge, practical skills and non-technical skills in a manner approaching the complexity of clinical practice. The individual and collective expertise in selecting the learning outcomes to target in our courses has been acquired through an iterative process of designing courses, retaining what works and eliminating what doesn't. This process in turn has helped to develop a better understanding of the strengths and limitations of simulation-based teaching.

Pressure is now coming from a variety of sources, whether external regulation or internal quality control, to provide evidence that participants on training programmes have reached a satisfactory standard. We can use the experience of selecting learning outcomes to help those charged with the responsibility of developing assessment instruments. We are moving beyond a world of high-stakes assessment that focuses on history taking, physical examination and planning interventions to a world of assessment of the management of sick patients with disordered physiology undergoing a wide variety and range of therapeutic interventions. We can approximate that environment with technological representation of human beings and so explore a set of abilities that have not featured widely in high-stakes summative assessment. In doing so we can give those responsible for development of assessment instruments a realistic sense of which parts of their curricula are likely to be suitable for simulation-based summative assessment.

A second challenge is to establish instruments and metrics to conduct the assessment (i.e. to make the measurement of those skills, behaviours, competencies, or abilities) in a fair and reliable manner. The literature is now full of papers that delineate the psychometric characteristics of different instruments in different simulation settings. While this work is by no means finished, we believe that there has in fact been sufficient work done to produce a number of useful instruments and metrics to measure a variety of important skills in a diverse set of clinical domains. While these simulation assessments are not perfect, they are no worse than our existing instruments and methods, and they certainly offer a unique window on performance that is not represented in our current assessment armamentarium.

The third challenge is to set standards to use with the instruments and metrics. Norcini and Guille define a standard as a statement about whether an examination performance is good enough for a particular purpose.<sup>3</sup> It is expressed as a special score that serves as the boundary between those who have met the standard and those who have not. Standards focus on the examinees' performances and judge them against a specific social or educational construct.<sup>3</sup> Norcini also describes the boundary in terms of an expression of professional values in the context of a test's purpose and content, the ability of the examinees and the wider social or educational setting.<sup>4</sup> The method for setting standards is a systematic way of gathering value judgements, reaching consensus and expressing that consensus as a single score on a test.<sup>4</sup> The mathematics of combining scores from sub-components of tests or from different raters can be complex, but users should never lose sight of the fact that these techniques are just ways of handling judgements. The credibility of standards will depend upon factors such as who sets them, the characteristics of the method they use, and the outcomes they produce.<sup>3</sup> Boulet et al use an "examinee centred approach." This consists of a panel of experts viewing a series of examinee performances (in this case audio-video recordings of performances) and making judgements concerning examinee proficiency or competence. The panel experts, following a training and calibration session, made an independent judgement of the quality of the examinee's performances on a binary scale of 0 (not qualified) or 1 (qualified). This process was repeated for each scenario over a range of performances. The paper highlights the amount of time and effort involved on the part of the expert panel to agree on a working definition prior to carrying out their judgements.

The mathematical handling of the data uses the experts' judgements to derive a standard—a cut-off value or "boundary." But the cut-off value does not have an independent objective reality separate from the process of judgment. Boulet et al describe the use of the point of "maximum disagreement between raters" as the place to choose the boundary cut-off value, but other values are possible and may be more relevant to the purpose of the assessment. Questions raised by the context and purpose of the assessment include: How serious are the consequences of letting through an examinee who may not be competent? How often can a competent candidate be allowed to fail? Clinicians are familiar with these issues as they apply to the definition of normal values for laboratory tests and their interpretation. Yet in the world of clinical competence and certification, these are ultimately complex "social values" that society as a whole must ultimately come to grips with. Our desire to be confident that clinical personnel are highly skilled must to some degree be traded off against our need to recruit, train, and sustain a sufficient number of them ready and willing to do the job.

Boulet et al have shown that the assessment process requires contributions from different sources—subject matter experts to provide judgement, those with educational expertise to design and manage the data handling, and the contribution of those responsible for setting up the simulation scenarios and techniques to allow reasonable and fair simulation-based assessment exercises to take place. The process of

developing and conducting summative assessment places demands that probably exceed the direct expertise of many of us involved in health care related simulation. Boulet's team, like others around the world, involves psychometricians and simulation-savvy clinicians and educators who collectively attack the three challenges of developing summative tests we outline above. We predict that there will be a growing need for such teams to lead the way in the complex processes of setting standards as healthcare embarks upon the journey from doing simulation-based training—primarily for early learners—to a career-long ongoing cycle of simulation-based training coupled with both formative and summative performance assessment.

## REFERENCES

1. Boulet JR, Murray D, Kras J, Woodhouse J. Setting performance standards for mannequin-based acute-care scenarios: an examinee-centered approach. *Sim Healthcare* 2008;3:72–81.
2. Berkenstadt H, Ziv A, Gafni N, Sidi A. Incorporating simulation-based objective structured clinical examination into the Israeli National Board Examination in Anesthesiology. *Anesth Analg* 2006;102:853–858.
3. Norcini J, Guille R. Combining Tests and Setting Standards. In: Norman GR, van der Vleuten CPM, Newble DI, eds. *International Handbook of Research in Medical Education*. Dordrecht, Netherlands: Kluwer Academic Publishers; 2002.
4. Norcini JJ. Setting standards on educational tests. *Med Educ* 2003;37:464–70.