Review article: Assessment in anesthesiology education
Article de synthèse: L'évaluation dans la formation en anesthésiologie

John R. Boulet, PhD · David Murray, MD

Abstract

Purpose *The purpose of this article is to provide a review of the competencies and associated assessment techniques relevant to the practice of anesthesiology. Although many of the competencies are difficult to define and measure specifically, advances in assessment techniques have provided more opportunities to gather meaningful performance data.*

Principal findings *Establishing the competence of anesthesiologists demands a host of measures, including standardized tests and less-structured peer evaluations. Simulation-based assessment will play an increasingly important role both in certification and in maintenance of certification for anesthesiologists.*

Conclusions *While there are many psychometric challenges associated with the assessments pertinent to the education of anesthesiologists, technological advances combined with an increased awareness of sound measurement principles will yield more meaningful competency measures that can be used to improve the practice of anesthesiology.*

Résumé

Objectif *L'objectif de cet article est de faire la synthèse des compétences et des techniques d'évaluation associées pertinentes à la pratique de l'anesthésiologie. Bien que de nombreuses compétences soient difficiles à définir et à mesurer précisément, les progrès observés dans les techniques d'évaluation nous ont donné des occasions supplémentaires de récolter des données de performance significatives.*

Constatations principales *Une importante gamme de mesures est nécessaire pour établir la compétence des anesthésiologistes, notamment des épreuves normalisées et des évaluations par les pairs moins structurées. L'évaluation basée sur la simulation jouera un rôle de plus en plus important, tant au niveau de la certification que du maintien du certificat pour les anesthésiologistes.*

Conclusion *Bien qu'il existe de nombreux défis psychométriques associés aux évaluations pertinentes à la formation des anesthésiologistes, les progrès technologiques combinés à une attention accrue aux principes de mesure solides entraîneront la création de mesures de la compétence plus significatives, lesquelles pourront être utilisées pour améliorer la pratique de l'anesthésiologie.*

J. R. Boulet, PhD (✉)
Foundation for Advancement of International Medical Education and Research, 3624 Market Street, 4th Floor, Philadelphia, PA 19104-2685, USA
e-mail: jboulet@ecfmg.org

J. R. Boulet, PhD
Educational Commission for Foreign Medical Graduates, Philadelphia, PA, USA

D. Murray, MD
Washington University School of Medicine, St. Louis, MO, USA

Many believe that the most direct approach to sustained improvement in patient care is to increase the skill and ability of healthcare professionals. It has been argued that the best available approach to improve patient safety involves fostering individual professional skills, expertise, values, responsibility, and accountability.[1] At present, the link between anesthesiologists' abilities and patient outcomes is mainly intuitive, but an expanding set of

assessment methodologies will afford new tools to measure skills effectively, assess clinical ability, and, over time, elevate practice standards. Residency programs, specialty and licensure boards, and hospital credentialling committees recognize that effective assessment is necessary to assure competence and to effect long-term improvements in practice. Therefore, even though the advanced skills of a specialist can often be difficult to evaluate, finding effective assessment methods, especially ones that can eventually lead to more capable practitioners, remains a high priority.

The lengthy process of becoming an anesthesiologist involves many types of evaluations. In both Canada and the United States, various assessments are employed to select individuals for medical school. Once accepted, students are regularly evaluated as part of the ongoing curriculum. In addition, licensing examinations are often taken both during and following medical school. The results from these examinations combined with other information may be used by program directors to select their residents. Once accepted for postgraduate training, additional assessments, some employing simulators and others involving multiple choice questions (MCQs), are used for formative educational needs. For those physicians seeking specialty board certification, further assessment by some approved authority is required. Finally, once established as an independent practicing anesthesiologist, assessment is an integral part of the maintenance of certification (MOC) and the maintenance of licensure (MOL) processes. Maintenance of certification is an ongoing process of education and assessment for board certified physicians to improve practice performance. Maintenance of licensure, sometimes referred to as revalidation, is a framework by which a regulatory authority can require physicians with active licenses to demonstrate periodically their ongoing clinical competence as a condition for licensure renewal.

The following article provides a broad overview of assessment in anesthesiology education. Since assessments are employed throughout an anesthesiologist's career, it is helpful to organize the discussion and review around the knowledge, skills, and abilities required for advanced specialty practice. Here, there are a number of potential (overlapping) frameworks that can be referenced including, amongst others, the Canadian Medical Education Directives for Specialists (CanMEDs roles)[2] and the Accreditation Council for Graduate Medical Education (ACGME) core competencies.[3-5] Both the CanMEDs roles and the ACGME core competencies define the abilities needed for practice. For the CanMEDs roles, the essential competencies are organized thematically around seven key physician roles: medical expert, communicator, collaborator, manager, health advocate, scholar, and professional. The six ACGME core competencies consist of patient care, medical knowledge, practice-based learning and improvement, interpersonal and communication skills, professionalism, and systems-based practice. To afford meaning to anesthesiology, especially with reference to assessment, these competencies need to be keyed to the particular practice characteristics of the profession.

Before discussing how specific competencies are relevant to anesthesiology and how they can be measured, a brief overview of different assessment methods and the qualities of "good" assessments is provided. As part of the section on "what can be measured", innovative approaches to assessment are highlighted. Since assessments can have both positive and negative consequences and can be challenging to administer and defend, their use and potential misuse in anesthesiology education are discussed throughout the document. Finally, given the difficulties in assessing the key competencies needed to be an effective anesthesia provider, some pressing measurement challenges are forwarded.

Types of assessments

In medical education, both at the undergraduate and post-graduate levels, many types of assessments are employed. These assessments, described in more detail elsewhere,[6-9] can be used for formative (training) or summative (certification, licensure) purposes. In general, assessments can be classified as either selected- or constructed-response. The most common selected response format is the MCQ. Here, candidates choose a response from a list that includes the correct alternative and a list of distractors. Multiple choice examinations are effective for measuring knowledge and, to some extent, clinical reasoning and clinical decision-making. Constructed-response formats, including practice-based observations,[10] are more varied and can consist of essay questions, oral presentations, objective structured clinical examinations (OSCEs), and various types of simulations, to name a few. Here, the person being assessed must construct a response through writing, orally, or performing a task (e.g., clinical procedure). Based on Miller's pyramid,[11] constructed-response formats are typically employed to assess whether a candidate knows what to do, shows what to do, or, at the highest level, actually does it.[5] While adequate knowledge and the ability to synthesize knowledge are often prerequisites for certain tasks, they are usually not sufficient for effective practice. For example, the development of an anesthetic plan requires a variety of clinical judgements and decisions based on an understanding of pharmacology, the patient history, physical examination results, and laboratory evaluations. Given the complexities of patient care, assessment formats other than MCQs, including many forms of

simulation and various workplace-based observational methods, are needed to ascertain whether a specialist is competent. Without these more "authentic" formats, it would not be possible to assess what a practitioner is actually able to do.

As a profession, anesthesiology has embraced simulation as a method to assess both procedural and "non-technical" skills, such as teamwork and situational awareness.[6,12] While cognitive-based examinations are still a fundamental component of the certification process, there has been a general recognition that simulated scenarios can provide an efficient and effective means for formative assessment. This is largely understandable because many of the most frequent causes of serious morbidity result when low-frequency events in real-life settings are ineffectively recognized and managed in the perioperative setting. In addition, with improvements in mannequin technology and advances in psychometric methods (e.g., scoring, standard setting), simulation-based assessments are slowly moving from the formative to the summative arena.[13,14] Combining various simulation modalities (e.g., standardized patients, task trainers, electromechanical mannequins) allows for a broader modelling of practice situations, making it possible to measure the multiple abilities or competencies required in anesthesia practice.[15-18] Nevertheless, while some very innovative scenario designs have been forwarded, there remains the need to ensure that student, resident, or practitioner assessments generate resulting scores or decisions that are meaningful and accurate. The qualities of "good" assessments are discussed in the next section.

Qualities of "good" assessments

Good assessments will necessarily have some positive impact on the person or persons being evaluated.[19] Ultimately, they should lead to more highly qualified practitioners and better patient care. In medicine, assessments are often employed to select the best candidates for a position (e.g., medical school, residency position) or to determine minimal competence. Regardless of their intended use, the scores or the associated decisions derived from the scores must be defensible. Ultimately, the quality of any assessment rests with the psychometric properties of the scores, namely, their reliability and validity.[20,21]

Reliability

Any assessment should yield reasonably precise scores. Depending on the nature of the assessment, the precision of the scores can be dependent on a number of factors, including the number of items/tasks, the choice (and number) of raters, and even the assessment site. When an individual undergoes an assessment, we are provided with his/her "observed" score. Generalizability or reliability is a measure of how well this observed score reflects "true" ability (i.e., the universe score – the hypothetical score if an individual is measured an infinite number of times). While beyond the scope of this article, it is important to ascertain the sources of measurement error in an assessment.[22] If MCQs are employed, error may be introduced by insufficient sampling of the content domain. If workplace-based assessments are employed, choice of rater or raters could impact the precision of scores. Where simulation scenarios are incorporated in the assessment, both the choice of tasks (simulation scenarios) and the choice of raters are likely to influence the precision of the scores. For all assessments, it is necessary to investigate the sources of measurement error. It should be noted, however, that all other things being equal, the precision of assessment scores will be highly dependent on testing time. In general, the more items on a MCQ examination or the more content-relevant tasks on a performance-based assessment, the greater the reliability of any estimates of ability.[23]

Validity

For assessment scores to be valid, they must reflect the trait or traits that one intends to measure. There are a number of guiding frameworks that can be referenced when developing strategies for gathering evidence to support the validity of assessment scores or associated decisions.[24] The *Standards for Educational and Psychological Testing* lists sources of validity under a number of broad headings: evidence based on test content, evidence based on response processes, evidence based on internal structure, evidence based on relations to other variables, and evidence based on consequences of testing.[A] When gathering evidence to support the validity of assessment scores, the intended inferences that one wishes to make based on the performance data must always be kept in mind.

While the practice of anesthesiology can involve many types of assessments, including many different formats, the steps taken to gather evidence to support the validity of the scores are similar. The first phase in the validation process often takes the form of matching the assessment content to the domain of practice. For example, if a knowledge-based MCQ examination is utilized (e.g., for anesthesiology board certification), the relevance of each content domain (e.g., inhalation anesthetic pharmacology) to practice must

---

[A] American Educational Research Association, American Psychological Association, National Council on Measurement in Education. Standards for Educational and Psychological Testing. 1999. Washington, DC, American Educational Research Association.

be established. Both logical and empirical analysis of items can be done to support content validity. Often, a job or practice analysis is undertaken to investigate the specific skills that are needed to perform adequately in the field.[25] Validity evidence based on response processes can take many forms. On an oral examination, for example, evidence needs to be gathered to show that the raters are using the evaluation criteria appropriately and are not being influenced by factors (e.g., sex, race, or training location) that are irrelevant to the intended interpretation of the scores. Gathering validity evidence based on internal structure can also be accomplished in a number of ways. On performance-based examinations (e.g., clinical simulations), for example, studies can be conducted to investigate how specific skills measured as part of the assessment are related. Here, depending on the types of clinical scenarios modelled, one might hypothesize that a procedural or practice acumen (e.g., the ability to place a double-lumen endotracheal tube or a thoracic epidural) would be related only minimally to communication skills.

Depending on the purpose of the assessment and the inferences that one wants to make based on the scores, evidence based on relationships with other variables is a key component of the validation process. Anesthesiology board certification examinations are an important determining factor in whether the specialist is ready for independent practice. Validity evidence for such assessments can take the form of predictive relationships between examination performance and practice performance and/or the documentation of performance differences between individual groups known to have differences in ability or experience.[26-29] Unfortunately, validity evidence based on consequences of testing is often ignored. Throughout the education of anesthesiologists, assessments are administered with the expectation that some benefit will be realized from the intended use of the scores or results. With the introduction of assessments for maintenance of certification in anesthesiology in the United States,[30] evidence needs to be gathered to substantiate that anesthesiologists remain qualified and patient outcomes improve. Performing these types of outcome studies is challenging, as it is often difficult to attribute patient outcomes to an individual practitioner. However, the strength of many validity arguments is severely diminished without evidence that the assessment leads to more capable practitioners and better patient care.

What can be measured?

There are a number of publications that describe the use of assessments in anesthesiology education.[6,28,31] Unfortunately, for the most part, these articles often do not make a specific link between assessment methods and the competencies required for practice as an anesthesiologist. In this section, the ACGME core competencies are used as an organizing framework for the discussion of applicable assessment methods. A similar exercise could be completed using the CanMEDs roles, but it would necessarily yield a comparable synthesis of assessment techniques and associated measurement issues. Overall, while it may be relatively straightforward to define the knowledge, skills, and aptitudes required for practice in the specialty of anesthesiology, measuring some of the associated competencies can be difficult, both logistically and psychometrically.

Patient care

In addition to those elements expected of all physicians, such as accurately gathering data, formulating a differential diagnosis, performing a relevant physical examination, and developing a safe evidence-based patient care plan, the practice of anesthesiology requires a set of skills and abilities that are particularly relevant to high-acuity settings. Attributes that are more fundamental to the practice of anesthesiology than to some other specialties include the need to: prepare and plan sequentially and efficiently combined steps to induce anesthesia, maintain vigilance, interpret monitoring data, remain situationally aware, conduct a rapid logical assessment, and make swift decisions.[4] From a patient care perspective, the perioperative and critical care environments often require or emphasize skills that differ from those typically needed or employed in other settings, such as those providing primary care. These skills must be assessed in a manner that reflects the realities of the specialty.

There are several ways to measure competencies related to patient care. These include direct observation,[10] chart reviews, and various other workplace-based evaluation methodologies.[32-34] Not surprisingly, based on the practice requirements for anesthesiology, simulation-based assessments can be particularly valuable to measure decision-making and high-acuity patient care skills in the compressed time line that frequently exists in settings such as the operating room, recovery room, or intensive care unit.[35,36] As mentioned previously, the long history of simulation in anesthesiology[37] coupled with advances in technology has effectively broadened the potential assessment domain for the specialty.[18] This affiliation has allowed for the measurement of both procedural[17] and non-technical skills such as communication, situational awareness, teamwork, and professionalism.[38] A thorough review of the use of simulation for assessment in anesthesiology can be found elsewhere.[6] However, even with the technical advances in simulation methodology, it should be emphasized that

multiple assessment techniques must be employed to measure patient care competencies effectively.

Since the measurement of patient care in anesthesiology can involve several assessment methods, numerous measurement problems surface. Workplace-based assessments that rely on observation of practitioners in clinical settings are subject to various biasing factors, including inadequate rater training, context effects, and inadequate sampling of behaviour.[39,40] Formal certification examinations (e.g., oral board examinations in anesthesiology), while often more rigorous in terms of scoring and standardization, can still suffer from a number of measurement problems. Even when the raters are sufficiently calibrated, candidates are often evaluated in a limited number of patient care situations, calling into question the generalizability of the performance to other settings or patient conditions. For example, the ability to manage an obstetric emergency for placental abruption effectively may not be a good predictor of the ability to evaluate an elderly patient with congestive heart failure who requires elective hip replacement. Without a broad sampling of behaviours across patient care situations, it may be difficult to make valid inferences concerning the abilities of those being assessed.

Medical knowledge

Medical knowledge is at the base of anesthesia practice. Without sufficient knowledge of the basic and clinical sciences, appropriate care is not possible. While many procedural skills and some clinical judgements may not demand an in-depth underlying knowledge of anatomic principles and physiological mechanisms, when variations and abnormalities are encountered, a sound knowledge base is required to choose the correct or most efficient approach or intervention. Since knowledge is a foundation for many of the other competencies, special care must be taken to ensure that it is measured adequately. Like other practitioners, an anesthesiologist must possess an adequate knowledge of biomedical, clinical, epidemiological, biomechanical, social, and behavioural sciences to make effective clinical judgements.

Compared with other competencies, the measurement of knowledge, most commonly through selected-response items, is relatively straightforward. Numerous articles have been written about the development and validation of MCQs and short answer questions, including patient management problems and other formats.[9,41,42] Since selected-response items take relatively little time to answer, measuring knowledge, from a testing perspective, can be efficient and yield reasonably precise estimates of ability. In anesthesiology, knowledge-based examinations are a fundamental part of the training, board certification, and maintenance of certification processes.[29] These types of assessments are standardized (same testing conditions for all candidates) and based on detailed content outlines, and they contain a broad sampling of items. As a result, it is possible to derive reasonably precise and valid measures of knowledge. However, as noted by Miller,[11] knowledge is at the base of the competency pyramid. It is also essential to measure the application of knowledge (e.g., assess the quality of information secured from the patient or other providers, judge the accuracy and usefulness of diagnostic screening procedures); this can be accomplished with a number of assessment methods, including computer-based case simulations.[43]

Practice-based learning and improvement

At the heart of practice-based learning and improvement is the growth in skills and insight which comes with experience.[44,45] During residency, the speed at which skill is acquired varies markedly, as does the process upon which expertise is developed. The timing of rotations and differences in interest, commitment, and confidence often make it difficult to determine whether a resident is progressing towards the goal of becoming an anesthesia consultant. Nevertheless, with additional experience and appropriate feedback, physicians with lower proficiency should gradually be better able to deal with the multitude of patient conditions encountered in practice. For anesthesiology residents, in particular, initial experiences in general anesthesia provide them with the groundwork to manage more complex specialty encounters effectively (e.g., cardiopulmonary bypass). For the specialist, practice-based learning and improvement, while potentially covering multiple skill sets, centres on the ability to enhance patient care. Amongst other requisites, an anesthesiologist must be able to interpret the meaning of different types of data, apply clinical decision rules, and use information technology to gather evidence to support or modify clinical decisions. Most importantly, the anesthesiologist must be able to implement practice-based improvement by tracking outcomes and reducing medical errors.

There are a number of ways to assess practice-based learning and improvement, including portfolios, patient records and chart reviews, and performance ratings of actual patient encounters.[46,47] Unfortunately, measuring this specific competency is fraught with measurement difficulties. First, regardless of the assessment technique, the evaluation of the resident or specialist, often based on "expert" ratings, can be highly subjective.[48] Likewise, the choice of information to include in the portfolio, patient records to evaluate, or patient encounters to observe can also impact the quality of the assessment. Those charged with assessing this specific competency must ensure that the sampling of performances (e.g., patient records) is

adequate. Second, a measurement of improvement requires that the interpretation of the results of any assessment can be compared with prior performance. When longitudinal judgements of quality are undertaken, the assessor must have an accurate frame of reference for judging the improvement, otherwise, it is impossible to make valid decisions concerning any increase in skills or abilities.

Interpersonal and communication skills

All physicians must be able to establish relationships, listen effectively, and talk about patient management options, including the discussion and disclosure of risk. While there are many definitions of communication skills, the essential elements include eliciting information, building rapport, and giving information. Anesthesiologists, like all practitioners, must also be able to document and synthesize clinical findings and diagnostic impressions effectively in written and electronic formats.

Interpersonal and communication skills in anesthesia can be complex, involving not only patients but also a host of other healthcare professionals. In high acuity settings, communication between professionals, or lack thereof, has been linked to patient safety.[49] The root cause of morbidity, while potentially dependent on many factors, such as not recognizing when to call for help or ineffective teamwork, can often be traced to poor communication amongst caregivers. To provide proper patient care in the operating room, intensive care unit, or other highly specialized care environments, anesthesiologists must possess effective communication and interpersonal skills.

There are numerous ways to measure communication skills. Most commonly, individuals are watched by colleagues or supervisors and evaluated using some form of rating scale. Alternatively, the opinions of patients can be solicited.[50,51] Unfortunately, as noted previously, all assessments that involve raters may be subject to bias. This is especially problematic for communication and interpersonal skills where specific constructs or traits are difficult to define and, arguably, are somewhat subjective with respect to interpretation. The plethora of communication rating scales and evaluation instruments supports this notion.[52-54] However, even when a well-constructed evaluation tool is used, those responsible for administering the assessment often provide little in the way of rater training. Without training, individual raters may base their evaluations on the quality of medical judgements and personal sentiments rather than key their scores to specific construct-related attributes.[55,56]

In anesthesiology, more structured forms of simulation-based assessment can also be employed to measure interpersonal and communication skills.[57,58] For these types of performance assessments, often employing both confederates (e.g., surgeons, nurses) and electromechanical mannequins, the administration conditions can be standardized and modelled to represent actual patient encounters. If constructed correctly, simulation-based evaluations provide a unique opportunity to measure communication skills that cannot be measured using other lower-fidelity assessment formats; these simulations cannot, however, replace the observation and evaluation of anesthesiologists in practice. While there is some evidence that doctor-patient communication skills measured in a simulated environment generalize to practice situations,[59] interpersonal and communication skills in anesthesia can be complex, involving not only patients but also a host of other healthcare professionals. As a result, the conditions under which communication skills among healthcare professionals, as measured in the simulated environment, generalize to actual practice situations have yet to be fully delimited.

While not as much the focus of assessment research as oral communication, written communication is also an important part of practice. The ability to document relevant history and physical exam findings and produce a differential diagnosis and management plan is currently measured as part of the certification and licensure of physicians in Canada and the United States.[13] For practicing anesthesiologists, it would make sense to assess this competency via chart reviews or, where possible, from electronic medical records (EMRs). However, unlike OSCEs where the patient presentation and conditions are fixed, there is no "gold standard" for establishing the adequacy of the documentation (charting) of the information or diagnostic hypotheses associated with "real" patient encounters. As a result, it can be difficult to make judgements concerning the quality of patient care from written reports.

Professionalism

Carrying out professional responsibilities, adhering to ethical principles, and being sensitive to a diverse patient population are key competencies for any specialist. In dealing with patients and other healthcare professionals, anesthesiologists must be altruistic and respectful, keeping the best interests of the patient at heart. However, while patients, physicians, and healthcare workers would all agree that "professionalism" is a desired trait, there is no clear consensus concerning the specific behavioural characteristics that could be used to delimit someone who is competent, based solely on professional attributes, from someone who is not. Moreover, while some criteria are relatively generic (e.g., ethically sound practice, social accountability), others (e.g., cultural sensitivity) may be context-specific and open to interpretation. Finally, and

most importantly, professionalism relates to many abilities,[60] making it difficult to obtain a pure measure of this competency. Nevertheless, there are numerous measurable aspects of professionalism, including working with colleagues in ways that serve the best interests of the patient, honouring patient boundaries, accepting personal errors, avoiding substances that may interfere with judgement when caring for patients, punctuality, organization, and preparedness.

There are several methods to measure professionalism and most involve some form of peer assessment or rating.[61-64] While professionalism rating scales are available[53,65] and have been employed in OSCEs and as part of peer evaluations,[66] they are often difficult to administer. Many aspects of professionalism are difficult to define, at least in terms of specific behaviours. Furthermore, when professional attributes are measured as part of a standardized assessment (e.g., OSCEs), they likely provide an inflated estimate of the practitioner's overall level of "professionalism". The manner in which a person behaves when being observed (or filmed) as part of a structured assessment can be quite different from how they may act in an everyday encounter with a patient or other healthcare worker. Peer assessments have also been advocated as a means to evaluate professionalism.[67] If there is a sufficient sampling of peers and the assessment is properly conducted, it is possible to separate those individuals who possess high moral and ethical standards from those who do not.[68]

Systems-based practice

Systems-based practice is manifested through actions that demonstrate an awareness of and responsiveness to the larger context of the system of healthcare and the ability to call on available resources effectively in order to provide care of optimal value.[51,69] Anesthesiologists are required to make appropriate patient care decisions relative to the characteristics of the healthcare system, function in inter-professional teams, make cost-effective decisions, overcome logistical barriers to patient care, and intervene in a timely and effective manner when patient safety may be compromised.

Given the diversity of skills associated with systems-based practice, many different types of assessments may be applicable. With the growth and improved sophistication of EMR systems, it should be possible to investigate and measure competence with regard to the provision of effective and efficient patient care, at least for some conditions and some providers. For anesthesiology, where some actions have direct measurable consequences (e.g., administering an anesthetic), the availability of the EMR can provide the means to establish cause and effect relationships, offering another tool to measure practice effectiveness and efficiency.

Arguably, one of the most important system-based practice competencies is teamwork. To improve patient safety and quality of care, anesthesiologists must forge interdependent relationships with many healthcare professionals. Also, as part of teams, they must be able to provide backup when other individuals fail to provide optimal care. From a systems-based practice perspective, an anesthesiologist's failure to manage malignant hyperthermia effectively or an inability to direct team members in responding to a difficult airway may be recognized as team failures, but the shortcomings may relate to the limitations of individual team members and their communication or teamwork skills. Although individual caregivers often cannot choose with whom they work, it remains important to measure both the team as a whole and individuals within the team. From the assessor's perspective, this process can provide valuable information on individual deficiencies, intra- and inter-professional difficulties, and system-based problems associated with the delivery of appropriate patient care.

Although peer assessment can be useful for evaluating inter-professional skills such as teamwork, the use of structured simulation scenarios or videotaped performances provides a standardized milieu in which to evaluate individual practitioners as they interact with patients and other healthcare workers.[58,70-72] Here, specific attributes (e.g., communication, leadership) can be codified, allowing for structured feedback. In medicine, research efforts have recently been directed at modelling team-based clinical scenarios and using these to measure both individual and group proficiencies.[73,74] These efforts will certainly add to the quality of measurement tools needed to obtain reliable and valid assessments of the skills associated with systems-based practice.

Measurement issues and future directions

The assessment of medical students, residents, and practicing physicians has certainly evolved over the last few decades. With respect to knowledge assessment, the use of selected-response items continues both during training and as part of the certification process. In addition to the typical MCQ format of choosing the correct answer from a list of distractors (A-type items), other formats are now employed (R-type items and G sets).[75] The utilization of these newer item formats provides an opportunity to measure higher-order thinking, including clinical decision-making. Likewise, in the field of performance evaluation, the introduction of various simulation modalities has greatly expanded the potential assessment domain. Unfortunately,

while advances in technology now allow for a more expansive measurement of the competencies needed for effective practice in anesthesiology, there are still many logistical and psychometric concerns that need to be addressed.[76]

Content underrepresentation and the generalizability of skills

Anesthesiology has taken a leading role in the development of simulation-based assessment. Nevertheless, while certain competencies (e.g., procedural skills in patient care) may be easier to measure in a standardized controlled environment, there are still many practice environments, conditions, and interactions that are difficult to model. Situations involving teams or the longitudinal management of patients present many logistical and measurement difficulties, including the separation of an individual's abilities from that of their coworkers, and the integration of patient histories over time. At present, while simulation affords many measurement opportunities across all the core competencies, it does not negate the need for evaluations of trainees and practitioners in "real" patient encounters.

A more pressing concern with simulation and all other assessment methods is the accrual of evidence to suggest that skills measured in one situation generalize to other situations.[77] For communication and interpersonal skills, at least for common doctor-patient interactions, it is likely that competence in one patient care situation generalizes to another. However, there may be situations, especially those involving acute care interventions, where the more general communication strategies, which are effective for doctor-patient communication, do not necessarily apply. Likewise, communication between an anesthesiologist and a surgeon may be categorically different from the communication between an anesthesiologist and a patient. Finally, at least for some situations, the context (e.g., situations involving poor patient prognosis) could have an appreciable impact on the measurement of certain competencies. As a result, regardless of the specific competency or competencies being evaluated, multiple measures gathered from multiple assessments at multiple intervals may be needed to yield stable ability estimates.

"Objective" vs "subjective" measurement

In medicine, there is often a distinction made between objective and subjective measures. Objective measures are typically based on analytical scoring rubrics (e.g., checklist items, key actions), correct answers (e.g., MCQ examination), or specific actions performed (e.g., checklists or key actions for a simulation scenario). Subjective measures generally involve expert ratings of some behavioural aspect

of performance (e.g., communication skills, professionalism). Depending on the competency being evaluated, objective or subjective measures may be more appropriate. However, it is unfortunate that the objective/subjective categorization schema is employed. Given the complexities of and interrelationships amongst some of the core competencies, subjective measures may often provide less biased, more generalizable, and appropriately valid indicators of ability.[78,79]

The use of objective performance measures for evaluating some competencies (e.g., patient care) is commonplace. For OSCEs, part-task trainers, and electromechanical mannequins, checklists or key actions are typically employed. However, while these types of rubrics can be scored objectively, their content can be open to debate. Even though expert panels are often employed to delimit checklist content, typically via some Delphi process,[80] the agreement on the actions necessary for patient care can vary as a function of the experience and expectations of the panellists and a lack of agreement about what constitutes best practice. Though there can be general agreement regarding checklist content and accurate scoring, the scores may still not reflect the intended ability. In many acute care situations, it is important to consider not only what the anesthesiologist does but also the order and timing of the actions. When the latter is not accounted for or when egregious actions cannot be factored into the scoring system, the use of objective measures may lead to questions concerning the validity of any resultant scores.

For other competencies (e.g., professionalism) the use of subjective measures would appear to be more appropriate. Rating scales, while sometimes subject to the nuances and biases of individual raters, are often the only reasonable method to gather meaningful data for some competencies. It is unfortunate that these ratings, however procured, are typically labelled as subjective measures. From a psychometric perspective, the validity of measures of some competencies, provided they are adequately defined, is often enhanced by employing more holistic evaluations where both positive and negative behaviours can be considered. More importantly, the subjectivity of the rating process can often be controlled through the specification of behavioural benchmarks and the incorporation of structured rater training regimes.[17] In many situations, the subjectivity of the ratings is simply a function of the raters not knowing exactly what is being measured.

Technical vs non-technical skills of anesthesiologists

The abilities of anesthesiologists are often crudely classified into two categories, technical skills and non-technical skills. Technical skills can encompass knowledge and procedures. Non-technical skills are more relevant to competencies such

as interpersonal and communication skills, professionalism, and systems-based practice and generally involve constructs that are difficult to define and measure.

In practice, anesthesiologists need both technical and non-technical skills. However, from a competency perspective, the integration of technical and non-technical skills is paramount. For example, some competencies (e.g., systems-based practice) demand a working knowledge of epidemiology, hospital administration, and consultation practices. Interleaving this knowledge with sound patient care practice can sometimes blur the boundary between technical and non-technical skills. Most importantly from an assessment perspective, while it may sometimes be less cumbersome to measure technical or non-technical skills of anesthesiologists in isolation, their combination with respect to the assessment and management of patients defines, at least in a global sense, competence within the specialty.

Minimal competency?

While several frameworks define the competencies needed to provide safe and effective patient care, relatively little work has been dedicated to defining minimal practice standards. For knowledge-based examinations, especially those used for summative purposes (e.g., board certification), there are a host of validated standard setting techniques.[81,82] For performance-based assessments (e.g., multi-scenario simulations), some work has been conducted to develop appropriate standard setting methodologies.[83] Nevertheless, outside the areas of standardized assessment, it is not clear how judgements of minimal competence should or could be made.

To evaluate some competencies, peer assessments, patient assessments, and portfolios are often employed. While these assessments are typically used formatively to gather information to provide feedback, questions concerning minimal competence may still arise. For example, if a 360° peer evaluation is employed to judge the professional attributes of resident anesthesiologists, is it sufficient to rank order only those residents who are being evaluated and to provide some form of remediation activity for those at the bottom of the class? Is there a minimal rating or ranking where secondary assessments are warranted? Overall, while many assessment techniques can be employed to measure the competencies of anesthesiologists, evaluators must put some thought into defining and establishing minimal performance standards.

Conclusion

The process of becoming an anesthesiologist demands that individuals develop and maintain specific competencies.

These competencies, however classified, can be measured using a variety of assessment tools, including simulation. It should be noted, however, that given the complexity of patient care in anesthesiology, it is often difficult to measure specific competencies in isolation. Moreover, at least from a psychometric perspective, some competencies (e.g., medical knowledge) are certainly easier to evaluate than others (e.g., systems-based practice, professionalism). Nevertheless, a series of assessments, if properly constructed, can ensure the adequacy of educational programs and the quality of physicians who enter and practice in the specialty. Those involved in educating, training, and certifying anesthesiologists must secure evidence to support the validity and reliability of their assessment scores or associated decisions based on the scores. Most importantly, data must be gathered to link the results of competency-based assessments and the contingent qualities of the practitioners to patient outcomes.

Key points

- Assessment plays a fundamental role in the education of anesthesiologists.
- Many different types of assessments are needed to measure the competencies of anesthesia providers.
- Development of sound assessment practices can help ensure the safe and effective provision of care.

References

1. *Hofer TP*, *Hayward RA*. Are bad outcomes from questionable clinical decisions preventable medical errors? A case of cascade iatrogenesis. Ann Intern Med 2002; 137(5 Part 1): 327-33.
2. *Frank JR*. The CanMEDS 2005 Physician Competency Framework. Better standards. Better physicians. Ottawa, Canada. The Royal College of Physicians and Surgeons of Canada. Available from URL: http://meds.queensu.ca/medicine/obgyn/pdf/CanMEDS2005.booklet.pdf (accessed August 2011).
3. *Accreditation Council for Graduate Medical Education*. ACGME Outcome Project. Available from: http://www.acgme.org/acWebsite/home/home.asp (accessed August 2011).
4. *Glavin R*, *Flin R*. Review article: The influence of psychology and human factors on education in anesthesiology. Can J Anesth 2012; 59: this issue. DOI:10.1007/s12630-011-9634-z.
5. *Bould MD*, *Naik VN*, *Hamstra SJ*. Review article: New directions in medical education related to anesthesiology and perioperative medicine. Can J Anesth 2012; 59: this issue. DOI:10.1007/s12630-011-9633-0.

6. *Boulet JR*, *Murray DJ*. Simulation-based assessment in anesthesiology: requirements for practical implementation. Anesthesiology 2010; 112: 1041-52.

7. *Davis MH*. OSCE: the Dundee experience. Med Teach 2003; 25: 255-61.

8. *Norcini JJ*, *McKinley DW*. Assessment methods in medical education. Teaching and Teacher Education 2007; 23: 239-50.

9. *Epstein RM*. Assessment in medical education. N Engl J Med 2007; 356: 387-96.

10. *Kogan JR*, *Holmboe ES*, *Hauer KE*. Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. JAMA 2009; 302: 1316-26.

11. *Miller GE*. The assessment of clinical skills/competence/performance. Acad Med 1990; 65(9 Suppl): S63-7.

12. *Edler AA*, *Fanning RG*, *Chen MI*, et al. Patient simulation: a literary synthesis of assessment tools in anesthesiology. J Educ Eval Health Prof 2009; 6: 3.

13. *Boulet JR*, *Smee SM*, *Dillon GF*, *Gimpel JR*. The use of standardized patient assessments for certification and licensure decisions. Simul Healthc 2009; 4: 35-42.

14. *Dillon GF*, *Boulet JR*, *Hawkins RE*, *Swanson DB*. Simulations in the United States Medical Licensing Examination (USMLE). Qual Saf Health Care 2004; 13 Suppl 1: i41-5.

15. *Nestel D*, *Kneebone R*, *Black S*. Simulated patients and the development of procedural and operative skills. Med Teach 2006; 28: 390-1.

16. *Amin Z*, *Boulet JR*, *Cook DA*, et al. Technology-enabled assessment of health professions education: consensus statement and recommendations from the Ottawa 2010 conference. Med Teach 2011; 33: 364-9.

17. *Bould MD*, *Crabtree NA*, *Naik VN*. Assessment of procedural skills in anaesthesia. Br J Anaesth 2009; 103: 472-83.

18. *Leblanc VR*. Review article: Simulation in anesthesia: state of the science and looking forward. Can J Anesth 2012; 59: this issue. DOI:10.1007/s12630-011-9638-8.

19. *Norcini J*, *Anderson B*, *Bollela V*, et al. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. Med Teach 2011; 33: 206-14.

20. *Cook DA*, *Beckman TJ*. Current concepts in validity and reliability for psychometric instruments: theory and application. Am J Med 2006; 119: 166.e7.

21. *Devitt JH*, *Kurrek MM*, *Cohen MM*, *Cleave-Hogg D*. The validity of performance assessments using simulation. Anesthesiology 2001; 95: 36-42.

22. *Boulet JR*. Generalizability theory: basics. *In*: Everitt BS, Howell D (Eds). Encyclopedia of Statistics in Behavioral Science. Chichester: John Wiley & Sons, Ltd; 2005. 704-11.

23. *Boulet JR, Swanson DB*. Psychometric challenges of using simulations for high-stakes assessment. *In*: Dunn WF (Ed). Simulators in Critical Care Education and Beyond. Des Plains, IL: Society of Critical Care Medicine; 2004: 119-30.

24. *Clauser BE*, *Margolis MJ*, *Swanson DB*. Issues of validity and reliability for assessments in medical education. *In*: Holmboe ES, Hawkins RE (Eds). Practical Guide to the Evaluation of Clinical Competence, 1st ed. Philadelphia, PA: Mosby/Elsevier; 2008: 10-23.

25. *Raymond MR*. Job analysis and the specification of content for licensure and certification examinations. Applied Measurement in Education 2001; 14: 369-415.

26. *Kearney RA*, *Sullivan P*, *Skakun E*. Performance on ABA-ASA in-training examination predicts success for RCPSC certification. American Board of Anesthesiology-American Society of Anesthesiologists. Royal College of Physicians and Surgeons of Canada. Can J Anesth 2000; 47: 914-8.

27. *Gale TC*, *Roberts MJ*, *Sice PJ*, et al. Predictive validity of a selection centre testing non-technical skills for recruitment to training in anaesthesia. Br J Anaesth 2010; 105: 603-9.

28. *Murray DJ*, *Boulet JR*, *Avidan M*, et al. Performance of residents and anesthesiologists in a simulation-based skill assessment. Anesthesiology 2007; 107: 705-13.

29. *McClintock JC*, *Gravlee GP*. Predicting success on the certification examinations of the American Board of Anesthesiology. Anesthesiology 2010; 112: 212-9.

30. The American Board of Anesthesiology. Maintenance of certification in anesthesiology (MOCA). Available from: http://www.theaba.org/Home/anesthesiology_maintenance (accessed July 2011).

31. *Ottestad E*, *Boulet JR*, *Lighthall GK*. Evaluating the management of septic shock using patient simulation. Crit Care Med 2007; 35: 769-75.

32. *Wilkinson JR*, *Crossley JG*, *Wragg A*, *Mills P*, *Cowan G*, *Wade W*. Implementing workplace-based assessment across the medical specialties in the United Kingdom. Med Educ 2008; 42: 364-73.

33. *Miller A*, *Archer J*. Impact of workplace based assessment on doctors' education and performance: a systematic review. BMJ 2010; 341: c5064.

34. *Norcini J*, *Burch V*. Workplace-based assessment as an educational tool: AMEE Guide No. 31. Med Teach 2007; 29: 855-71.

35. *Berkenstadt H*, *Erez D*, *Munz Y*, *Simon D*, *Ziv A*. Training and assessment of trauma management: the role of simulation-based medical education. Anesthesiol Clin 2007; 25: 65-74.

36. *Hunt EA*, *Walker AR*, *Shaffner DH*, *Miller MR*, *Pronovost PJ*. Simulation of in-hospital pediatric medical emergencies and cardiopulmonary arrests: highlighting the importance of the first 5 minutes. Pediatrics 2008; 121: e34-43.

37. *Cooper JB*, *Taqueti VR*. A brief history of the development of mannequin simulators for clinical education and training. Qual Saf Health Care 2004; 13 Suppl 1: i11-8.

38. *Flin R*, *Patey R*, *Glavin R*, *Maran N*. Anaesthetists' non-technical skills. Br J Anaesth 2010; 105: 38-44.

39. *Schuwirth LW*, *Southgate L*, *Page GG*, et al. When enough is enough: a conceptual basis for fair and defensible practice performance assessment. Med Educ 2002; 36: 925-30.

40. *Downing SM*, *Haladyna TM*. Validity threats: overcoming interference with proposed interpretations of assessment data. Med Educ 2004; 38: 327-33.

41. *Wass V*, *Van der Vleuten C*, *Shatzer J*, *Jones R*. Assessment of clinical competence. Lancet 2001; 357: 945-9.

42. *Epstein RM*, *Hundert EM*. Defining and assessing professional competence. JAMA 2002; 287: 226-35.

43. *Dillon GF*, *Clyman SG*, *Clauser BE*, *Margolis MJ*. The introduction of computer-based case simulations into the United States Medical Licensing Examination. Acad Med 2002; 77(10 Suppl): S94-6.

44. *Moore DE Jr*, *Pennington FC*. Practice-based learning and improvement. J Contin Educ Health Prof 2003; 23 Suppl 1: S73-80.

45. *Lynch DC*, *Swing SR*, *Horowitz SD*, *Holt K*, *Messer JV*. Assessing practice-based learning and improvement. Teach Learn Med 2004; 16: 85-92.

46. *Carraccio C*, *Englander R*. Evaluating competence using a portfolio: a literature review and web-based application to the ACGME competencies. Teach Learn Med 2004; 16: 381-7.

47. *Harden RM*. Trends and the future of postgraduate medical education. Emerg Med J 2006; 23: 798-802.

48. *Graham J*, *Hocking G*, *Giles E*. Anaesthesia non-technical skills: can anaesthetists be trained to reliably use this behavioural marker system in 1 day? Br J Anaesth 2010; 104: 440-5.

49. *Canas M*, *Moreno R*, *Rhodes A*, *Grounds RM*. Patient safety in anesthesia. Minerva Anestesiol 2010; 76: 753-7.

50. *Evans RG*, *Edwards A*, *Evans S*, *Elwyn B*, *Elwyn G*. Assessing the practising physician using patient surveys: a systematic review of instruments and feedback methods. Fam Pract 2007; 24: 117-27.

51. *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: 301-9.

52. *Duffy FD, Gordon GH, Whelan G, et al*. Assessing competence in communication and interpersonal skills: the Kalamazoo II report. Acad Med 2004; 79: 495-507.

53. *Symons AB, Swanson A, McGuigan D, Orrange S, Akl EA*. A tool for self-assessment of communication skills and professionalism in residents. BMC Med Educ 2009; 9: 1.

54. *Ong LM, de Haes JC, Hoos AM, Lammes FB*. Doctor-patient communication: a review of the literature. Soc Sci Med 1995; 40: 903-18.

55. *Nicolai J, Demmel R*. The impact of gender stereotypes on the evaluation of general practitioners' communication skills: an experimental study using transcripts of physician-patient encounters. Patient Educ Couns 2007; 69: 200-5.

56. *Schirmer JM, Mauksch L, Lang F, et al*. Assessing communication competence: a review of current tools. Fam Med 2005; 37: 184-92.

57. *Mercer SJ, Moneypenny MJ, Guha A*. Communication and simulation for anaesthetists. Anaesthesia 2009; 64: 1259-60.

58. *Park CS*. Simulation and quality improvement in anesthesiology. Anesthesiol Clin 2011; 29: 13-28.

59. *Whelan GP, McKinley DW, Boulet JR, Macrae J, Kamholz S*. Validation of the doctor-patient communication component of the Educational Commission for Foreign Medical Graduates Clinical Skills Assessment. Med Educ 2001; 35: 757-61.

60. *Smith AF, Greaves JD*. Beyond competence: defining and promoting excellence in anaesthesia. Anaesthesia 2010; 65: 184-91.

61. *Lynch DC, Surdyk PM, Eiser AR*. Assessing professionalism: a review of the literature. Med Teach 2004; 26: 366-73.

62. *Wilkinson TJ, Wade WB, Knock LD*. A blueprint to assess professionalism: results of a systematic review. Acad Med 2009; 84: 551-8.

63. *Veloski JJ, Fields SK, Boex JR, Blank LL*. Measuring professionalism: a review of studies with instruments reported in the literature between 1982 and 2002. Acad Med 2005; 80: 366-70.

64. *Hodges BD, Ginsburg S, Cruess R, et al*. Assessment of professionalism: recommendations from the Ottawa 2010 Conference. Med Teach 2011; 33: 354-63.

65. *Norcini JJ, Blank LL, Duffy FD, Fortna GS*. The mini-CEX: a method for assessing clinical skills. Ann Intern Med 2003; 138: 476-81.

66. *van Zanten M, Boulet JR, Norcini JJ, McKinley D*. Using a standardised patient assessment to measure professional attributes. Med Educ 2005; 39: 20-9.

67. *Norcini JJ*. Peer assessment of competence. Med Educ 2003; 37: 539-43.

68. *Massagli TL, Carline JD*. Reliability of a 360-degree evaluation to assess resident competence. Am J Phys Med Rehabil 2007; 86: 845-52.

69. *Heffron MG, Simspon D, Kochar MS*. Competency-based physician education, recertification, and licensure. WMJ 2007; 106: 215-8.

70. *Hingle ST, Robinson S, Colliver JA, Rosher RB, McCann-Stone N*. Systems-based practice assessed with a performance-based examination simulated and scored by standardized participants in the health care system: feasibility and psychometric properties. Teach Learn Med 2011; 23: 148-54.

71. *Gaba DM*. Crisis resource management and teamwork training in anaesthesia. Br J Anaesth 2010; 105: 3-6.

72. *Wright MC, Phillips-Bute BG, Petrusa ER, Griffin KL, Hobbs GW, Taekman JM*. Assessing teamwork in medical education and practice: relating behavioural teamwork ratings and clinical performance. Med Teach 2009; 31: 30-8.

73. *Volk MS, Ward J, Irias N, Navedo A, Pollart J, Weinstock PH*. Using Medical Simulation to Teach Crisis Resource Management and Decision-Making Skills to Otolaryngology Housestaff. Otolaryngol Head Neck Surg 2011; 145: 35-42.

74. *Morgan PJ, Pittini R, Regehr G, Marrs C, Haley MF*. Evaluating teamwork in a simulated obstetric environment. Anesthesiology 2007; 106: 907-15.

75. *Case SM, Swanson DB*. Constructing Written Test Questions for the Basic and Clinical Sciences, Third Edition (revised). Philadelphia, PA: National Board of Medical Examiners; 2010 .

76. *Edler AA*. The use of simulation education in competency assessment: more questions than answers. Anesthesiology 2008; 108: 167.

77. *Weller JM, Jolly B, Robinson B*. Generalisability of behavioural skills in simulated anaesthetic emergencies. Anaesth Intensive Care 2008; 36: 185-9.

78. *Hodges B, Regehr G, McNaughton N, Tiberius R, Hanson M*. OSCE checklists do not capture increasing levels of expertise. Acad Med 1999; 74: 1129-34.

79. *Hodges B, McNaughton N, Regehr G, Tiberius R, Hanson M*. The challenge of creating new OSCE measures to capture the characteristics of expertise. Med Educ 2002; 36: 742-8.

80. *Scavone BM, Sproviero MT, McCarthy RJ, et al*. Development of an objective scoring system for measurement of resident performance on the human patient simulator. Anesthesiology 2006; 105: 260-6.

81. *De Champlain AF*. Ensuring that the competent are truly competent: an overview of common methods and procedures used to set standards on high-stakes examinations. J Vet Med Educ 2004; 31: 61-5.

82. *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.

83. *Boulet JR, Murray D, Kras J, Woodhouse J*. Setting performance standards for mannequin-based acute-care scenarios: an examinee-centered approach. Simul Healthc 2008; 3: 72-81.