



Competency Assessment

6

Anjan Shah, Samuel DeMaria, and Andrew Goldberg

Introduction

Competency assessment is a comprehensive and often controversial topic given the challenges inherent to the fair assessment of a practitioner's knowledge, skills, and attributes. Currently, simulation has been identified as a critical educational tool but is increasingly being used for high-stake competency assessments. High-fidelity simulation offers flexibility, realism, and inherent patient safety that makes it ideal for the assessment of undergraduate, graduate, and postgraduate anesthesiology providers, whether in the setting of residency training (or anesthesiology assistant or nurse anesthetist training), Maintenance of Certification in Anesthesiology (MOCA) courses, and even the retraining of physicians (or certified registered nurse anesthetists) seeking reentry to clinical practice [1]. Over the years, significant research and experience have been accumulated on assessment procedures in general, as well as the specifics of simulation-based assessment [2–4]. The goal of this chapter is not only to lay out the importance of assessment to the field of anesthesiology and a framework for how to approach simulation-based assessment but also to understand the complexities that must be overcome when dealing with practitioner assessment.

Competency Assessment

Competency assessment of any practitioner is multifactorial and poses significant challenges with regard to controlling the assessment between individuals. Since there are a wide

variety of tools that can be used for assessment purposes (each one with different strengths and weaknesses), reproducibility becomes a central concern. As such, two critically important concepts must be considered when attempting to utilize an assessment tool, namely reliability and validity.

Reliability

In general, reliability refers to the consistency of a measure when utilized under similar conditions. *Test-retest reliability* refers to the degree to which test scores are consistent from one administration event to the next, in which there is a single rater using the same methods and instruments under the same testing conditions. When relating it to assessment in academic performance, it refers to the ability of a method of assessment in consistently producing the scores, when administered multiple times under comparable conditions. This is a critical point to consider if an assessment tool is going to become universal.

Consider an analogy of two people, A and B, throwing darts at a dartboard where different sections have different point values assigned to them. Person A consistently hits the center mark, while Person B consistently hits the bottom right corner of the dart board. Both of these people, undergoing two individual assessments, would be classified as reliable. If these same subjects undergo the same exact assessment several times, but at different times, they each would receive the same exact score as their previous tries if the test had good test-retest reliability (Fig. 6.1).

Validity

Validity refers to the degree of accuracy of what is being assessed. The purpose of validation is to gather evidence that evaluates whether a decision is useful. An analogy to the validity argument is an investigator examining a crime

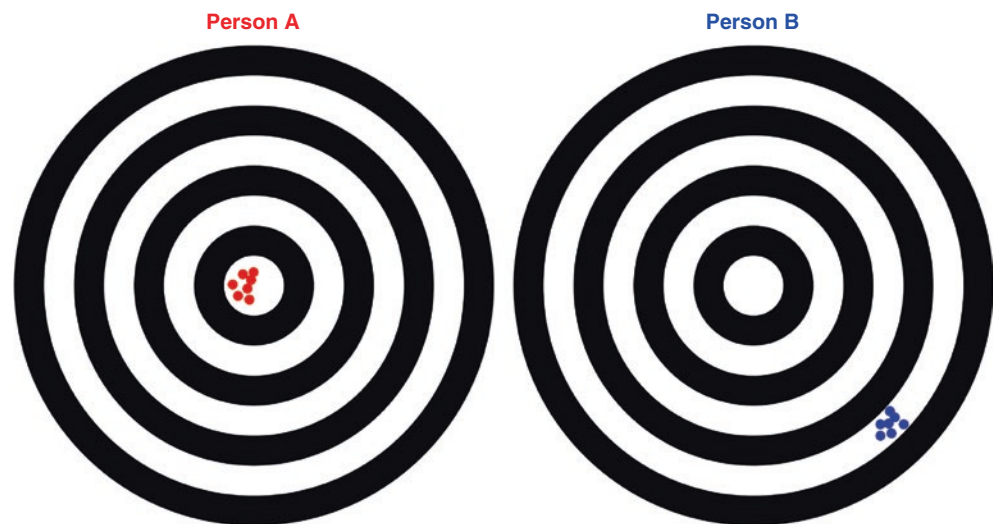
A. Shah · S. DeMaria (✉)

Department of Anesthesiology, Perioperative and Pain Medicine,
Icahn School of Medicine at Mount Sinai, New York, NY, USA
e-mail: Samuel.demaria@mssm.edu

A. Goldberg

Department of Anesthesiology, James J Peters VA Medical Center,
New York, NY, USA

Fig. 6.1 Two individuals, Person A and Person B, throwing darts at separate targets. Both demonstrate high reliability as the results are consistent over seven tries. Accuracy depends on what exactly is being measured. Person A is more accurate than Person B if the center of the target was the goal



scene: the investigator is looking for a wide variety of evidence that may link the defendant to the crime, such as DNA evidence on hair or blood, fingerprints on the weapon used, footage from nearby cameras, eye-witness interviews, interviews with known acquaintances, etc. The prosecuting attorney must then organize all the data collected and make an argument to the jury that the defendant is guilty of the said crime by presenting all the evidence and interpretations of the evidence, in the hopes of persuading the jury into making the wanted decision [5, 6]. Unlike reliability, validity is not a characteristic of the data collected from an assessment but rather a characteristic from the interpretations made from the collected data as representations of the truth. Although the process of building a successful assessment tool is rigorous and involves much more than just optimizing and controlling for reliability and validity, weakness in these two domains will negatively impact the results of an assessment. The use of simulations (including simulators, virtual-reality devices, part-task trainers, and standardized patients) in healthcare education as an assessment tool provides high reliability from the ease of manipulation of physical parameters and delivery of consistent information, as well as providing strong validity evidence by controlling parameters in the weaker inferences in the validity argument (to be explained in greater detail below).

Kane's Validity Framework

According to Kane, when building an assessment tool, the process of validation cannot begin until both the purpose of an assessment and the use of its scores are specified. Once the purpose has been clearly stated, Kane lays out a two-step process to build a validity argument—stating the claims to be made in what he calls the interpretation/use argument (IUA) [7] and then evaluating each of the claims by moving

through the four inferences: scoring, generalization, extrapolation, and implication:

1. *Scoring inference*—this relates to an observation about a performance. Evidence accumulated in regard to this inference evaluates whether a standardized protocol was used in establishing scores for the encounter. This includes a set scoring rubric applied correctly, as well as the exam being performed under appropriate and specified conditions.
2. *Generalization inference*—in practicality, there is a finite number of questions or stations that one can make observations on, but in theory, there is an infinite amount. Generalization inference relies on taking the scores from a select sample of items in the assessment(s) and applying them to the bottomless pool in the assessment universe. For qualitative assessments, this includes forming an accurate and descriptive narrative from singular pieces of qualitative data. As Cook et al. mentions, interrater variability for qualitative assessments may be a differing perspective and give alternative insights into a performance rather than an error for numeric scores as with quantitative measures [6, 8, 9]. Evidence gathered under this inference addresses the actual construction of the assessment and assumes that reliability issues of internal consistency were checked.
3. *Extrapolation inference*—the real goal of competence assessments is to be able to predict performance in the real-world clinical setting. Evidence here is used to confirm or refute the relationship between scores on the assessment and the outcomes that stakeholders are interested in.
4. *Decision inference*—evidence gathered here will lead to a final decision regarding the purpose of the assessment. The decision will depend on the stakeholders at play and the assumption that the implications of the said decision (both intended and unintended) were considered.

A Framework for Assessment: Outcomes and Levels of Assessment, Stages of Development, and Context

Once all of these inferences have been considered, the next step in creating a successful assessment is determining the framework in which it will be administered. There are a wide variety of assessment tools available, and they are now generally classified into broad categories (see Table 6.1). It is important to consider several aspects of the evaluation/assessment prior to choosing an appropriate assessment modality. The overall purpose of an assessment should be stated clearly as this provides the context in which the assessment will occur, as well as the stakes of the assessment. In addition, other components to consider in this framework include the specific outcomes to be assessed, the appropriate levels of assessment, and the developmental stage of those who are being assessed.

Outcomes of Assessment

As will be mentioned in greater detail later, in the United States, the Accreditation Council of Graduate Medical Education (ACGME) has listed six core competencies (patient care and procedural skills, medical knowledge, practice-based learning and improvement, interpersonal and communication skills, professionalism, and system-based practice) that a general physician must be competent in by the end of medical education [10]. Specific skills within each competency can be tailored to specific specialties and their needs and challenges. For example, referring to *The Anesthesiology Milestone Project*, one can see specific outcomes expected of anesthesiology residents in each of the six core competency domains, and each one is graded from level 1 to level 5: preanesthetic evaluation, crisis management,

management of a critically ill patient in a nonoperative setting, coordination of patient care, team and leadership skills, etc. [11]. The ACGME has also created a list of assessment methods matched to each outcome being assessed within a core competency as a guide to those performing assessments [12]. Consequently, one path to choosing an assessment method would be to clearly identify an outcome to be measured and to match it with an appropriate modality.

Levels of Assessment

There are four levels of assessment in the model first described by George Miller: The first level, *knows*, represents the knowledge base of medical facts and physiology; level 2, *knows how*, is the application of the knowledge in order to make decisions regarding a management plan; the third level, *shows how*, tries to look at how exactly the student would tackle a problem when faced with a patient; the final and fourth level, *does*, is meant to evaluate the learner in an actual clinical practice [13]. The four levels of assessment are also commonly known (and illustrated) as “Miller’s Pyramid” (see Fig. 6.2). Once again, specific methods of assessment evaluate the different levels of assessment with varying degrees of efficiency. Therefore, aligning an assessment method to a level on Miller’s Pyramid is another way of choosing between different assessment modalities (see Table 6.2).

Stages of Development

The process of learning and education in general is a continuum, a gradual progress that shows improvement within different competencies. Just like the different stages in training for a physician (medical school, internship, residency,

Table 6.1 Categories of assessment methods and specific examples

Assessment methods	Examples
Performance based	Simulation-based assessments
	Objective Structured Clinical Examinations (OSCEs)
Written and oral	Long and short cases
	Multiple-choice questions
	Oral exams
	True/false items
Clinical observation	Matching items
	Long and short essay questions
	360° evaluations/feedback
Miscellaneous	Direct observation of procedural skills
	Mini-clinical evaluation exercise
	Self-assessments
	Peer assessments
	Logbooks
	Patient surveys

Primary focus of the chapter is highlighted in green

fellowship, attending physician), an anesthesiology assistant (anesthesiology assistant students, certified anesthesiology assistants), and a certified registered nurse anesthetist (student registered nurse anesthetist, certified registered nurse anesthetist), subjects in different stages of development learn differently. There are multiple ways to describe the stages of development that a learner is advancing through, and different organizations in various parts of the world use unique models. One approach, the RIME scheme (Reporter, Interpreter, Manager, Educator), was originally designed for internal medicine residency in the United States, and later the

Dreyfus brothers created an alternative and original model for skill acquisition for learners in general (novice, competent, proficient, expert, master) that was later summarized by Michael Eraut into the widely accepted stages: novice, advanced beginner, competent, proficient, expert (Fig. 6.3) [14, 15, 16].

Context

As mentioned previously, the purpose of an assessment is of paramount importance. In the outcome-based educational model, formative feedback (or “assessment *for* learning”) is as important as the traditional summative reasons (or “assessment *of* learning”). Moreover, George Miller stated, “Tests of knowledge are surely important, but they are also incomplete tools in this appraisal if we really believe there is more to the practice of medicine than knowing” [13]. Furthermore, different stakeholders may make different decisions based on the same results from rating scales and scores. As such, assessment for its own sake should be avoided, and an a priori justification is crucial to guiding the process and informing the assessment itself.

Rating Instruments and Scoring

There are numerous rating scores and checklist forms validated for use, and they can often be used interchangeably with most methods of assessment. Examples of rating instruments include ANTS (Anesthetists’ Non-Technical Skills),

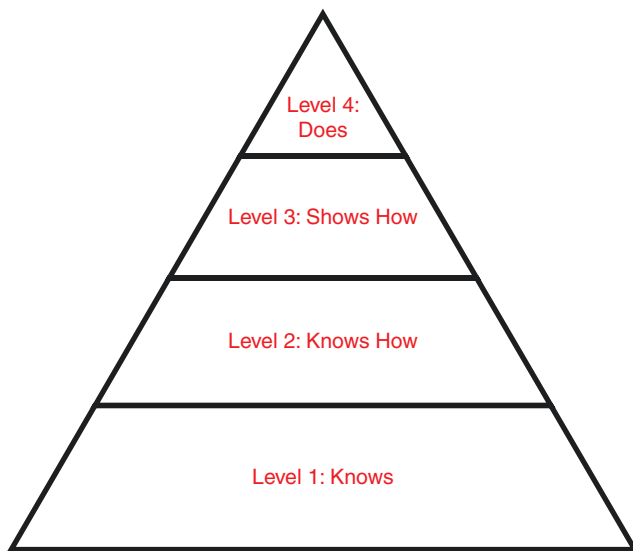


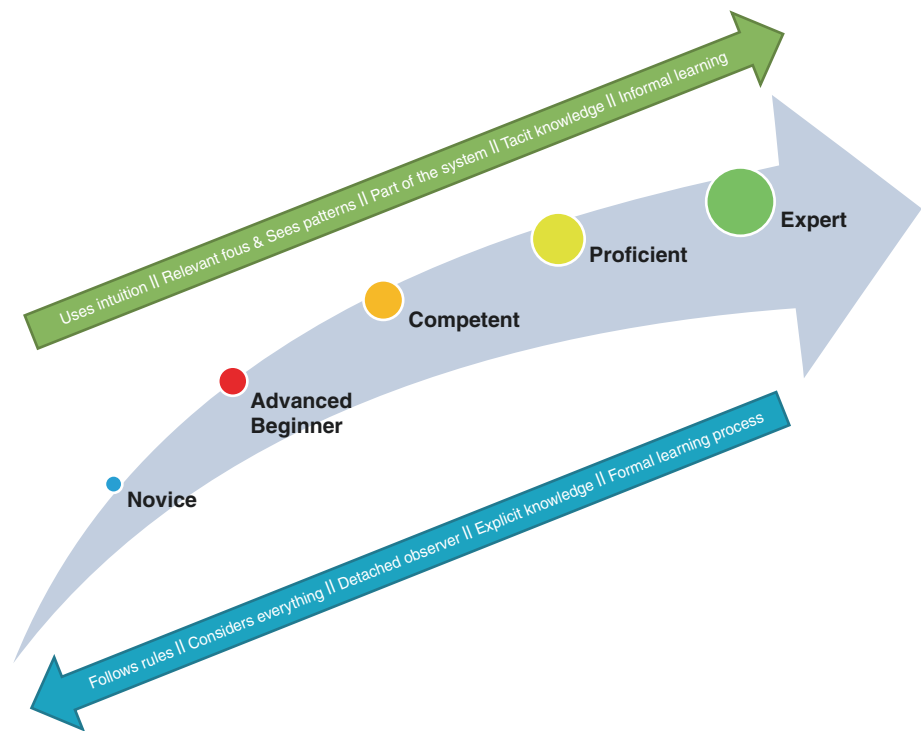
Fig. 6.2 Based on George Miller’s Pyramid for levels of assessment [13]

Table 6.2 Aligning the level of assessment with a matching assessment methodology

Level of assessment	Assessment methods	Examples
1 – Knows	Written and oral	Multiple-choice questions
		Oral exams
		True/false items
		Matching items
	Performance based	Simulation-based assessments
2 – Knows How	Written and oral	Oral exams
		Long and short essay questions
	Performance based	Simulation-based assessments
3 – Shows How	Performance based	Simulation-based assessments
		Objective structured clinical examinations
4 – Does	Clinical observation	Direct observation of procedural skills
		360° evaluations/feedback
		Mini-clinical evaluation exercise
	Miscellaneous	Medical record audits
		Peer assessments
		Self-assessments
		Logbooks

Highlighted in green, simulation-based assessments under the category of “Performance based” is primarily used to evaluate level 3 in the “Miller’s Pyramid” scheme—“shows how.” However, there is also value in using simulations to assess levels 1 and 2 (“Knows” and “Knows How,” respectively), as shown in yellow

Fig. 6.3 Model of skill acquisition (and the basis of action learners). (Based on the Dreyfus brothers' original model, later summarized by Michael Eraut [15, 16])



NOTSS (Non-Technical Skills for Surgeons), and RIME (Reporter, Interpreter, Manager, Educator), to name a few. Rating instruments undergo demanding procedures during development, and those developed are said to be “validated.” However, it is important to note that the context of an assessment is extremely important in the validation process and that adjustments must be made when using previously developed rating instruments (even if the outcomes of a competency being measured are similar) when used under different circumstances.

Competency Assessment of Practitioners Using Simulation

When assessing competence in a clinical setting, there is a layer of variance in reliability that cannot normally be controlled, and this is largely due to the factors introduced by the patient being studied (or the specific physiological changes present from certain diseases) or the clinical task at hand. Abrahamson and Barrows had the foresight to recognize these difficulties and created what is now recognized as the standardized patient (SP), one of the first true simulation modalities used for assessment (see also Chap. 10). For decades, SPs have been recognized as a means of delivering history and physical findings consistently and have proven key to simulation-based assessments [17]. In addition to SPs, Abrahamson created Sim One, the first computer-enhanced mannequin that was used to train anesthesiology residents in endotracheal intubation (see also Chaps. 1, 11, and 12). He

showed that residents achieved a higher level of proficiency with intubation in overall fewer days in training, as well as with fewer attempts within the operating room, remarking that this leads to increased patient safety [18]. Over the years, technological advances in computer-enhanced mannequins allow manipulation of multiple physiologic parameters so that anesthesiology trainees (as well as members of other specialties) can become acquainted with the scenarios most commonly encountered within the operating room, as well as introduced to the rare scenarios with potential for major morbidity and mortality, all while in a controlled environment. The ability to reproduce the same parameters, accurately, across multiple administrations of the assessment to different groups of subjects is a great strength of using simulators in assessments. However, simulation adds a different layer to the already complex process of competency assessment, through separate challenges regarding validity and fidelity of the simulated environment.

For an assessment being performed for a specific purpose, it is important to note that the individual properties of various assessment methods will determine in part which components of the validity argument are the weakest. As Kane states, “Validity evidence is most effective when it addresses the weakest parts of the interpretive argument ... The most questionable assumptions deserve the most attention” [19]. In regard to observational methods for conducting assessments, including the use of simulated environments, the *generalization* component of the validity argument is often questioned, due to construct underrepresentation, when it is believed that a single encounter

in the test world is not generalizable to the real world [20]. However, this threat can be overcome by increasing the volume of observations in one assessment. For example, increasing the number of different cases that the subject is exposed to and assessed on, a better prediction can be made between the test world and the real world regarding behavior and performance in a variety of clinical encounters [20].

Unlike using written tests as an assessment modality, where *extrapolation* is a major threat to the validity argument, simulations add strength to this component of the validity argument via realism of the scenarios. The simulated environment is one that attempts to reproduce an actual clinical environment that the subjects will encounter in practice, specifically the conditions that anesthesiology residents and attending residents experience in the perioperative care of their patients. *Fidelity* is the degree of exactness to which this simulated environment parallels real-world circumstances. Certain aspects that contribute to increased or high-fidelity simulations are as follows: (1) physical appearance of mannequins tailored to specific scenarios; (2) extra support staff playing the roles of surgeons, nurses, technicians in and around the operating room environment; (3) presence of appropriate equipment to simulate an operating room, such as anesthesia machine, operating table, surgical equipment, proper draping, IV poles, etc. In a simulation lab, subjects are expected to interact with the computer-enhanced mannequin and with the other support staff as they normally would if they were a real patient or real interdepartmental colleagues.

Several simulators have been developed and are used to aid in the learning of specific technical skills by a trainee in anesthesiology. Examples of these part-task trainers include head mannequins for direct laryngoscopy, virtual reality (VR) simulators for fiberoptic bronchoscopy, or surgical trainers. With each of these devices, a singular technical skill was shown to be more easily acquired when compared to groups that did not use the devices [18, 21]. However, the critical piece missing from these devices that are high in engineering fidelity is the interplay of psychological fidelity, which deals with the actual skills and behaviors required in real clinical situations [22]. The ability to create an environment rich in both engineering fidelity and psychological fidelity is invaluable as it allows the assessor to witness and grade nontechnical skills, such as communication, as well as technical skills in a fluid and busy environment.

Regardless of the level of anesthesiology practitioners, it is important to assess how their technical and nontechnical skills are affected or change in an environment filled with various distractors. Imagine an operating room environment where the surgeon is placing pressure to get the operation started and upon induction, the patient rapidly desaturates and all eyes are on the anesthesiology provider who will make attempts to secure the airway—one can simulate this and even more complex scenarios in the simulation lab. In

fact, it has been shown that having a simulated environment that triggers extreme emotional responses within the subjects can increase future performance [23].

Often the underlying assumption is that the closer the simulation is to the real-life environment, the better the assessment of performance will be in predicting clinical behaviors. However, as one can imagine a plastic mannequin, even the most advanced model, certainly has limitations. Furthermore, other aspects of the simulated environment can distract the subject, creating “simulation artifact” and can interfere with the assessment of performance. Therefore, creating too rigid of a scenario may not necessarily be optimal depending on how a participant interacts in the simulated environment, and “sticking to a script” may also negate a scenario’s validity with regard to generalization into real-world clinical practice (see also Chap. 3).

The Role of Simulations in Competency Assessment

Using the framework outlined above for assessment, we will now focus on specific simulations (including SPs, computer-enhanced mannequins, virtual reality simulators) and their current and evolving role in assessment in medical education, specifically as they relate to the specialty of anesthesiology. To be sure, individuals and practitioners in different stages of training learn differently, necessitating distinct assessment programs for each level, including undergraduate medical education, anesthesiology resident training, certification or continuing education of attending anesthesiologists, and even reentry of attending anesthesiologists. In addition, there are similar programs targeting anesthesia assistants, student registered nurse anesthetists, certified registered nurse anesthetists, and reentry of certified registered nurse anesthetists into practice.

Undergraduate Trainees in Anesthesiology

We have witnessed the national impact of simulations upon medical education with the implementation of Step 2 Clinical Skills (Step 2 CS) portion of the United States Medical Licensing Examination (USMLE). This simulation exam is for summative purposes and consists of multiple stations that assess students’ ability to perform a history and physical examination upon SPs, who provide highly reliable and accurate histories for specific disease processes, as well as objective physical exam findings [24]. SPs provide a score to each student using standardized checklists, and this provides input in a high-stake decision of whether the medical students will “pass” or “fail” and whether they may move on to the next phase of medical education in residency. Once

this Step 2 CS component of the USMLE was announced, medical schools around the United States began the development of programs with the training of SPs to simulate this very exam and to facilitate the teaching and assessment of the clinical skills, including history taking, performing physical examinations, and creating differential diagnoses.

As mentioned previously, nearly all medical schools now employ a program to help prepare the students for various clerkships and for the Step 2 CS exam by means of the SP. These same SPs can be used to help foster a preoperative history and physical examination to help prepare for a clerkship in anesthesiology. For centers that have head mannequins and part-task trainers, holding workshops for basic yet vital technical skills such as bag-mask ventilation and intravenous line placement will aide these students over various clerkships and residencies regardless of specialty. More advanced workshops requiring more sophisticated simulators for endotracheal intubation, advanced cardiac life support (ACLS), central line placement, neuraxial/regional anesthesia for these medical students can also promote interest in the field of anesthesiology. One program has developed a six-week externship for third-year medical students, which includes didactics and procedural and simulation education, where they had a statistically significant increase in applications for the field of anesthesiology by the program's completion [25]. As the medical students are progressing through the anesthesiology clerkship, the use of periodic high-fidelity simulation scenarios as a means of assessment can help gauge student interest, as well as student adherence to readings and to establish if they are meeting the various stated goals and objectives. In addition, varying proportions of medical students who pass through an anesthesiology clerkship and who have interest in applying for a residency inevitably speak with the program director in anesthesiology at the host institution. The summative and formative feedback from various simulation assessments, in addition to the more basic and general exam scores from USMLE Steps 1 and 2, can help program directors better direct the applicants.

In addition to the increased use of simulation in medical school and specifically the potential benefits of teaching by simulation during anesthesiology clerkships, schools for student registered nurse anesthetists (SRNAs) have also found value in incorporating simulation within the curriculum. The importance of nontechnical skills (NTS) on anesthetist performance and the resultant excellence in care and patient safety outcomes has been reviewed and validated [26, 27]. The previously mentioned ANTS rating score has been used during simulations to assess these very skills in new trainees and practicing anesthesiologists. However, there has been little incorporation of simulation into the SRNA curriculum in the United States. Among various international institutions, there has recently been incorporation of simulation into SRNA curriculum where they have also created

and validated modified rating instruments (NANTS-no and N-ANTS) suited for their specific assessments [28, 29]. In the United States, a program for SRNAs has moved toward the development and incorporation of an Objective Structured Clinical Examination (OSCE) for summative assessment via simulation for first-year SRNAs to ensure competence prior to entering their clinical year [30]. A follow-up study by Wunder was conducted to see the effect of a 3-hour intervention on first-year SRNAs on their NTS; results showed a statistically significant increase in post-test scores as measured by six high-fidelity scenarios simulating crisis [31].

Anesthesiology Residents

In respect of outcomes and the ACGME Toolbox, experts have placed simulations as a preferred or “most desirable” modality to evaluate clinical skills, knowledge, and attitudes in areas involving patient care and procedural proficiency, as well as interpersonal and communication skills. Similarly, when thinking about levels of assessment in “Miller’s Pyramid” model, simulations tend to emphasize the *shows how* level, in which anesthesiologists in training can show how they would perform a skill, whether it be technical or nontechnical.

Key to simulation-based assessment for anesthesiology residents is the ability to provide fidelity of a very complex environment with varying degrees of workload, time pressure, and nontechnical challenges. Most importantly, it adds to patient safety by having new residents in anesthesiology learn in, and periodically be assessed in, a controlled, high-fidelity environment.

Anesthesiology as a specialty assumes proficiency in numerous technical and nontechnical skills, in addition to a broad knowledge base regarding human physiology and pharmacology; these include airway management (whether basic bag-mask ventilation or direct laryngoscopy and fiberoptic bronchoscopy skills), ACLS, central-line placement, performing neuraxial blocks and peripheral nerve blocks, team leadership skills, interpersonal and communication skills, and overarching crisis resource management skills, to name a few essential components. It is important to assess these skills as many of them deal with emergency situations in which a potential outcome is the death of the patient being cared for if not performed properly and in a timely manner. Current restrictions on duty hours imposed by the ACGME for anesthesiology residents and having a minimum of eight consecutive hours off between shifts [32], along with a finite amount of operating room cases with technical and nontechnical skills needing to be deployed, are some limitations that exist when learning and practicing said tasks. Here again, we suggest the use of simulation for the training and assessment of anesthesiology practitioners to increase their efficiency, proficiency, and ultimately patient safety.

Several studies have utilized simulation-based assessment for anesthesiology residents, where the instruments used were to assess explicit procedural skills and those of communication and collaboration, with the majority of the simulation studies devoid of evidence to support the validity of the performance measures [33–39]. The authors of one paper created tested a behaviorally anchored rating scale used during a simulation-based assessment to help identify critical gaps in anesthesia performance and to increase patient safety [40]. Two trained faculty (blinded to outside anesthesiology trainee program and level of training) applied the behaviorally anchored scale in a multiscenario setting and used surveys completed by the residents, fellows, facilitators, and raters to gain feedback on the overall assessment system. The results showed evidence that supported the reliability and validity of the assessment scores that included high generalizability, and the feedback from the surveys illustrated that the multiscenario simulation-based assessment was “useful, realistic, and representative of critical skills required for safe practice” [40].

Many anesthesiology residency programs that have a simulation center or program in place are incorporating its use and have integrated it into the standard training curriculum. It is particularly important to give the Clinical Anesthesia Level 1 (CA-1) residents this exposure early on in their training so that any gaps in knowledge and skills identified from the formative assessment can be used to further tailor and stimulate growth within those residents. Some critics have stated that the outcomes from a simulation scenario may impact the scoring and validity as some learners may become emotionally invested in negative outcomes and potentially employ avoidance behaviors when similar situations arise in the future (whether in more simulated environments or clinical practice). However, in a particularly rare but critical case scenario of pipeline contamination of oxygen supply during a simulated intraoperative environment, Goldberg et al. demonstrated that a negative outcome to the patient during a simulated independent practice (one where the facilitator is hands off and is simply there to drive the scenario and watch it unfold) led to better retention of clinical skills upon retesting the scenario 6 months later as compared to those who performed in a simulated supervised practice (one where an attending intervened to “save” the patient) [41]. In addition, simulation-based assessments can be performed that are targeted to more specific and advanced areas within the specialty of anesthesiology, to continue the educational growth of the anesthesiology residents on specific rotations during residency.

Regardless of the stage of development of the learner, one can tailor a simulated environment to assess competencies, as well as constructive data in a formative assessment. For example, to assess specific skills in a novice anesthesiology resident, one might separate them into individual compo-

nents, such as preanesthetic evaluation, airway intubation, and hemodynamic management intraoperatively. A resident in the advanced or expert stage of development may get all the same components tested but in a more intricate environment that involves an acutely decompensating patient, which requires multitasking.

Practicing Anesthesiologists

Whether the assessment being performed is considered high stakes or low stakes, whether it is for summative (assessment *of* learning) or formative (assessment *for* learning) purposes, the results from one individual modality alone never solely determine the outcome. It is important to note George Miller’s words in that “...no single assessment method can provide all the data required for judgement of anything so complex as the delivery of professional services by a successful physician” [13]. Rather, we make a strong case to use simulation as another assessment modality, an extra tool to assist in gathering the evidence to lead to the appropriate decision.

To work as a trainee or faculty within the hospital setting, these practitioners (especially those within the department of anesthesiology) are required to have active Basic Life Support (BLS) and ACLS certification. Currently, all American Heart Association (AHA) courses for the certification of BLS and ACLS involve the use of simulations, including part-task training for evaluating chest compressions and airway management skills on mannequins, as well as high-fidelity scenarios for mega-codes. In fact, studies have demonstrated an increased retention of skills and knowledge of ACLS when using full-scale, high-fidelity environments as compared to the standard part-task mannequins of the past [42–45].

Another clear example of utilizing simulations for high-stake situations deals with the evolving nature of the certification process of anesthesiologists after the completion of residency. The traditional exam consisted of two parts: an advanced written exam portion and an oral exam portion. Now, physicians must take three different assessments of different modalities: a written exam and what is now known as the “APPLIED” exam, which consists of the traditional oral exam, as well as a third portion that follows the OSCE format. The OSCE started in March of 2018, and the goal with this addition has been to “assess two domains that may be difficult to evaluate in written or oral exams—communication and professionalism and technical skills related to patient care” [46]. The nontechnical skills regarding communication and professionalism include informed consent, discussing various treatment options, working through peri-procedural complications, navigating ethical issues, communication with other professionals, practice-based learning

and improvement; the technical skills being assessed include the interpretation of a variety of simulated monitors, interpretation of various views of echocardiography, and application of ultrasonography [47]. It seems that preparation for this high-stake summative exam, which now includes a larger portion of simulations via both simulators and SPs, will best be achieved by the increased, periodic use of simulations by individual institutions that cover the various aspects of the examination.

Maintenance of Certification

Nowadays, the initial certification upon passing the high-stake examinations and becoming a new anesthesiology attending is time limited. This means that anesthesiology practitioners must partake in periodic evaluations to demonstrate continued and up-to-date knowledge in the field in order to become recertified. This process is known as the Maintenance of Certification in Anesthesiology (MOCA). This recertification process currently consists of multiple components, and traditionally, part 4 of the MOCA consisted of a simulation course at a center endorsed by the American Society of Anesthesiologists (ASA). With the evolution of times, this last portion of the MOCA now includes a wide variety of activities, including being an institutional/departamental leader of a quality improvement project, clinical pathway development leader, or self-directed case discussion/presentation of Mortality & Morbidity (M&M) case, to name a few [48]. However, undergoing the simulation course will likely remain a popular choice for those seeking recertification because “simulation experiences stimulate active learning and motivate personal and collaborative practice improvement changes” [48]. Furthermore, the simulation course offers the most points per hour and is beneficial in terms of time commitment. A study looking at practice improvement plans over a period of 3 years after anesthesiologists participated in MOCA part IV simulation course demonstrated that 94% of these practitioners successfully applied some or all of their planned improvements in practice [49].

There exists a minimum of certain requirements for the MOCA courses imposed by the ASA, such as duration of the course, content, and ratio of faculty to attendees, but the specifics of the content and structure/organization of the MOCA courses are left up to the discretion of the endorsed simulation center. One course design is described by the Mount Sinai Human Emulation, Education and Evaluation for Patient Safety and Professional Study (HELPS) Center. They developed a course using a variety of educational formats, which included traditional lectures on topics such as airway management, small group activities on part-task training mannequins and virtual-reality simulators to break the

ice among the attendees, team-building exercises, and large “grand simulation” scenarios that ties everything together in high-fidelity simulations followed by debriefing [50].

Physician Reentry

In addition to simulations being integrated into the curriculum during anesthesiology residency, they are also being used for the purpose of retraining anesthesiology practitioners from returning to practice after a leave of absence. Relative physician shortages through a mismatch of supply and/or demand are a reality in modern medicine and can be attributed to a multitude of factors, both personal and economic [51, 52]. Although there is no standardized curriculum for reentry program and therefore allows flexibility for customization to the individual, a lot of the programs provide opportunities of observership that may assess the *knows* and *knows how* levels but not much more. The effectiveness of simulation-based assessment has been demonstrated for physician reentry across a variety of fields [53–56]. However, the use of a high-fidelity and high-stake simulation-based assessment “to assess the individual practitioner’s deficits and provide a means to tailor the educational program to fill skills and knowledge gaps without risk of patient harm” is a unique method and consists of a two-part process: (1) a two-day assessment involving two standardized written tests (Anesthesiology Knowledge Test and AHA ACLS), as well as several repetitive simulations of varying complexities and covering various concepts (common and rare but critical scenarios) that are scored using a global rating Likert scale 1 to 5; (2) retraining phase that consists daily simulations covering distinct learning objectives over a variable and flexible 1- to 6-week duration, as well as operating room observation with one simulation faculty member [1, 57]. A case series that was published by DeMaria et al. demonstrated the success of the program with 73% of participants having successfully reentered into active practice for at least 1 year [57].

Rating Instruments in Simulation-Based Assessment

We have discussed the general criteria for good assessment, briefly mentioned different modalities of assessment, and then focused on the use of simulation for competency assessment, with examples specific to the specialty of anesthesiology. The assessment is not complete without a composite score that can be used to make a decision about the learners. As stated earlier, there are a host of global rating scales and various checklists already present, and it is beyond the scope of this chapter to list and analyze each one but rather to give a comparison. There was a

systematic review performed of simulation-based assessments, in which the authors reviewed the evidence of the checklists and global rating scales used; the results from their study showed that the global-rating scale (GRS) to checklist correlation was 0.76, a similar interrater reliability between the two methods; and GRS had higher inter-item and interstation reliabilities than checklists [58]. So checklists are certainly a good alternative to GRS, but if a scenario has multiple tasks being measured, a separate checklist needs to be developed and used for each task, whereas the GRS can be used across several tasks.

Conclusion

Using simulation as an educational modality is gaining traction in the medical field, and their use in competency assessment is growing with it. Using the framework described as applied to simulation-based assessments, the purpose of the assessment to be performed must clearly be stated. In regard specific outcome to be assessed and the level of assessment to be evaluated, simulations are known to be the strongest in the clinical skills and interpersonal and communication skill areas and *shows how* level; however, the complexity of the scenario and questions asked within the scenario or in the debrief can elucidate and assess other aspects and levels of assessment. Simulation-based assessments (ranging from the use of SPs to computer-enhanced mannequins) are invaluable in the field of anesthesiology, where patient safety is addressed daily through invasive procedures, maintaining vigilance for the common and rare yet potentially fatal scenarios, along with knowledge and implementation of ACLS. This method of assessment of competencies in anesthesiology residents ensures little to no harm for patients while allowing for high reliability, as well as strengthening of the weaker components of the validity argument (generalization and extrapolation inferences) by manipulation of variables and organizing and conducting assessments with the appropriate level of fidelity. The use of simulation by the department of anesthesiology at individual teaching institutions as a method of teaching and assessing anesthesiology practitioners for both technical and nontechnical skills throughout the stages of training, and for new trainees and physicians seeking reentry into clinical practice, is all encouraged as the formative feedback will prove invaluable to both the assessor and those being assessed. Especially now with the movement by the American Board of Anesthesiology in incorporating simulations to provide summative feedback in the high-stake initial certification process, along with the ongoing MOCA requirements, anesthesiology practitioners should be familiarized with the simulated environments in all the components that will be tested.

References

1. Goldberg A, et al. High-stakes simulation-based assessment for retraining and returning physicians to practice. *Int Anesthesiol Clin.* 2015;53(4):70–80.
2. McGaghie WC, et al. A critical review of simulation-based medical education research: 2003–2009. *Med Educ.* 2010;44(1):50–63.
3. Issenberg SB, et al. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Med Teach.* 2005;27(1):10–28.
4. Weller JM, et al. Psychometric characteristics of simulation-based assessment in anaesthesia and accuracy of self-assessed scores. *Anaesthesia.* 2005;60(3):245–50.
5. Cook DA. When I say... validity. *Med Educ.* 2014;48(10):948–9.
6. Cook DA, et al. A contemporary approach to validity arguments: a practical guide to Kane's framework. *Med Educ.* 2015;49(6):560–75.
7. Kane MT. Validating the interpretations and uses of test scores. *J Educ Meas.* 2013;50(1):1–73.
8. Gingerich A, Regehr G, Eva KW. Rater-based assessments as social judgments: rethinking the etiology of rater errors. *Acad Med.* 2011;86(10 Suppl):S1–7.
9. Yeates P, et al. Seeing the same thing differently: mechanisms that contribute to assessor differences in directly-observed performance assessments. *Adv Health Sci Educ Theory Pract.* 2013;18(3):325–41.
10. Accreditation Council for Graduate Medical Education. Next Accreditation System (NAS) and Milestones – a resident perspective. 2014 [cited 2016 09/25]. Available from: <http://www.acgme.org/portals/0/pdfs/resident-services/9nasresidentmay2014.pdf>.
11. The Accreditation Council for Graduate Medical Education and The American Board of Anesthesiology. The Anesthesiology Milestone Project. 2015 [cited 2016 09/20]. Available from: <https://www.acgme.org/Portals/0/PDFs/Milestones/AnesthesiologyMilestones.pdf>.
12. Accreditation Council for Graduate Medical Education and The American Board of Medical Specialties. ACGME competencies: suggested best methods for evaluation (Toolbox of assessment methods). 2000 [cited 2016 09/26]. Available from: <https://www.partners.org/Assets/Documents/Graduate-Medical-Education/ToolTable.pdf>.
13. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med.* 1990;65(9 Suppl):S63–7.
14. Pangaro L. A new vocabulary and other innovations for improving descriptive in-training evaluations. *Acad Med.* 1999;74(11):1203–7.
15. Dreyfus, S.E. A five-stage model of the mental activities involved in directed skill acquisition. 1980 [cited 2016 09/28]. Monograph available online. Available from: <http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA084551&Location=U2&doc=GetTRDoc.pdf>.
16. Eraut M. Developing professional knowledge and competence. London: RoutledgeFalmer; 1994. p. 124.
17. Barrows HS, Abrahamson S. The programmed patient: a technique for appraising student performance in clinical neurology. *J Med Educ.* 1964;39:802–5.
18. Abrahamson S, Denson JS, Wolf RM. Effectiveness of a simulator in training anesthesiology residents. *J Med Educ.* 1969;44(6):515–9.
19. Kane MT. An argument-based approach to validity. *Psychol Bull.* 1992;112(3):527–35.
20. Downing SM, Haladyna TM. Validity and its threats. In: Downing SM, Yudkowsky R, editors. *Assessment in health professions education.* New York: Routledge; 2009. p. 21–56.
21. Goldmann K, Steinfeldt T. Acquisition of basic fiberoptic intubation skills with a virtual reality airway simulator. *J Clin Anesth.* 2006;18(3):173–8.
22. Maran NJ, Glavin RJ. Low- to high-fidelity simulation – a continuum of medical education? *Med Educ.* 2003;37(Suppl 1):22–8.

23. Demaria S Jr, et al. Adding emotional stressors to training in simulated cardiopulmonary arrest enhances participant performance. *Med Educ*. 2010;44(10):1006–15.
24. USMLE. Bulletin of Information. 2016 [cited 2016 09/30]. Available from: <http://www.usmle.org/pdfs/bulletin/2016bulletin.pdf>.
25. Baker KS, Cormican D, Seidman PA. Summer anesthesiology externship: demonstrating the ability of early clinical involvement to educate and increase specialty interest. *J Educ Perioper Med*. 2012;14(4):E063.
26. Fletcher GC, et al. The role of non-technical skills in anaesthesia: a review of current literature. *Br J Anaesth*. 2002;88(3):418–29.
27. Glavin RJ. Excellence in anesthesiology: the role of nontechnical skills. *Anesthesiology*. 2009;110(2):201–3.
28. Flynn FM, Sandaker K, Ballangrud R. Aiming for excellence – a simulation-based study on adapting and testing an instrument for developing non-technical skills in Norwegian student nurse anaesthetists. *Nurse Educ Pract*. 2016;22:37–46.
29. Lyk-Jensen HT, et al. Assessing nurse anaesthetists' non-technical skills in the operating room. *Acta Anaesthesiol Scand*. 2014;58(7):794–801.
30. Wunder LL, et al. Objective structured clinical examination as an educational initiative for summative simulation competency evaluation of first-year student registered nurse anesthetists' clinical skills. *AANA J*. 2014;82(6):419–25.
31. Wunder LL. Effect of a nontechnical skills intervention on first-year student registered nurse anesthetists' skills during crisis simulation. *AANA J*. 2016;84(1):46–51.
32. Accreditation Council for Graduate Medical Education. Specialty-specific Duty Hour Definitions. 2012 [cited 2016 09/29]. Available from: http://www.acgme.org/Portals/0/PDFs/DH_Definitions.pdf.
33. Mudumbai SC, et al. External validation of simulation-based assessments with other performance measures of third-year anesthesiology residents. *Simul Healthc*. 2012;7(2):73–80.
34. Berkenstadt H, et al. Deficits in the provision of cardiopulmonary resuscitation during simulated obstetric crises: results from the Israeli Board of Anesthesiologists. *Anesth Analg*. 2012;115(5):1122–6.
35. Ben-Menachem E, et al. Objective structured clinical examination-based assessment of regional anesthesia skills: the Israeli National Board Examination in anesthesiology experience. *Anesth Analg*. 2011;112(1):242–5.
36. Flin R, et al. Anaesthetists' non-technical skills. *Br J Anaesth*. 2010;105(1):38–44.
37. Fehr JJ, et al. Simulation-based assessment of pediatric anesthesia skills. *Anesthesiology*. 2011;115(6):1308–15.
38. Blum RH, et al. A method for measuring the effectiveness of simulation-based team training for improving communication skills. *Anesth Analg*. 2005;100(5):1375–80, table of contents
39. Malec JF, et al. The mayo high performance teamwork scale: reliability and validity for evaluating key crew resource management skills. *Simul Healthc*. 2007;2(1):4–10.
40. Blum RH, et al. Simulation-based assessment to identify critical gaps in safe anesthesia resident performance. *Anesthesiology*. 2014;120(1):129–41.
41. Goldberg A, et al. Learning through simulated independent practice leads to better future performance in a simulated crisis than learning through simulated supervised practice. *Br J Anaesth*. 2015;114(5):794–800.
42. Wayne DB, et al. A longitudinal study of internal medicine residents' retention of advanced cardiac life support skills. *Acad Med*. 2006;81(10 Suppl):S9–S12.
43. Wayne DB, et al. Mastery learning of advanced cardiac life support skills by internal medicine residents using simulation technology and deliberate practice. *J Gen Intern Med*. 2006;21(3):251–6.
44. Wayne DB, et al. Simulation-based education improves quality of care during cardiac arrest team responses at an academic teaching hospital: a case-control study. *Chest*. 2008;133(1):56–61.
45. Rodgers DL, Securro S Jr, Pauley RD. The effect of high-fidelity simulation on educational outcomes in an advanced cardiovascular life support course. *Simul Healthc*. 2009;4(4):200–6.
46. The American Board of Anesthesiology. The American Board of Anesthesiology: APPLIED (Staged Exams). 2016 [cited 2016 09/30]. Available from: [http://www.theaba.org/Exams/APPLIED-\(Staged-Exam\)/About-APPLIED-\(Staged-Exam\)](http://www.theaba.org/Exams/APPLIED-(Staged-Exam)/About-APPLIED-(Staged-Exam)).
47. The American Board of Anesthesiology. Applied Examination. [cited 2016 12/13/2016]. Available from: <http://www.theaba.org/PDFs/APPLIED-Exam/APPLIED-OSCE-ContentOutline>.
48. The American Board of Anesthesiology. MOCA Part 4 Requirements. [cited 2016]. Available from: <http://www.theaba.org/PDFs/MOCA/MOCA-2-0-Part-4-Requirements>.
49. Steadman RH, et al. Practice improvements based on participation in simulation for the maintenance of certification in anesthesiology program. *Anesthesiology*. 2015;122(5):1154–69.
50. Levine AI, et al. Simulation-based Maintenance of Certification in Anesthesiology (MOCA) course optimization: use of multi-modality educational activities. *J Clin Anesth*. 2012;24(1):68–74.
51. Grace ES, et al. Physicians reentering clinical practice: characteristics and clinical abilities. *J Contin Educ Heal Prof*. 2011;31(1):49–55.
52. Jewett EA, Brotherton SE, Ruch-Ross H. A national survey of 'inactive' physicians in the United States of America: enticements to reentry. *Hum Resour Health*. 2011;9:7.
53. Gallagher CJ, Tan JM. The current status of simulation in the maintenance of certification in anesthesia. *Int Anesthesiol Clin*. 2010;48(3):83–99.
54. DeMaria S Jr, Levine AI, Bryson EO. The use of multi-modality simulation in the retraining of the physician for medical licensure. *J Clin Anesth*. 2010;22(4):294–9.
55. Levine AI, et al. Role of simulation in U.S. physician licensure and certification. *Mt Sinai J Med*. 2012;79(1):140–53.
56. Steadman RH, Huang YM. Simulation for quality assurance in training, credentialing and maintenance of certification. *Best Pract Res Clin Anaesthesiol*. 2012;26(1):3–15.
57. DeMaria S Jr, et al. Simulation-based assessment and retraining for the anesthesiologist seeking reentry to clinical practice: a case series. *Anesthesiology*. 2013;119(1):206–17.
58. Ilgen JS, et al. A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment. *Med Educ*. 2015;49(2):161–73.