

Designing and Implementing the Objective Structured Clinical Examination in Anesthesiology

Maya Jalbout Hastie, M.D., Jessica L. Spellman, M.D., Parwane P. Pagano, M.D., Jonathan Hastie, M.D., Brian J. Egan, M.D., M.P.H.

ABSTRACT

Since its description in 1974, the Objective Structured Clinical Examination (OSCE) has gained popularity as an objective assessment tool of medical students, residents, and trainees. With the development of the anesthesiology residents' milestones and the preparation for the Next Accreditation System, there is an increased interest in OSCE as an evaluation tool of the six core competencies and the corresponding milestones proposed by the Accreditation Council for Graduate Medical Education. In this article the authors review the history of OSCE and its current application in medical education and in different medical and surgical specialties. They also review the use of OSCE by anesthesiology programs and certification boards in the United States and internationally. In addition, they discuss the psychometrics of test design and implementation with emphasis on reliability and validity measures as they relate to OSCE. (*ANESTHESIOLOGY* 2014; 120:196-203)

SINCE its introduction in the late 1970s, the Objective Structured Clinical Examination (OSCE) has been used as an assessment tool in medical education, physician training, and certification exams. There has been an increased interest in OSCE in recent years, accompanying planned changes in residents' education, evaluation, and certification.

The Accreditation Council for Graduate Medical Education (ACGME) defined in 1999 six core competencies that outlined the scope of medical residents' development and described the required domains within which to assess all residents. These have become well established and incorporated into graduate medical education. More recently, the ACGME elaborated incremental concrete steps, called milestones, by which to measure progress within these domains.¹ The milestones are a set of specialty-specific educational outcomes to be achieved at defined intervals. In the Next Accreditation System, annual residency program review will include an assessment of residents' progress along the milestones.¹ The ACGME will implement the Next Accreditation System in anesthesiology training by June 2014.

The anesthesiology milestones will delineate five developmental levels, each level characterized by the additional knowledge, skill, or behavior required for a physician at that stage. The assessment of residents' progress along the milestones may incorporate written and oral examination, observation in the clinical setting, OSCE, or any combination

thereof. In addition, the American Board of Anesthesiology plans to introduce OSCE into the final part of the applied examination, including use of standardized patients, mannequins, or computer-based assessment. Accordingly, anesthesiology training programs and residents must understand the design, applications, and limitations of OSCE to prepare for the upcoming changes in program accreditation and practitioner certification.

What Is an OSCE?

Miller² proposed a hierarchical framework for assessing physicians long before the milestones were introduced. "Knowing" is the most fundamental form of understanding, and factual knowledge is amenable to assessment by written tests. When a physician combines knowledge with clinical judgment, this "knowing how" incorporates data to make an informed decision about patient management. The anesthesiology oral exam format aims to assess this competence. "Showing how" is what residents will do in practical situations, and this lends itself to assessment by direct clinical observation. Clinical observation, however, may be lacking in scope and in objectivity. OSCE and simulation scenarios can provide valuable additional assessments of this level of knowing, which includes clinical judgment and "practical skills."³ Ultimately, residents should be able to translate this

This article is featured in "This Month in Anesthesiology," page 1A. Corresponding article on page 4.

Submitted for publication April 30, 2013. Accepted for publication October 9, 2013. From the Department of Anesthesiology, Columbia University Medical Center, New York, New York.

Copyright © 2013, the American Society of Anesthesiologists, Inc. Lippincott Williams & Wilkins. *Anesthesiology* 2014; 120:196-203

knowledge into the clinical setting and demonstrate their skills and knowledge with actual patients (“doing”). Assessment of this level of behavior remains the most challenging to accomplish reliably and accurately.²

Formalized by Harden *et al.*⁴ in 1975, the OSCE consists of a series of stations, each 5 to 10 min in duration. These stations provide an array of clinical scenarios and tasks that require procedural skills or data interpretation. The presence of two or more examiners using a standardized checklist promotes objectivity and consistency in the assessment of the trainee in clinical problem solving and patient-management skills.^{2,5-7} OSCE incorporates several methods of assessment such as multiple-choice questions, open-ended questions, simulation as well as the classical description of standardized patients.^{2,5,6} The hallmark of OSCE is the focus on assessment of clinical competence,⁵ or in Miller’s classification, the ability of trainees to demonstrate their knowledge in practice. By ensuring both the uniformity of the administered exam and the exposure to different examiners, the OSCE provides a more comprehensive evaluation of the trainee.⁵ Further, the OSCE design promotes objectivity by using multiple examiners, establishing evaluation criteria, and incorporating reproducible scoring sheets.⁸ The use of standardized patients further promotes uniformity in the delivery of the exam as well as testing of rare clinical scenarios.⁴ Finally, the OSCE design provides for immediate feedback to both the resident and the educator after completion of the stations,^{8,9} thus reinforcing the learning⁹ and identifying areas of weaknesses in the training or in the exam itself, thereby making it a valuable formative evaluation tool.⁸

OSCE in Medical Education

Because of these useful characteristics, the OSCE has been incorporated into several certifying international agencies’ certification processes since 1994. The Clinical Skills Examination portion of the United States Medical Licensing Examination is a model of OSCE that has been used for the assessment of foreign medical graduates since the late 1990s,¹⁰ and subsequently for the assessment of American medical students since 2004.¹¹ In addition, most U.S. medical schools have included OSCE in their curricula.¹² Although most reports of OSCE have described its use in medical student education, OSCE has been applied in graduate medical education either to complement information obtained from existing evaluation methods or to circumvent the subjective nature of other forms of assessment.¹³ Several reports have evaluated the feasibility, reliability, and validity of using OSCE in the medical and surgical fields.¹⁴⁻²⁰ In reported experiences with OSCE, various formats have included a combination of methods including written problem-solving stations,¹⁵ standardized patients,^{14-16,20} hands-on stations requiring demonstration of technical skills, or interpretation of laboratory results.²⁰ These differences in

format and content target various objectives expected of diverse groups of trainees.

OSCE was found to be a reliable tool to assess physical examination skills,^{18,19} clinical judgment and diagnosis,^{15,20} interpretation of radiological and laboratory findings,²⁰ technical skills,¹⁷⁻²⁰ and even billing and documentation exercises.¹⁶ Communication skills, such as end-of-life decisions, disclosure of medical errors,^{14,21} and patient feedback,¹⁶ have also been assessed.

OSCE in Anesthesiology

Relatively few reports exist on the use of OSCE in anesthesiology training or certification. Anesthesiology OSCE has been used for assessing physical exam skills, clinical and history-taking skills, airway management,^{3,22} resuscitation,^{3,23} blood product transfusion,²⁴ anatomy, as well as statistics.³

OSCE was included in the final examination of the Royal College of Anaesthetists in the United Kingdom in the mid-1990s.³ The exam consists of 17 stations used to assess a range of skills such as resuscitation, handling and troubleshooting anesthesia equipment, data interpretation, history taking and communication, physical examination, identification of anatomy, and understanding and using statistical tools.³ It was subsequently incorporated into the primary part of this two-part certification exam.²⁵

The Israeli National Board Examination in Anesthesiology first included OSCE in 2003. The exam format and process were created in a joint effort by the Israeli Board of Anesthesiology Examination Committee, the Israel Center for Medical Simulation, and the Israeli National Institute for Testing and Evaluation.²² The examination content was designed using the approach described by Newble⁶ in which OSCE developers (1) identify the clinical competencies to be assessed, (2) design the tasks to evaluate the competency, and (3) create a blueprint for the test to be administered.⁶ In the Israeli experience, expert opinion determined the clinical skills to be assessed. The examination task force selected representative tasks, which were subsequently incorporated into five simulation-based OSCE stations: regional anesthesia, trauma management, resuscitation, intensive care medicine/ventilation, and critical events in the operating room.²⁶ The passing threshold required completion of previously determined critical actions in addition to successful demonstration of 70% of the checklist items.²² This high pass rate was observed despite the use of a predetermined passing threshold, which was reported to result in lower pass rates when compared with expert consensus methods, such as Angoff method or borderline method.²⁷

The Israeli Board of Anesthesiology later included a regional anesthesia station. The development of this station was identified as a highly iterative process, requiring revisiting for optimal design and delivery.²⁶ The authors also recognize the need for consultation with experts for psychometric rigor,²⁶ as will be discussed later in this review.

In 2010 the Royal College of Physicians and Surgeons of Canada incorporated simulation assistance with video presentations to its oral examination in anesthesiology.²⁸ The test format and content continue to evolve, underscoring the importance of revisiting the original OSCE design throughout the process.

Psychometrics of OSCE

Psychometrics of any assessment tool can be evaluated along four measures: feasibility, objectivity, reliability, and validity.^{12,29} The majority of the reported experience with OSCE in medical education has focused on only one or a few aspects of psychometrics, without a systematic approach to reliability and validity. Many reports have highlighted the development and application of the examination form itself and underemphasized the importance of evaluation of the examination. Achieving reliability and content validity of OSCE is of particular importance when the OSCE is used as a measure of trainees' progress in residency or as part of the board-certification process.^{2,6} However, OSCE design and feasibility need to be taken into consideration as well.

Feasibility

Instructional systems designs traditionally follow a multi-stage, iterative model frequently referred to by the acronym ADDIE: Assess, Develop, Design, Implement, and Evaluate.^{30,31} The process is illustrated in figure 1.

Assessing Needs

Planning for any new educational program should start with needs assessment of the learner, the organization, and other regulating agencies. Hence, when planning an OSCE, residency training programs should take into consideration their residents' needs, their departmental and organizational goals, the ACGME design for curriculum, and the American Board of Anesthesiology design for board certification. It is tempting to assume that the needs of all stakeholders converge on preparing residents for the planned final step of their board certification. However, program-specific needs should be assessed before designing OSCE, such as whether the OSCE will be part of the formative and summative evaluations of the trainees and whether the results will affect a resident's progression through residency. Needs and interests can be identified both by forming advisory groups³¹ of program directors, key faculty involved in education, and resident representatives, as well as by surveying the experience of other programs and other specialties. In addition, expected changes in credentialing, such as the inclusion of OSCE in the American Board of Anesthesiology certifying exam, may prompt the development of additional means of evaluation.

Developing a Program

Goals and objectives of the OSCE are developed to address the competencies and milestones identified during the

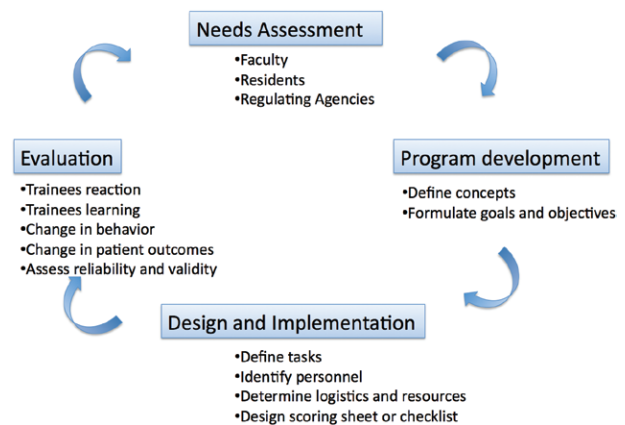


Fig. 1. The process of Objective Structured Clinical Examination design begins with needs assessment of the program, its faculty, the involved trainees, and the requirements from regulating agencies. During program development phase, key concepts to be evaluated are identified and specific goals and objectives are formulated. In the design and implementation phase, specific tasks and their corresponding scoring sheets are designed, and location, personnel, equipment, finances, and duration of the Objective Structured Clinical Examination stations are decided. Throughout the process, continuous evaluation of the program is performed and the program altered as needed.

needs-assessment phase. On the basis of the identified needs, the instructional program's goals and objectives are formulated explicitly and clearly, and are shared with the program, the learners, and the faculty. Objectives are specific and detailed explanations of the stated goals, and are usually described using Bloom's taxonomy.³² Bloom's taxonomy, originally published in 1956,³² and later revised and refined,^{33,34} allowed for a common language to be used in education and for assessment of educational endeavors.^{32,33} In the taxonomy of education, cognitive processes are viewed as a linear progression from least to most complex and are defined by the use of "verbs" to illustrate the category. Learners progress from knowing to "understanding" the concepts and their relationships, "applying" the learning, "analyzing" the principles and their organization, "synthesizing" the information and producing a plan of action, and finally "evaluating" the learning and the situation.^{32,33} Objectives therefore describe the skills, knowledge, and attitudes that will be assessed by the OSCE as well as their level of complexity, depending on the trainee level. This is particularly relevant in adequately defining tasks to match the anticipated milestones. Setting clear goals and objectives is important for the design of the learning activity, and for the evaluation of its progress. These should be revisited frequently to avoid inflexibility in the design, to incorporate other previously omitted goals and objectives, and to redefine those that are not relevant.³¹

Design and Implementation

In the design phase, specific tasks are elaborated to accurately assess the stated goals and objectives, and a plan for

implementation is put in place. The process is refined in this phase by defining the tasks to be included, the faculty involved in testing, and the logistics of the OSCE implementation. Identifying the skills, knowledge, or attitudes to be evaluated by the OSCE is key because the format of OSCE stations needs to be tailored to the task being assessed.² It has been suggested that OSCE is best suited to assess clinical and practical skills, rather than attitudes or factual knowledge,⁶ but several reports describe its use in assessment of communication skills as well.^{13,14,16} Tasks are then constructed to assess the given competencies, keeping in mind that performance on one task may not reflect performance on other tasks, even those that are closely related.⁶ A number of tasks should therefore be planned for each broad competency in order to ensure validity of the test.⁶ In designing the activity, methods for standardization of OSCE should be sought such as the use of an objective checklist for evaluation of participants, training of the examiners to ensure interrater reliability, as well as establishing the policies and procedures for administration of the exam. The logistics of OSCE include deciding on duration and format of each station, as well as the duration of the entire examination. Details such as location of testing and training of faculty involved in the OSCE should be addressed. Last, space, time, and cost are important considerations in OSCE design. Design and implementation are financially costly, especially when standardized patients are employed.^{12,29} Adhering to a timeline for design and implementation will avoid delays and frustrations and also allows for timely evaluation of the activity.³¹

Evaluation

Evaluation of the instructional design is an ongoing process throughout implementation and design, as well as after completion of the activity. Kirkpatrick proposed a four-step evaluation model, with each step providing increased complexity: reactions, learning, behavior, and results.³⁵ The first level describes the participants' attitudes, satisfaction, and emotional response to the learning activity. This can be assessed by surveying participating residents to evaluate their subjective response to the exam. Studies have shown mixed responses to OSCE regarding the level of stress experienced by trainees, their overall enjoyment of the activity, and their perception of the content validity of the test.^{7,8,12} The second level of the evaluation model is a measure of change in the learning of residents, demonstrated either by better performance on retesting or by improved performance on other measures of assessment such as in-training exams, oral exams, and others. Correlation between use of OSCE and improved performance would also serve as a validity measure of the designed test. However, establishing such correlations has been historically difficult.³ At the third level of evaluation, a change in behavior is sought in clinical practice, as evaluated by faculty. Finally, the ultimate outcome one hopes to establish is that the instructional design will lead

to improved patient care; however, this is both difficult to define and to measure.

Constant reflection and evaluation needs to accompany the process all throughout, leading to repeated reappraisal of the program, its purpose, and its design. Input from learners, the evaluators, and the program should be incorporated.³¹ In addition, although OSCE is used primarily as an assessment tool, the exam can uncover areas of weaknesses in the curriculum, which could subsequently be addressed by the program and the trainee.¹³

Reliability and Objectivity

Evaluation of OSCE reliability and validity has been investigated in several reports since the 1990s. Reliability is often referred to as the consistency of a test, that is, the reproducibility of the exam score.²⁹ Validity, however, is conceptualized as the accuracy of the exam score, or the extent to which the test measures what it purports to measure.²⁹ These simplistic definitions, however, hide the greater complexity inherent in test assessment.

Test reliability comprises several components: interrater reliability, internal reliability, test–retest reliability, and intermethod reliability. *Interrater reliability* is a measure of the degree of agreement between different raters when grading the same examinee at a specific station. It has been suggested that interrater reliability can be improved by using a standardized scoring checklist with the objectives clearly stated.^{2,4,36} Indeed, this is one of the reported strengths of the OSCE methodology. However, controversy persists over whether the checklist methodology of Harden performs as well as global ratings on measures of reliability.²⁹ Checklist scoring systems fail to acknowledge the ability of a trainee to solve a given clinical problem by implicit pattern recognition rather than by a step-wise approach.¹² In addition, standardization of the scoring is challenging and can lead to either a high pass rate or a high failure rate, depending on the method used for deciding on acceptable performance.²⁷ Reaching expert consensus is considered a better standardized approach than an arbitrarily chosen cutoff, because performance evaluation assesses the learner's engagement with the task rather than simply comparing them with peers.⁸

There are two main approaches to standardization: item-based (criterion-referenced) and trainee-based (norm-referenced). In item-based standardization, such as the Angoff method, a panel of experts decides how a borderline candidate is likely to perform on any given task,²⁷ and this is used as a guide to evaluate the actual performance of trainees. Trainee-based methods evaluate the overall performance of the trainees rather than focusing on specific tasks; the mean of all borderline scores achieved by trainees on a task is considered the passing score for the given station.⁶ Although Wass *et al.*³⁷ found norm-referenced scoring to be superior to criterion-referenced scoring, this finding has not been consistent. Finally, although interrater reliability is an important

goal, care must be taken when considering the test “reliable” based solely on a high degree of correlation in the scores between different evaluators.

The *internal reliability* of the exam is at least as important yet more difficult to achieve.³⁶ Internal reliability is characterized by the extent to which performance across different test stations remains consistent. In the context of a typical OSCE, high internal reliability implies that the scores obtained on various items are capturing the knowledge, skills, and attitudes from the same conceptual domain, or closely interrelated domains.³⁸ Although the reliability can be influenced by many factors including motivation, stress, attention, and distraction,^{12,29} the overarching concern is that the exam score conveys the degree of mastery of the intended competency. Internal reliability can be improved by increasing the number of stations^{12,29,36} or by focusing the content of the different stations on assessing the same conceptual domain or competence.^{8,39} Some evidence suggests that a series of OSCEs administered over time, when evaluated collectively, demonstrate improved reliability.¹²

Quantitative assessment of internal reliability is often reported as Cronbach alpha. This statistic can be conceptually understood as the average correlation score obtained after examining all possible divisions of the test items into two groups.^{38,40} The magnitude of this score from 0 to 1 is a reflection of the internal consistency of the exam scores. Values greater than 0.8 are desirable, especially in high-stakes situation such as in pass–fail decisions.¹⁵ However, one must also be aware that the resulting number does not reflect an inherent property of the exam itself,³⁶ but rather of the scores obtained on that exam. Accordingly, it is a reflection of the specific population of examinees who took the exam. If a test is piloted in a group of postgraduate year 1 trainees (or junior attendings) and then subsequently used for a group of postgraduate year 4 trainees, reliability may change. In addition, a trainee’s performance on one task is not a good predictor of his or her subsequent performance on other tasks.³⁶

Finally, *test–retest reliability* is an estimate of the trainee’s likelihood of achieving a similar score on a given station on repeat performance.³⁹ Performance on a given station is marginally improved by increasing time allowed for a task, but is significantly improved when feedback is provided.⁷ Achieving reliability is the necessary first step in establishing the validity of an examination.⁴⁰

Validity

One cannot assume that because a test meets acceptable criteria for reliability, it is valid.⁴⁰ Validity, in its simplest form, asks whether the exam provides accurate feedback about an examinee’s skill level in the domain or “construct” of interest. This may be assessed by examining several subcategories of validity. *Face validity* is, simply stated, the extent to which the exam has the appearance of measuring what it is intended to measure.²⁹ The assessment of face validity can be

done without the in-depth examination of exam content or the input of content experts and is, therefore, the least rigorously evaluated form of validity.⁴¹ *Content validity*, however, requires that subject-matter experts review (or design) exam items to ensure they reflect not only the topic or domain of interest but that they also adequately capture the entirety of the subject matter in that domain.²⁹ In order to draw conclusions about the level of expertise in the desired knowledge, skills, or attitudes, the exam items have to be adequately representative of the full spectrum of those elements within that domain. The goal of content validity in OSCE construction may be achieved with the use of a “blueprint,” which carefully deconstructs the competency to be tested in the design phase.⁶ As might be expected in light of this complexity, it has been suggested that an OSCE with fewer than 10 stations is less likely to achieve content validity.^{6,8} In addition, overall competence of the trainee, as related to level of training or to overall expertise, can influence performance on a specific station,⁴² and an OSCE with high construct validity may differentiate between the learners’ level of training.^{8,15,19}

Validity can also be measured by the relationship between exam performance and measures external to the exam.⁴¹ *Concurrent validity* refers to the correlation between performance on OSCE for a given competence and contemporaneous performance on an alternative, well-established method of evaluation of the same competence.²⁹ *Predictive validity* provides a similar measure of comparison, though it looks to correlate exam scores with an alternative assessment performed at a future time. Examining performance correlations on OSCE to other test modalities can help avoid the problem of case specificity, where the tasks tested are specific and fail to represent the general competency to be assessed.⁶

However, when measures of concurrent validity have been reported in the literature, results have varied between 0.1 and 1, with a majority of studies reporting a correlation factor of less than 0.7.¹² This limited concurrent validity may be related to the type of competency being assessed.^{12,29} Correlation coefficients, and hence concurrent validity, can be improved when specific subsets of an OSCE exam reflecting specific conceptual subsets of a competency are better matched to the comparison exam of interest. For example, Matsell *et al.*⁷ showed that OSCE had a high concurrent validity in pediatric residents in areas of knowledge and patient management compared with standardized multiple-choice tests; but OSCE did not correlate with other evaluation methods, such as faculty evaluation, in tasks assessing clinical skills and problem-solving ability. For this reason, some researchers have recommended that OSCE be used as an additional evaluation tool rather than an alternative to current assessment methods of clinical competencies.^{6,29,36}

When considering the use of OSCE in the assessment of residents’ progression along the ACGME milestones, as well as performance on the American Board of Anesthesiology certification exam, there is an underlying assumption that an OSCE provides superior predictive validity. Moreover, it is

Table 1. OSCE Applications in Anesthesiology

Concept	Station	Tasks
Cardiac physiology	ECG and arrhythmias	<ul style="list-style-type: none"> • Identify ECG rhythm strips • Provide differential diagnosis • Select appropriate medical therapy • Provide CPR
	Pressure-volume loops	<ul style="list-style-type: none"> • Identify pressure volume loop components • Draw pressure volume loops in different valvular heart diseases • Discuss effect of different vasoactive medications on the pressure-volume loop
Assist devices	Intraaortic balloon pump	<ul style="list-style-type: none"> • Interpret blood pressure tracing with intraaortic balloon pump • Explain indications and contraindications of intraaortic balloon pumps • Communicate and demonstrate troubleshooting of intraaortic balloon pumps
Hemodynamics and calculations	Intracardiac pressures	<ul style="list-style-type: none"> • Perform calculations for SVR, PVR • Identify waveforms of CVP, PA tracings
Cardiac anatomy	Coronary artery anatomy and distribution	<ul style="list-style-type: none"> • Correlate ECG changes with coronary distribution • Name left ventricular wall segments on simulated TEE
Pulmonary physiology	Airway management during one lung ventilation	<ul style="list-style-type: none"> • Understand indications and contraindications for lung isolation • Place double-lumen tube; confirm placement with auscultation and with fiberoptic scope • Analyze causes for hypoxemia during one-lung ventilation • Interpret arterial blood gases
Regional anesthesia	Epidural anesthesia management	<ul style="list-style-type: none"> • Obtain informed consent from standardized patient • Identify anatomy for epidural catheter placement
Airway management	Difficult airway management	<ul style="list-style-type: none"> • Formulate plan with unexpectedly difficult airway • Manage difficult airway • Demonstrate the use of different intubating devices

Important concepts to be evaluated are identified. Stations are designed to assess specific aspects of the desired concept. Tasks are specified according to the developed goals and objectives.

CPR = cardiopulmonary resuscitation; CVP = central venous pressure; ECG = electrocardiogram; OSCE = Objective Structured Clinical Examination; PA = pulmonary artery; PVR = pulmonary vascular resistance; SVR = systemic vascular resistance; TEE = transesophageal echocardiography.

this superior prediction of clinical performance that appears to be subtly implied by the recent enthusiasm for OSCE. But as OSCE is introduced into multiple training programs, one must ensure adequate internal reliability. Strong content validity should be achieved through rigorous planning and blueprinting. The true magnitude of predictive validity of OSCE with regard to postgraduate clinical performance remains to be seen (and its future assessment should not be forgotten). Designing these exams by anesthesiology programs will be a challenging endeavor and, necessarily, a highly iterative process, requiring frequent reevaluation of exam psychometrics. For this reason, Bromley *et al.*³ as well

as others have recommended that economies of scale in exam development should be employed, both to minimize costs of development as well as to maximize quality, sharing well-designed and tested stations between institutions.³⁶

Applications

In practical terms, OSCE may comprise as few as five stations and as many as twenty stations, depending on the specialty, with each station requiring 5 to 10 min for completion. This allows simultaneous administration to a larger group of trainees in a limited time frame. In addition, different formats can

Table 2. Blueprint Design

Stations	Patient Care	Medical Knowledge	Professionalism	Interpersonal and Communication Skills	Practice-based Learning and Improvement	Systems-based Learning
Difficult airway management	+	+		+		+
Arterial blood gases management	+	+				
Epidural anesthesia management	+	+	+	+	+	
Intraaortic balloon pump management	+	+				+

Example of the use of blueprint in Objective Structured Clinical Examination design relating specific stations to the competencies they assess. Similarly, a blueprint can be used to relate tasks to specific concepts or to milestones.

Table 3. Strengths and Weaknesses

Strengths	Weaknesses
Prepares residents for ABA examination	Can be costly
Helps evaluate trainees	Can be time consuming
Can assess milestones	Need for establishing validity
Correlates with other testing modalities	Difficulty in evaluating OSCE (quality, reliability, validity)

Some of the identified strengths and potential weaknesses of OSCE.

ABA = American Board of Anesthesiology; OSCE = Objective Structured Clinical Examination.

be combined within the OSCE stations, such as the use of standardized patients, laboratory data, equipment or slides, as previously described by Newble.⁶ In the Department of Anesthesiology at Columbia University Medical Center, we have started a limited use of OSCE for assessment of clinical skills, technical skills, and patient-management skills. Clinical skills of interest to anesthesiology programs may include taking medical history, obtaining informed consent for invasive procedures, and interpreting electrocardiograms, radiographic studies, and hemodynamic data. Technical skills can be assessed using simulation for airway management, double-lumen endotracheal tube placement, and basic echocardiography image acquisition and interpretation. Accordingly, some stations such as interpretation of laboratory or electrocardiogram studies could be completed without the presence of an examiner, and the results can be collected in a written format.¹³ However, technical stations, such as demonstrating the placement of a double-lumen tube or management of a difficult airway, require the presence of examiners. An objective assessment of the trainee is made by completing a predetermined checklist to evaluate for completion of all required steps. Agarwal *et al.*¹³ suggested further division of the stations into basic and advanced skills. Such a refinement may facilitate the assessment of progress along the developmental milestones according to trainee level. A sample of specific examples of the potential use of OSCE is provided in table 1, detailing the suggested tasks for assessing the described concepts. Table 2 illustrates the use of a blueprint for designing an OSCE in anesthesiology to assess core competencies.

Conclusion

As George Miller notes “no single assessment method can provide all the data required for judgment of anything so complex as the delivery of professional services by a successful physician.”² When well designed, OSCE is a reliable tool with only “modest” validity⁸ and should be accordingly viewed as a valuable, although insufficient addition to residency assessment. OSCE allows for a flexible, yet structured examination characterized by objective evaluation of trainees,⁵ preparing them for the board examination, as well as providing programs with means of regular trainee and program assessment. However, programs should not rely solely

on OSCE to provide by itself a comprehensive assessment of a trainee’s competence,^{3,12} but rather view it as complementary to existing exam modalities. In addition, the cost and the logistics of designing and implementing the OSCE should be considered carefully. Future studies should look at the use of OSCE both as a formative and summative assessment tool to evaluate its effect on the learning and behavioral outcomes of trainees, and compare it with other established methods of assessment. Some of the recognized strengths and potential weaknesses of the OSCE are listed in table 3. As more programs begin designing and implementing OSCE to accompany the changing accreditation system, experience with OSCE in anesthesiology education may be shared between programs and in the literature to further enrich the education community, to fill the knowledge gap about the applications of OSCE in anesthesiology, and to better prepare the trainees and residency programs for the ACGME competencies and milestones.

Acknowledgments

Support was provided solely from institutional and/or departmental sources.

Competing Interests

The authors declare no competing interests.

Correspondence

Address correspondence to Dr. M. J. Hastie: Department of Anesthesiology, Columbia University Medical Center, 622 West 168th Street, PH 5-505, New York, New York 10032. mj2081@cumc.columbia.edu. This article may be accessed for personal use at no charge through the Journal Web site, www.anesthesiology.org.

References

- Nasca TJ, Philibert I, Brigham T, Flynn TC: The next GME accreditation system—Rationale and benefits. *N Engl J Med* 2012; 366:1051–6
- Miller GE: The assessment of clinical skills/competence/performance. *Acad Med* 1990; 65(9 suppl):S63–7
- Bromley LM: The Objective Structured Clinical Exam—Practical aspects. *Curr Opin Anaesthesiol* 2000; 13:675–8
- Harden RM, Stevenson M, Downie WW, Wilson GM: Assessment of clinical competence using objective structured examination. *Br Med J* 1975; 1:447–51
- Harden RM: What is an OSCE? *Med Teach* 1988; 10:19–22
- Newble D: Techniques for measuring clinical competence: Objective structured clinical examinations. *Med Educ* 2004; 38:199–3
- Matsell DG, Wolfish NM, Hsu E: Reliability and validity of the objective structured clinical examination in paediatrics. *Med Educ* 1991; 25:293–9
- Carraccio C, Englander R: The objective structured clinical examination: A step in the direction of competency-based evaluation. *Arch Pediatr Adolesc Med* 2000; 154:736–41
- Hodder RV, Rivington RN, Calcutt LE, Hart IR: The effectiveness of immediate feedback during the objective structured clinical examination. *Med Educ* 1989; 23:184–8

10. Whelan GP: Educational Commission for Foreign Medical Graduates: Clinical skills assessment prototype. *Med Teach* 1999; 21:156–60
11. Papadakis MA: The Step 2 clinical-skills examination. *N Engl J Med* 2004; 350:1703–5
12. Turner JL, Dankoski ME: Objective structured clinical exams: A critical review. *Fam Med* 2008; 40:574–8
13. Agarwal A, Batra B, Sood A, Ramakantan R, Bhargava SK, Chidambaranathan N, Indrajit I: Objective structured clinical examination in radiology. *Indian J Radiol Imaging* 2010; 20:83–8
14. Chipman JG, Beilman GJ, Schmitz CC, Seatter SC: Development and pilot testing of an OSCE for difficult conversations in surgical intensive care. *J Surg Educ* 2007; 64:79–87
15. Cohen R, Reznick RK, Taylor BR, Provan J, Rothman A: Reliability and validity of the objective structured clinical examination in assessing surgical residents. *Am J Surg* 1990; 160:302–5
16. Franzese CB: Pilot study of an Objective Structured Clinical Examination (“the Six Pack”) for evaluating clinical competencies. *Otolaryngol Head Neck Surg* 2008; 138:143–8
17. Maker VK, Bonne S: Novel hybrid objective structured assessment of technical skills/objective structured clinical examinations in comprehensive perioperative breast care: A three-year analysis of outcomes. *J Surg Educ* 2009; 66:344–51
18. Merrick HW, Nowacek GA, Boyer J, Padgett B, Francis P, Gohara SF, Staren ED: Ability of the objective structured clinical examination to differentiate surgical residents, medical students, and physician assistant students. *J Surg Res* 2002; 106:319–22
19. Sloan DA, Donnelly MB, Schwartz RW, Strodel WE: The Objective Structured Clinical Examination. The new gold standard for evaluating postgraduate clinical performance. *Ann Surg* 1995; 222:735–42
20. Stewart CM, Masood H, Pandian V, Laeeq K, Akst L, Francis HW, Bhatti NI: Development and pilot testing of an objective structured clinical examination (OSCE) on hoarseness. *Laryngoscope* 2010; 120:2177–82
21. Hodges B, Turnbull J, Cohen R, Bienenstock A, Norman G: Evaluating communication skills in the OSCE format: Reliability and generalizability. *Med Educ* 1996; 30:38–43
22. Berkenstadt H, Ziv A, Gafni N, Sidi A: Incorporating simulation-based objective structured clinical examination into the Israeli National Board Examination in Anesthesiology. *Anesth Analg* 2006; 102:853–8
23. Berkenstadt H, Ben-Menachem E, Dach R, Ezri T, Ziv A, Rubin O, Keidan I: Deficits in the provision of cardiopulmonary resuscitation during simulated obstetric crises: Results from the Israeli Board of Anesthesiologists. *Anesth Analg* 2012; 115:1122–6
24. Corrie K, Wiles M, Flack J, Lamb J: A novel objective structured clinical examination for the assessment of transfusion practice in anaesthesia. *Clin Teach* 2011; 8:97–100
25. McIndoe AK: Modern anaesthesia training: Is it good enough? *Br J Anaesth* 2012; 109:16–20
26. Ben-Menachem E, Ezri T, Ziv A, Sidi A, Brill S, Berkenstadt H: Objective Structured Clinical Examination-based assessment of regional anesthesia skills: The Israeli National Board Examination in Anesthesiology experience. *Anesth Analg* 2011; 112:242–5
27. Kaufman DM, Mann KV, Muijtjens AM, van der Vleuten CP: A comparison of standard-setting procedures for an OSCE in undergraduate medical education. *Acad Med* 2000; 75:267–71
28. Blew P, Muir JG, Naik VN: [The evolving Royal College examination in anesthesiology]. *Can J Anaesth* 2010; 57:804–10
29. Barman A: Critiques on the objective structured clinical examination. *Ann Acad Med Singapore* 2005; 34:478–82
30. Allen WC: Overview and evolution of the ADDIE training system. *ADHR* 2006; 8:430–41
31. Galbraith MW, Sisco B, Guglielmino LM: Administering Successful Programs for Adults: Promoting Excellence in Adult, Community, and Continuing Education, Original edition. Malabar, Krieger, 1997, pp vii-187
32. Bloom BS: Taxonomy of Educational Objectives: The Classification of Educational Goals, 1st edition. New York, Longmans, Green, 1956
33. Krathwohl DR: A revision of Bloom’s taxonomy: An overview. *Theory Pract* 2002; 41:212–8
34. Anderson IW, Krathwohl DR: A Taxonomy for Learning, Teaching, and Assessing : A Revision of Bloom’s Taxonomy of Educational Objectives, Complete edition. New York, Longman, 2001, pp 25–92
35. Alliger GM, Janak EA: Kirkpatrick’s levels of training criteria: Thirty years later. *Pers Psychol* 1989; 42:331–42
36. Newble DI, Swanson DB: Psychometric characteristics of the objective structured clinical examination. *Med Educ* 1988; 22:325–34
37. Wass V, Van der Vleuten C, Shatzer J, Jones R: Assessment of clinical competence. *Lancet* 2001; 357:945–9
38. Cortina JM: What is coefficient alpha? An examination of theory and applications. *J Appl Psych* 1993; 78:98–4
39. Roberts J, Norman G: Reliability and learning from the objective structured clinical examination. *Med Educ* 1990; 24:219–23
40. Tavakol M, Dennick R: Making sense of Cronbach’s alpha. *Int J Med Educ* 2011; 2:53–5
41. DeVon HA, Block ME, Moyle-Wright P, Ernst DM, Hayden SJ, Lazzara DJ, Savoy SM, Kostas-Polston E: A psychometric toolbox for testing validity and reliability. *J Nurs Scholarsh* 2007; 39:155–64
42. Warf BC, Donnelly MB, Schwartz RW, Sloan DA: The relative contributions of interpersonal and specific clinical skills to the perception of global clinical competence. *J Surg Res* 1999; 86:17–23