Using OSCE-Based Evaluation: Curricular Impact over Time


Abstract: The Objective Structured Clinical Examination (OSCE) is becoming more widely used for performance assessment in dentistry. The department of pediatric dentistry at Baylor College of Dentistry (BCD) began incorporating the OSCE into its curriculum in 1995. This article describes the evolution of the department’s use of the OSCE and its impact on teaching and the curriculum. The discussion focuses on logistics and station design, curricular content and order, student anxiety, writing and scoring exams, and curriculum assessment. BCD has found that using an OSCE-based testing format is time-consuming and labor-intensive, but provides unprecedented feedback about students’ understanding and pinpoints areas of confusion. The demands of an OSCE-based testing format reveal that students can master, to the level of competency, only a finite amount of information in a given time period. The timed, interactive aspects of the OSCE create high levels of student anxiety that must be addressed. Writing and scoring OSCE items are different from traditional test items. The OSCE is a valuable mechanism to assess the students’ progress toward competency. This review of the process of incorporating OSCEs into a curriculum is the foundation for future assessment of the OSCE and its use for curricular improvement.

Dr. Zartman is Assistant Professor, Dr. McWhorter is Associate Professor, and Dr. Seale is Regents Professor and Chairman, Department of Pediatric Dentistry, Baylor College of Dentistry, Texas A&M University System Health Science Center; Dr. Boone is Associate Professor of Science and Environmental Education and Coordinator of the Program in Science and Environmental Education, Indiana University. Direct correspondence to Dr. Rosemarie Zartman, P.O. Box 660677, Department of Pediatric Dentistry, Baylor College of Dentistry, Texas A&M University System Health Science Center, Dallas, TX 75266-0677; 214-828-8456 phone; 214-828-8132 fax; rzartman@tambcd.edu.

Key words: OSCE, assessment of competency, Rasch analysis, curriculum

Submitted for publication 8/22/02; accepted 10/14/02

The Objective Structured Clinical Examination (OSCE) requires that examinees rotate through a series of stations in which they are required to perform a variety of clinical tasks. It has been used successfully in medicine since Harden and Gleeson introduced it in 1975 as a procedure to assess clinical competence at the bedside using timed stations. Harden and Gleeson described two types of stations: procedure stations and question stations. Procedure stations involving history-taking often used simulated patients as an alternative to actual patients. Review of the medical literature reveals that a variety of different formats have been used in administering the OSCE, and much of the literature focuses on its use in specialty programs.

Performance assessments such as the OSCE are only now becoming more widely used in dentistry. In 1994, the National Dental Examining Board of Canada (NDEB) began to include an OSCE as part of the certification process. OSCEs have also been used for summative in-course assessments at the Royal London School of Medicine and Dentistry and to provide feedback in restorative dentistry midway through the fourth year at Leeds Dental Institute in the United Kingdom. Several publications have focused on the use of standardized patients (SPs) in dental education. Johnson and Kopp described the use of an SP-based format for comprehensive treatment planning and emergency cases. Stilwell and Reisine used a patient-instructor (PI) program to teach and give individual feedback to a class of third-year dental students. Using carefully trained lay people to consistently portray a particular patient (standardized patient) and to assess students’ performance is a viable method for teaching and evaluating clinical interviewing skills. It can also be used to provide formative feedback throughout a predoctoral curriculum; however, there is nothing in the dental literature supporting this use.

The Department of Pediatric Dentistry at Baylor College of Dentistry (BCD) incorporated OSCEs into the curriculum in 1995. Results of what
we have learned by using the OSCE have been the subject of four Faculty Development Workshops presented at Annual Sessions of the American Dental Education Association since 1997. The purpose of this paper is to describe the evolution of the use of an OSCE-based testing format and the impact it has had on our program.

Logistics and Station Design

When we initially began to use the OSCE format, the logistics of developing and administering the examination were somewhat daunting and consumed all our energy and efforts. We soon discovered that three crucial factors must be considered at the beginning of the process: 1) the size of the class being examined, 2) the amount of space available, and 3) the time allowed for administering the exam. All other logistical considerations can be manipulated to accommodate these fixed variables.

Timed interactive stations are the hallmark of the OSCE format. Station types reported in the literature vary from short, simple, static exhibits to lengthy sophisticated stations with a standardized patient. At BCD, we have identified three categories of stations: 1) interactive, nonproctored stations where the student examines materials and provides a written answer to the question; 2) observed stations where the student performs a task in front of a passive observer who judges his or her performance; and 3) proctored, interactive stations where the student interacts with a proctor trained to act as a standardized patient or parent and also serve as a judge of the student’s performance. The two station types using proctors require trained personnel who must be present for the entire exam. Some proctors must be dentists, while others can be nondentists. We found that recruitment from outside our discipline is possible because the proctors are trained and standardized prior to the exam. Ultimatey, the size and makeup of the workforce available to participate in the examination determines numbers and types of proctored stations that can be used.

Once the space available, time allotted to the exam, class size, and available workforce have been determined, the numbers, lengths, and types of stations can be finalized. For example, at BCD, for a midterm examination of third-year students, we have a space allotment of two 100-station laboratories and one lecture room, an exam time of two hours, a class size of eighty-seven, and an available workforce of twenty composed of full- and part-time faculty, graduate students, department staff, and individuals recruited from outside the department.

For exam security, we examine all students at one time. The class is randomly divided in half, and during the first hour, one-half (forty-three) of the students take a written exam (not described further in this paper) while the other half (forty-four) take the OSCE. The groups are reversed during the second hour. Therefore, the total time allocated for the administration of the OSCE is sixty minutes. This constricted time frame is dictated by curricular scheduling outside of our control; though challenging, it is adequate for a meaningful examination experience, as this is one of three OSCEs administered during the term.

To rotate forty-four students through the OSCE in sixty minutes, the total must be divided into smaller, more manageable groups. Allowing five minutes for students to find their places and receive instructions and ten minutes to transfer them at the end of the exam, we have approximately forty-five minutes to actually administer the examination. Since we use four-minute stations, there can be eleven stations in forty-five minutes. The result is four groups of eleven students each. All four groups take the OSCE simultaneously; therefore, all stations must be duplicated four times, and four different proctors must be standardized for each proctored station. In this scenario, given a workforce of twenty, there could be as many as five proctored stations. All proctored stations have a performance checklist of designated tasks that assist the proctor in judging the student’s performance. Nonproctored stations require written answers that must be graded after the exam is completed. Keys for these stations are prepared in advance to facilitate standardization in grading.

The space allocated for the four OSCEs is in the two laboratories, and two groups of eleven students each are assigned to each laboratory. Due to the close proximity of OSCE stations, each has a three-sided, foam-board barricade for privacy. Figure 1 depicts a rotation chart for stations for two identical OSCEs being administered simultaneously in one laboratory.

With time and experience, the logistics of developing and administering the OSCEs consumed less energy, and we began to focus on the valuable feedback provided through observing the students and grading their responses.
The need for curricular modification in pediatric dentistry at BCD became evident within our first year of experience with the OSCE-based testing format. The interactive nature of the OSCE, which requires students to use written and verbal communication as well as problem-solving skills, provided us with unprecedented feedback about their understanding and mastery of concepts. Published reports on the use of standardized patients in dental OSCE stations have also identified this improved level of feedback concerning areas of student misunderstanding as a finding associated with the format.1,8

The fact that OSCE questions are open-ended eliminates the advantage of recognition or guessing possible when taking multiple choice tests; thus, the students’ answers reflect their core comprehension. Our most overwhelming and alarming discovery was how confused our students were over what we believed to be the most basic concepts. As we searched for the source of their confusion, we quickly realized our curriculum was overloaded with nonessential information. Concomitant with our initial OSCE discoveries was the movement in dental education to competency-based teaching. Our experiences with the OSCE forced us to evaluate and streamline our curriculum using the competency-based concept of “necessary to know” versus “nice to know.”

Achievement of competency requires mastery of basic information necessary for a beginning practitioner. Testing for mastery of information involves assessment of higher-order critical thinking and problem-solving skills that require in-depth command of concepts. Our experiences with OSCEs revealed multiple areas of confusion that required reteaching and retesting. In order to produce a student who gave evidence of mastery, concepts from the elaboration theory were employed.9 All important fundamental concepts were introduced early in the second-year course, so they could be built upon as the student progressed. A traditional preclinical technique course consisting of twelve hours of lecture and thirty-six hours of laboratory was converted to forty-eight one-
hour modules taught to small groups of five to six students. Each module was hands-on and interactive in format, and all of the basic pediatric dentistry concepts were introduced during this course. The goal was to produce a “safe beginner” who could enter the clinic in the third year. Traditional lectures were replaced in the third and fourth years with small-group, interactive seminars that used multiple approaches to reinforce and build on basic concepts.

The OSCE format is now used throughout the pediatric dentistry curriculum to identify areas of misunderstanding and confusion so that reteaching and retesting can take place. A total of seven OSCEs are given throughout the pediatric dentistry didactic curriculum, and focus groups are routinely held to assess students’ attitudes and concerns about the exam. Time is made available to review exams and provide feedback to students about their performance.

**Student Anxiety Concerning the OSCE**

Changes in curriculum format can result in anxiety for students as well as for those involved in implementing the changes. When the faculty does not pay attention to feedback and subtle cues from students, it can sometimes be overlooked that students are experiencing their own stressors. During the first administrations of the OSCE, there were multiple anecdotal reports from station proctors that some students had shaking hands and quivering voices at the proctored, interactive stations. There was even one incident where a student experienced such a heightened level of anxiety that she could not complete the exam.

These observed behaviors led the faculty to include focus group sessions concerning the OSCE with each group of students in the second-year pediatric dentistry course. Each group of six students met for fifty minutes with a facilitator who was an educational specialist employed by the college and who did not participate in the course or exam. The facilitator followed a list of open-ended questions about the course and the OSCE to elicit student comments. By using small groups of students with a facilitator not involved in the course or exam, our intent was to encourage frank, uninhibited feedback. The facilitator asked the students: Was the OSCE straightforward? Did you have enough time at each station? How can the OSCE be better described so you will know how to study? A majority of the student groups (nine out of fourteen) said that they needed more time at some of the stations. Students expressed concern that the format was still unfamiliar. Some reported “feeling panic” when the classmate in front of them accidentally walked away with materials from that station or if a classmate did not leave the station quickly enough. A moderate number of student groups (six out of fourteen) felt that some of the questions were vague. Other comments described frustration with the quality of some of the materials used at stations; this frustration caused them to lose focus at subsequent stations, thereby increasing their anxiety. It was apparent that what we might interpret to be small details could be sources of stress that could have a snowball effect on students. In addition to focus group feedback, the most common anecdotal feedback to faculty was that students reported feeling very threatened by the observed nature of the proctored OSCE stations and the required roleplay at interactive stations.

Using the focus group feedback, the faculty created practice OSCE modules to be included within the course for the purpose of lowering stress among students. Practice OSCEs were used within a lesson module during the course and in a stand-alone module that closely simulated the exams given during the course. For the practice OSCE module, a classroom was set up to simulate (on a smaller scale) the administration of an actual test. Five stations were set up with the same types of materials used in a real OSCE, including props such as dental models and radiographs, answer sheets for students to record their answers, foam board barricades, and arrows attached to the tables to help direct students from station to station. The faculty coordinating the practice OSCE gave directions to the students and operated a timer to cue them to move through all five stations. This practice OSCE did not count for a grade, but after the exercise, students reviewed correct answers with the faculty leader. The faculty member also used this time to answer students’ questions and give advice on how to study for these exams. The focus was on the OSCE format and logistics rather than on reviewing course information. The outcome was a student who had more practice with this new format and was more familiar with the types of questions that would appear. With better-prepared and more relaxed students, when the actual exams were graded, we obtained a better sense of students’ understanding because reduction of anxiety levels had improved their performance.
The grading of the OSCE was also a source of anxiety for students. Students expressed concerns that some proctors graded harder than others. Students indicated that some of the questions were too difficult to interpret and/or answer within the four-minute time limit of the stations. These issues required the faculty to look at ways to improve the standardization of the proctors and the grading process. Our standardization of proctors involves writing and using scripts for proctors who talk to students so that each student technically is being asked the same question. When there were multiple proctors for the same station, these individuals met to practice scenarios of students and proctors interacting at that station. This helped refine the proctor’s response during the interactions. We realize that we have more work to do, but what we have learned has affected our test construction as well as what we do with the test answers we collect from students. The test answer data reveal valuable information when analyzed.

**Scoring Exams/Writing Exams**

The performance data obtained from students at OSCE stations include written answers, proctor or observer evaluations of performance of a task, and student materials left at stations for grading after the exam is over. The resulting data are very different from traditional test data that can often be immediately scored electronically. The OSCEs are scored by either the proctor or by a faculty member who scores all of the students’ answers for a given nonproctored station. Having different scorers for an examination can create the need for employing checks on standardization of the scoring process. For proctored and observed stations scored during the exam, it is essential to create forms called “proctor sheets” (checklists) that include all the tasks the student should demonstrate to receive credit for performing the prescribed procedure. This causes the faculty to break down a given clinical skill or treatment into all the pieces or subskills that make up the total action.

For a dental student treating a pediatric patient, for example, giving local anesthesia includes the sub-skills of knowing the correct tooth to anesthetize, using behavioral management techniques, knowing where to place the needle, etc. (Figure 2). A variety of skills—including communication, motor, and cognitive—are all involved in the same station. For all three types of stations, faculty must determine what the expected answer will be and then break the answer down into its smallest subskills or items before the exam is given. This process allows the creation of reliable proctor sheets, and for the answers scored after the exam, it facilitates creation of a grading key that helps maintain consistency among faculty when grading the entire group. When the exam is scored as a distribution of items, then a judge (proctor or faculty) only needs to decide if the student performed the item correctly, partially correctly, or incorrectly. Sometimes students provide answers that are partially correct, but are not the expected responses on our exam key. These instances have created a need to scrutinize each question before it becomes part of the exam. Several faculty members thus read a proposed question and make revisions to the question. This process of review and revision of the question continues until all are satisfied that it should not be misinterpreted.

The creation of exam questions, key, and proctor sheets takes place over several weeks. Since our items are scored as correct, partially correct, or incorrect, we assign each a specific point value. For example, an item may be worth a total of four points. If a student answers the item totally correctly, he or she receives four points toward the total score. If the faculty judge determines the answer is partially correct, the student earns two points. If the student answered the item incorrectly, he or she receives no points for that item. Students’ total raw scores are tabulated and then converted to a traditional grade on a letter scale. One advantage of this method is that difficult items that test higher order skills or items inherent to the skill can be assigned higher point values than the easier items. This is known as “weighting” of the exam questions.

The student receives feedback in the form of a grade. Faculty receive feedback in the form of student responses to items on a clinically based, performance-based examination. Student responses are converted into data spreadsheets relatively easily because they are in the form of items that fall into one of three categories: correct, partially correct, or incorrect. Each category is designated by a number 2, 1, or 0 in the spreadsheet for each item. With the help of a measurement expert, the data can be analyzed. Rasch analysis allows us to produce maps that demonstrate how the tested group of students has performed. We have coded items that appear on this type of map to better describe them. For instance,
PLACE NAME LABEL   Proctor Initials ________

QUESTION 7

Preinjection

7a. Tell-Show-Do (T-S-D) general explanation before putting tooth to sleep  Y  N

Anesthesia Technique

Topical Application

7b. T-S-D/age-appropriate language  Y  N
7c. Dry tissue  Y  N
7d. Apply small amount of topical to site of needle penetration  Y  N
7e. Cover topical with gauze  Y  N
7f. Wait 1 min before injecting  Y  N

Injection

7g. T-S-D/age-appropriate language  Y  N
7h. Distracts/constantly talks/jiggles lip  Y  N
7i. Anesthetizes the buccal area of Tooth E  Y  N
7j. Anesthetizes the lingual area of Tooth E  Y  N
7k. Uses interpapillary injection for lingual  Y  N

Post Injection

7l. T-S-D/age-appropriate language describes sensation  Y  N
7m. Does not say "shot," "hurt," "pain," etc.  Y  N

COMMENTS:

Figure 2. Proctor sheet for OSCE station at which the student demonstrates local anesthesia administration
an item can be described by the content area (domain) that we are testing, such as the “restorative” domain. The item can be described by the skill it tests, such as a communication skill. The item can also be described by the faculty’s assessment of the difficulty of the item. The maps more clearly show item difficulty that is not readily apparent through the simple calculation of percentages. This information paired with the item descriptors (domain, skill, faculty assessment of difficulty) has become the backbone of our process toward exam and course improvement.

Curriculum Assessment

Earlier we discussed how we used OSCE data concerning student performance to manage our curriculum by reducing its scope and revising its sequencing. The most recent development from our use of OSCE-based testing has been the realization that data obtained have broader application than merely grading student performance. As we have become more skilled at designing examinations to facilitate data analysis, we have come to the realization that these analyses can be used to assess the overall effectiveness of the curriculum. By definition, curricular assessment should be used to determine whether the curriculum is doing what it is supposed to do: produce a competent beginning practitioner. Most information in the dental literature concerning curricular assessment is directed either at whole school curriculum in preparation for accreditation site visits or at individual courses based on student evaluations.11,12 There is a dearth of information about how to assess, over time, the effectiveness of a discipline-based curriculum to produce the desired result.

At BCD, a number of approaches to curriculum assessment have been developed using a Rasch analysis of OSCE data from our students’ performance throughout their three years in the pediatric dentistry curriculum. These have led to the development of purposeful assessment techniques (PAT), which were discussed in detail in an earlier paper.13 Briefly, PAT are those techniques that help provide valid and reliable measures that are informative to students, faculty, and administrators and allow one to work toward development of linear measurement scales that are “person-free” and “item-free.” These are being used to assess the effectiveness of our pediatric dentistry curriculum.

Summary

Published reports concerning use of OSCE-based testing formats, both in the medical and dental literature, indicate that OSCEs have many formats and uses. This paper described the effects an OSCE-based testing format has had on our predoctoral curriculum in pediatric dentistry over a seven-year period. The following statements summarize the lessons learned:

1. The use of an OSCE-based testing format is time-consuming and labor-intensive, requires extensive resources, and provides unprecedented feedback about students’ understanding and areas of confusion concerning basic concepts.

2. The demands of an OSCE-based testing format reveal that students can master, to the level of competency, only a finite amount of information in a given time period. Therefore, competency-based education requires a streamlined curriculum.

3. The timed, interactive aspects of the OSCE create high levels of student anxiety that can impact performance among students unfamiliar with the format. For students’ performance to be reflective of their actual knowledge, strategies must be implemented to desensitize students to the OSCE testing format.

4. Writing and scoring OSCE items are different from traditional test items and require standardization of judges and graders, weighting of items, and mixing items by skill set, difficulty, and domain being tested.

5. The OSCE is a valuable mechanism to assess curriculum, not only for the content but also its effectiveness. A Rasch analysis of OSCE-based testing data provides a means to develop and use tools for this purpose.

REFERENCES

