The Use of an Objective Structured Clinical Examination With Postgraduate Residents in Pediatrics

Robert I. Hilliard, MD, EdD, FRCPC; Susan E. Tallett, MBBS, FRCPC

Objective: To evaluate the usefulness of an objective structured clinical examination (OSCE) for assessing and providing feedback to postgraduate residents in pediatrics.

Design: A 5-station OSCE given in 1996, based on the educational objectives of a general pediatric training program. Each station assessed the residents’ interviewing and history-taking skills with a standardized patient. The results were correlated with those of the in-training evaluation reports.

Setting: The Department of Paediatrics, University of Toronto Faculty of Medicine, Toronto, Ontario.

Participants: Forty-three of 61 pediatric residents: 14 first-year, 12 second-year, 8 third-year, and 9 fourth-year residents.

Main Outcome Measures: Scores for each of the 5 stations were broken down into 15 points for the checklist, 5 for the global assessment, and 10 for the written postencounter question, for a total score of 150. The in-training evaluation report ratings were converted to a 5-point numerical scale, ranging from 1 (unsatisfactory) to 5 (outstanding).

Results: The mean OSCE score for the 43 pediatric residents was 104.9. Although the residents in their senior year scored higher, there was no statistically significant difference among the 4 years for the total OSCE score or for any of the 5 stations. The fourth-year residents’ scores on the postencounter questions were significantly (P<.05) higher than the first-year residents’ scores. Two residents scored less than 60%. The internal consistency of the 5-station OSCE was limited (r=0.69). Residents received verbal feedback at the conclusion of the OSCE, and they received their scores when they were calculated. The mean overall in-training evaluation report score for all 61 pediatric residents was 3.9. There was a moderate, but statistically significant, correlation between the overall mean OSCE results and the overall mean in-training evaluation report scores (r=0.45).

Conclusion: The OSCE can provide a useful formative evaluation of postgraduate residents, but the usefulness of the evaluation data and the feedback must be balanced with the logistic difficulties and expense.

Arch Pediatr Adolesc Med. 1998;152:74-78
SUBJECTS AND METHODS

In 1996, the Postgraduate Medical Education Committee of the Department of Paediatrics, University of Toronto Faculty of Medicine, agreed that all pediatric residents should participate in an OSCE, if possible, given the limitations of their clinical schedules, and that the results of the examination would be included in each resident’s evaluation file. The OSCE was administered during a weekday (March 29, 1996) to all of the 61 general pediatric residents who were able to take part in the examination.

THE OSCE

The couplet station format was used for the OSCE: a clinical encounter followed by a postencounter probe (PEP) station at which students answered open-ended or multiple-choice questions based on the clinical scenario just completed. At each station, the residents’ performance was assessed with a standardized checklist and an examiner’s global rating (the examiners included one of us [S.E.T.]).

The 5 stations developed for the OSCE were based on important educational objectives for the general pediatric residency program. Each station was given an equal weighting; each checklist was scored out of 15 points, each global assessment out of 5 points, and each PEP out of 10 points. These 5 problems were presented: (1) a father concerned about his 12-year-old son’s short stature; (2) a mother concerned about her 8-year-old daughter, who had hematuria; (3) a 16-year-old girl with secondary amenorrhea; (4) a 13-year-old boy with abdominal pain; and (5) a teenage mother concerned about her 8-year-old daughter, who had hematuria;

Twenty-three faculty members acted as examiners (including one of us [S.E.T.]). To accommodate all the residents, 2 equivalent circuits of 5 stations were used simultaneously, and each circuit was repeated twice. Fifteen empty rooms were required. There were 4 equivalent standardized patients and parents were trained with the help of the Standardized Patient Program of the Department of Family and Community Medicine, University of Toronto Faculty of Medicine.

ADMINISTRATION OF THE OSCE

The reliability of the OSCE was calculated with a coefficient of internal consistency (Cronbach’s α).

The content validity of the OSCE was assessed by a comparison of the problems in the 5 stations with the educational objectives of the general pediatric training program. Because no criterion standard for this test exists, the construct validity of the OSCE was assessed with an analysis of variance (ANOVA) that compared the residents’ results, grouped by year, among the 4 years of training. The premise of this hypothetical educational construct is that students with additional training should do better on an assessment of their performance.

The results of the OSCE were compared with those of the ITERs to assess the concurrent validity of the OSCE. Means, SDs, ANOVA scores, and Pearson product moment correlation coefficients were computed. The results were considered significant at P < .05.

RESULTS

The OSCE was administered to 43 of the 61 general pediatric residents: 14 in year 1, 12 in year 2, 8 in year 3, and 9 in year 4 (31 female and 12 male residents). Eighteen residents did not participate because they were away from the hospital, participating in elective courses or taking holidays; had been on call the previous night; or could not be released from their clinical responsibilities.

The mean scores of the residents on each of the 5 stations in each year of training are summarized in Table 1. There were no significant (P > .05) differences in the total OSCE scores among the residents grouped by their years of training. There were no significant (P > .05) costs for the standardized patient, the standardized patient trainer, and the refreshments served.

On completion of all 5 stations, the residents were given the expectations for each station and an outline of the checklist. After all the examinations were marked, the residents received their total score for the OSCE and their individual scores for each of the 5 stations. They were also given the mean (±SD) of the scores for all the residents who participated in the examination and for the residents in their year of training, plus any written comments made by the examiners.

COMPARISON OF THE OSCE WITH THE ITER

At the end of the 1995-1996 academic year, we reviewed the rotational ITERs. The ITERs’ descriptive assessments of the residents’ clinical skills, knowledge, clinical judgment, professional attitudes, and overall assessment were each converted to a numerical score on a 5-point Likert scale: 1, unsatisfactory; 2, below expectations; 3, meets expectations; 4, above expectations; and 5, outstanding. All the ITERs for each resident were reviewed, and the mean scores were calculated for the overall assessment and for the individual ITER items. The ITERs evaluate the performance of residents according to the residents’ level of training. The Pearson product moment correlation coefficient was used to compare the results of the OSCE with the results of the monthly ITERs.

ANALYSIS OF THE DATA

The costs to administer this OSCE were recorded (ie, the...
differences in the total OSCE scores for female and male residents.

The mean scores for the residents on the checklist, the global assessment, and the PEP were further analyzed. There were few significant differences among the scores for each of the 4 years of residency: the fourth-year residents scored higher than the first-year residents on the written questions \((P<.05)\). The 31 women scored significantly higher on the global assessments than the 12 men \((P<.05)\).

The scores for 1 of the fourth-year residents were clearly outliers. This resident had the lowest score on the OSCE, did not complete the year, and dropped out of training. If the scores for this resident were taken out of the analysis, the fourth-year residents scored significantly higher than the first- or second-year residents \((P<.05)\).

Because of the few residents involved in this OSCE and the few stations used, no pass or fail score was calculated. But 2 residents, 1 in the first year and 1 in the fourth year, had total scores of less than 60%, the traditional pass mark for undergraduate medical students at the University of Toronto Faculty of Medicine.

**Comparison of the Results of the OSCE with the Results of the Monthly ITERs**

There were 491 ITERs for the 61 residents, or a mean of 8.0 ITERs per resident. The mean scores are listed in Table 2. Overall, the mean score was 3.9; 4 residents were rated outstanding \((a score \geq 4.50)\); 52, above expectations \((a score of 3.50-4.49)\); and 5, meets expectations \((a score of 2.50-3.49)\). There was little variation in the faculty ratings of residents’ performance; faculty tended to rate residents higher on professional attitudes than on knowledge. There were no significant \((P \geq 0.05)\) differences among the 4 classes of residents because faculty completing the ITERs evaluated the performance of residents according to the residents’ level of training.

There were no significant \((P \geq 0.05)\) differences in ITER scores between the 42 female and the 19 male residents. There were no differences in the ITER scores for the 43 residents who participated in the OSCE and the 18 residents who did not take part in the OSCE.

There were significant correlations \((P<.05, Pearson product moment correlation coefficient) between OSCE scores \((total and components) and ITER scores \((overall and components)\). When the total OSCE scores were compared with the overall ITER scores, the correlation coefficient was 0.45, an understandably modest correlation because the 2 assessed different aspects of clinical competence (Figure). The strongest correlation was between the global assessment score and the overall ITER score \((r=0.57)\).

**The Reliability and Validity of the OSCE**

The reliability of the 5-station OSCE, as measured by the coefficient of internal consistency or Cronbach \(\alpha\), was 0.69.

The 5 stations used tested the different aspects of general and subspecialty pediatrics that were thought to cover the essential educational objectives for a postgraduate pediatric training program. However, because only 5 stations were used, the content validity of this OSCE was limited.

The total OSCE did not demonstrate construct validity, that is, this OSCE did not demonstrate statistically significant differences among residents in each of the 4 years of training. If the examination had construct validity, the scores would have been greater for residents with more experience and training. However, fourth-year residents scored higher than first-year residents on the PEP questions, which indicates that the fourth-year residents’ knowledge base may be better than that of the first-year residents.

---

### Table 1. Mean Scores for All Residents for All Stations of the 1996 Objective Structured Clinical Examination*

<table>
<thead>
<tr>
<th>Year (No. of Residents)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (14)</td>
<td>20.8±3.0</td>
<td>18.8±2.8</td>
<td>21.2±2.9</td>
<td>21.2±2.2</td>
<td>20.4±2.9</td>
<td>102.5±8.8</td>
</tr>
<tr>
<td>2 (12)</td>
<td>21.1±3.1</td>
<td>19.1±3.1</td>
<td>19.0±2.7</td>
<td>22.7±2.5</td>
<td>19.7±2.6</td>
<td>101.6±7.5</td>
</tr>
<tr>
<td>3 (8)</td>
<td>21.0±3.2</td>
<td>20.7±3.1</td>
<td>21.9±2.9</td>
<td>22.8±1.7</td>
<td>22.6±3.5</td>
<td>109.0±9.9</td>
</tr>
<tr>
<td>4 (9)</td>
<td>22.7±3.6</td>
<td>21.3±3.2</td>
<td>20.6±3.2</td>
<td>23.1±3.2</td>
<td>21.8±2.7</td>
<td>109.4±12.8</td>
</tr>
<tr>
<td>Total (43)</td>
<td>21.4±3.2</td>
<td>19.8±3.1</td>
<td>20.6±3.0</td>
<td>22.3±2.5</td>
<td>20.9±3.0</td>
<td>104.9±9.9</td>
</tr>
</tbody>
</table>

*All scores are given as the mean (±SD). The reliability (coefficient of internal consistency) equals 0.69.

### Table 2. Mean Scores for the In-Training Evaluation Reports*

<table>
<thead>
<tr>
<th>Year (No. of Residents)</th>
<th>Knowledge</th>
<th>Clinical Skills</th>
<th>Judgment</th>
<th>Professional Attitude</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (17)</td>
<td>3.7±0.3</td>
<td>3.8±0.3</td>
<td>3.8±0.3</td>
<td>4.2±0.3</td>
<td>3.9±0.3</td>
</tr>
<tr>
<td>2 (16)</td>
<td>3.6±0.3</td>
<td>3.7±0.3</td>
<td>3.7±0.3</td>
<td>4.0±0.2</td>
<td>3.9±0.3</td>
</tr>
<tr>
<td>3 (15)</td>
<td>3.7±0.2</td>
<td>3.8±0.2</td>
<td>3.8±0.1</td>
<td>4.1±0.2</td>
<td>3.9±0.2</td>
</tr>
<tr>
<td>4 (13)</td>
<td>3.7±0.6</td>
<td>4.0±0.4</td>
<td>3.9±0.5</td>
<td>4.3±0.7</td>
<td>4.1±0.5</td>
</tr>
<tr>
<td>Total (61)</td>
<td>3.7±0.4</td>
<td>3.8±0.3</td>
<td>3.8±0.3</td>
<td>4.1±0.4</td>
<td>3.9±0.3</td>
</tr>
</tbody>
</table>

*All scores are given as the mean (±SD). The residents were scored on a 5-point scale, where 1 indicates unsatisfactory; 2, below expectations; 3, meets expectations; 4, above expectations; and 5, outstanding.
It is clear from this study that the OSCE can be used to assess the clinical performance of postgraduate residents in pediatrics. The OSCE is a valid and reliable format uniquely capable of assessing many fundamental clinical skills that are not being assessed in a rigorous way in most postgraduate programs. Because of the few stations used, the reliability and content validity of this OSCE are limited. However, this OSCE was used as a formative assessment to provide feedback to pediatric residents at the University of Toronto Faculty of Medicine and not as a summative assessment to pass or fail residents for the year’s rotation. The OSCE score was correlated with the mean score of the residents’ ITERs, providing some measure of the concurrent validity of the OSCE. This OSCE evaluated our residents’ clinical skills by means of the examiners’ observation of the residents’ interviewing and history-taking skills in a structured format, which is uncommon for faculty completing the monthly ITERs. The highest correlation was between the examiners’ global score and the mean score of the ITERs, perhaps because these are the most comparable—both measure the staff’s general assessment or feeling of the resident’s overall competence.

The OSCE has advantages and disadvantages as a method of ongoing evaluation of the residents. The OSCE provided the postgraduate program director with a reliable assessment of the clinical skills of residents who participated in the examination. It identified 2 residents whose performance was less than acceptable compared with that of other residents in their year. The OSCE also provided the residents with useful feedback about their performance. The faculty also benefited from being examiners. It helped them learn how to critically appraise residents’ clinical skills and gave them some structure for the comparison of residents.

There were some difficulties and disadvantages of the OSCE. There were logistic difficulties in organizing all the residents, the staff, the rooms, and the trained standardized patients. It was difficult to find 15 rooms that were not being used. We were able to use a hospital ward that had been closed. Other centers administer the OSCE on weekends in clinic areas that are not used, but we had agreed to administer the OSCE during a weekday and not to require the residents to come to the hospital when they were off call.

Objective structured clinical examinations are expensive to administer. Our expenses for this OSCE, which included the costs of training and hiring standardized patients (Canada, $70 per resident; United States, $50 per resident), were comparable with those of other studies. However, this expense is considerably more than that of any other formative assessment being used.

The usefulness of the OSCE for our program director was limited because not all the residents could take part. Only 43 of 61 pediatric residents participated in the OSCE. The remaining 18 residents were away from the hospital, participating in elective courses, taking holidays, or completing off-site rotations; were on call, having been up the previous night; or were on rotations that did not permit them to get away (such as rotations in the pediatric intensive care unit, neonatal intensive care unit, or emergency department). Finally, this OSCE was limited to interviewing and history-taking skills. We did not have any stations that assessed physical examination skills.

Feedback would be more effective if it were given after each station; this feedback might include the residents’ scores and comments from the examiner or standardized patient. However, immediate feedback would limit the usefulness of the PEP questions if the residents were given the outline for the station before they were to write the PEP. Immediate feedback might also detract from the residents’ performance at subsequent stations if they were to dwell on their performance at one station, especially if they had done poorly at a station. Research has shown that physicians perform better when they repeat a station when they have been given immediate feedback, but whether candidates perform better or worse on subsequent or new stations when given immediate feedback has not been studied.

We elected to give residents feedback only after they had completed all 5 stations of our OSCE, to mimic more closely the experience of the Medical Council of Canada Qualifying Examination. Residents expressed appreciation for this feedback, but more experience with the OSCE
in postgraduate programs is needed to be able to maximize the feedback to residents.

HOW CAN OSCE STATIONS BE USED TO ASSESS THE PHYSICAL EXAMINATION SKILLS IN PEDIATRICS?

Our stations were all interviewing, history-taking, and giving advice stations that used adult or teenage standardized patients or parents. We did not assess residents' physical examination skills. Other OSCE stations that use adult standardized patients assess candidates' physical examination skills. In these stations, standardized patients are trained to demonstrate certain findings on physical examination, such as back pain or abdominal pain; a checklist is developed to assess whether candidates perform the essential components of the physical examination.

Some pediatric centers have used standardized patients, normal newborns or patients with similar findings (eg, heart murmur caused by a ventricular septal defect and hepatosplenomegaly). We were concerned that a child might not act consistently with 12 to 15 different residents performing a physical examination. We were more concerned about the ethical issues of using children for educational purposes, especially when they may be examined by 10 to 15 different physicians. Perhaps other methods, such as videos or mannequins, could be used to assess examination skills without having young children repeatedly examined. Video stations demonstrating a neurologic examination or a developmental assessment, or audiovisual stations demonstrating an examination of the cardiovascular system, might be used as stations to assess residents' observational skills.

WHAT ETHICAL ISSUES ARISE WHEN CHILDREN ARE USED AS SUBJECTS FOR AN OSCE STATION?

The key to standardized patients participating in OSCEs is that they volunteer, are trained, and are paid to portray a patient with a specific clinical problem. Can this be done ethically with children? For instance, is it ethical to have a newborn examined by several residents during an OSCE if the examination does not contribute to the newborn's medical care? In our OSCE, we used 13-year-old boys as standardized patients who gave their history to 12 different residents. We believed that they would be able to act reliably as patients and that they were competent to give their consent to be standardized patients. Other centers already cited that have pediatric physical examination stations have used either children who do not have any specific findings or children who all have the same finding, for example, a murmur of a ventricular septal defect. If a child is to be examined by 10 to 20 different students or physicians, who gives consent and does the child understand what is expected of him or her? In providing consent for medical treatment, parents usually give consent “in the best interest of the child,” weighing the benefits and the harms of the treatment. Is it ethical to ask parents to consent to the use of their children in OSCEs when there may be no benefit to the child? We decided not to use any children in our OSCE and would like other centers to discuss the ethical aspects of volunteers giving consent to participate in OSCE stations.

The OSCE is a reliable and valid method for the assessment of the clinical competence of postgraduate pediatric residents. When used in conjunction with other evaluation formats, the OSCE can provide an objective assessment of a pediatric resident’s progress. The extent to which the evaluation of a resident’s clinical skills and the feedback of this small OSCE adds to other formative assessment methods, given the expense and logistic difficulties of administering the OSCE, needs further study and discussion. We will continue to use the OSCE but will study ways to improve the usefulness of the OSCE as a formative assessment of our pediatric residents.

Accepted for publication June 12, 1997.

This study was supported by the Paediatric Consultants, Department of Paediatrics, the Hospital for Sick Children, Toronto. We thank the members of Editorial Services, the Hospital for Sick Children, for their assistance with this article.

Corresponding author: Robert I. Hilliard, MD, EdD, FRCP, Division of General Paediatrics, the Hospital for Sick Children, 355 University Ave, Toronto, Ontario, Canada MSG 1X8.

REFERENCES