Objective Structured Assessment of Cataract Surgical Skill

George M. Saleh, MRCSEd, MRCOphth; Vinod Gauba, PAMEd, MRCOphth; Arijit Mitra, DO, MRCSEd; Andre S. Litwin, MBBS; Andrew K. K. Chung, FRCOphth; Larry Benjamin, FRCOphth

Objective: To evaluate the Objective Structured Assessment of Cataract Surgical Skill scoring system.

Methods: An objective performance rating tool was devised. This instrument is comprised of standardized criteria with global rating and operation-specific components, each rated on a 5-point Likert scale. The total potential score was 100. Complete phacoemulsification cataract extraction operations were recorded through the operating microscope of surgeons with a range of experience (group A, 50 procedures; group B, 50-249 procedures; group C, 250-500 procedures, and group D, >500 procedures). These were then scored by independent expert reviewers masked to the grades of the surgeons. The U test was used to evaluate statistical significance.

Results: We evaluated 38 surgical videotapes of 38 surgeons (group A, 11 surgeons; group B, 10 surgeons; group C, 5 surgeons; and group D, 12 surgeons). Mean±SD overall scores were as follows: group A, 32.0±5.3; group B, 55.0±12.6; group C, 89.0±4.7; and group D, 90.0±11.1. Statistically significant differences were found between groups A and B (P=.002) and groups B and C (P=.003), but not between groups C and D (P>.99).

Conclusion: The Objective Structured Assessment of Cataract Surgical Skill scoring system seems to have construct validity with cataract surgery and, thus, may be valuable for assessing the surgical skills of junior trainees.

Arch Ophthalmol. 2007;125:363-366

There are, however, few formal tools for evaluating the surgical competence of ophthalmology trainees, and often the style and consistency of feedback can vary as they progress through different training centers. Consequently, ophthalmic surgical training programs have changed considerably, striving to make assessments more transparent and objective. In part, this has been because of the recognition that deficiencies in training and performance are more difficult to identify and correct without objective data. In the clinical setting, tools such as the Objective Structured Clinical Examination have proved successful, but formal methods of operative assessment have developed more slowly. Reznick and his colleagues developed the Objective Structured Assessment of Technical Skill (OSATS) for evaluation of general surgical trainees, which was later validated for other surgical specialties. In this study, we used a modified version of OSATS as applied to cataract surgery to formally assess discrete segments of surgical tasks and global performance and to examine construct validity.

Author Affiliations:
Departments of Ophthalmology, Royal Surrey County Hospital, Guildford, England (Messrs Saleh, Mitra, and Litwin); Leeds General Infirmary, Leeds, England (Messrs Gauba and Chung); and Stoke Mandeville Hospital, Aylesbury, England (Mr Benjamin). Mr Mitra is now with Wolverhampton and Midland Counties Eye Infirmary, Wolverhampton, England.

For editorial comment see page 395

METHODS

The Objective Structured Assessment of Cataract Surgical Skill (OSACSS, Figure 1) was used as the grading system. It consists of global and phacoemulsification task-specific elements. The global rating system was adapted from the OSATS tool previously validated for assessment of technical skills in the operating room and in simulated operations. The task-specific component consisted of a checklist par-
ticular to cataract surgery. Generic end points were developed that allowed the system to be applied to different techniques (eg, divide and conquer vs pure chop). Several pilot studies were performed with input from senior surgical trainers throughout England. Based on this feedback, the tool was refined to weight the scoring appropriately so that different parts of the operation, with differing degrees of complexity, were deemed to be assessed proportionately. Each of the global and task-specific components was rated on a 5-point Likert scale, with scores ranging from 1 (poor performance) to 5 (good performance). There are 30 available marks for the global portion and 70 available marks for the task-specific portion. These are summed, for a maximum potential score of 100.

Ethical committee approval was sought and obtained before commencing the study. Data were collected from 8 eye units throughout England. Four grades of surgeons were recruited with a varying range of experience. Surgeons in group A had performed fewer than 50 complete procedures; those in group B had performed 50 to 249 complete procedures; in group C, 250 to 500 complete procedures; and in group D, more than 500 complete procedures. Full cataract surgical operations were recorded through the operating microscope. Case patients with ophthalmic comorbidity, poor dilation, mature cataract, intraoperative complication, or other complexity identified at preoperative assessment were excluded. All cases recorded were deemed by the authors to have been suitable for the most junior group to undertake, thus promoting consistency. The videotapes were sent to an independent technician who digitized them, standardizing the size and color and removing any logos or other characteristics that could identify the surgeon or training unit. The videotapes were anonymized and randomized. Three independent expert reviewers then applied the grading system to the videotapes. The test was used to evaluate statistical significance.

Figure 1. The Objective Structured Assessment of Cataract Surgical Skill instrument.
RESULTS

Thirty-eight surgical videotapes from 38 surgeons were evaluated (group A, 11 surgeons; group B, 10 surgeons; group C, 5 surgeons; and group D, 12 surgeons). The global, task-specific, and total scores for each group are given in the Table. The task-specific scores are derived from 14 cataract-specific stems, and the global scores are derived from another 6 stems. The total scores are calculated by combining the totals of the 20 stems (Figure 1).

For the total scores, statistically significant differences were found between groups A and B (P = .002) and groups B and C (P = .003), but not groups C and D (P > .99). For the task-specific scores, statistically significant differences were found between groups A and B (P = .002) and groups B and C (P = .006), but not groups C and D (P = .79). For the global scores, statistically significant differences were found between groups A and B (P = .001) and groups B and C (P = .002), but not groups C and D (P = .97). Total scores, global scores, and task-specific scores, plotted against surgeon experience, are shown in Figure 2.

The task-specific and global scores for the 4 groups were individually compared with each other. There was a significant improvement in all scores between group A (<50 complete procedures) and group B (50-249 complete procedures; <.001 < P <.008). A significant improvement was noted for many of the scores between group B and group C (250-500 complete procedures; .003 < P < .44) but not for the draping, viscoelastic, hydrodissection, phacoemulsification probe and second instrument insertion, nucleus sculpting, nucleus rotation, irrigation and aspiration, and wound closure scores. There was no significant improvement in scores between group C and group D (>500 complete procedures; .23 < P < .96).

**Table.** Overall OSACSS Scores

<table>
<thead>
<tr>
<th>Group*</th>
<th>No. of Surgeons</th>
<th>Mean ± SD Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>38</td>
<td>65.0 ± 24.9</td>
</tr>
<tr>
<td>Task-specific score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>11</td>
<td>23.0 ± 3.4</td>
</tr>
<tr>
<td>B</td>
<td>10</td>
<td>42.0 ± 9.4</td>
</tr>
<tr>
<td>C</td>
<td>5</td>
<td>63.0 ± 3.1</td>
</tr>
<tr>
<td>D</td>
<td>12</td>
<td>63.0 ± 7.5</td>
</tr>
<tr>
<td>Total</td>
<td>38</td>
<td>48.0 ± 19.45</td>
</tr>
<tr>
<td>Global score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>11</td>
<td>7.0 ± 2.6</td>
</tr>
<tr>
<td>B</td>
<td>10</td>
<td>14.0 ± 3.5</td>
</tr>
<tr>
<td>C</td>
<td>5</td>
<td>26.0 ± 1.9</td>
</tr>
<tr>
<td>D</td>
<td>12</td>
<td>27.0 ± 3.85</td>
</tr>
<tr>
<td>Total</td>
<td>38</td>
<td>18.0 ± 8.1</td>
</tr>
</tbody>
</table>

Abbreviation: OSACSS, Objective Structured Assessment of Cataract Surgical Skill.

*Group A, surgeons who had performed fewer than 50 complete procedures; group B, 50 to 249 complete procedures; group C, 250 to 500 complete procedures; and group D, more than 500 complete procedures.
In this study, the OSACSS was used to assess anonymized surgical recordings. The results show that this tool is effective for differentiating between surgeons of varying experience up to 250 procedures, but is less sensitive for more senior surgeons. This may be, in part, because there are fewer experienced surgeons and, in part, because of convergence of surgical skill requiring a more sensitive or detailed tool to detect the difference. These findings are in contrast to the original OSATS studies that found that the correlation between faculty ranking and OSATS scores was high for senior residents but low for junior residents. It seems, therefore, that the OSACSS used in this study may well be complementary to the original evaluation.

Three criteria were particularly useful in differentiating between group A, with fewer than 50 complete procedures, and group B, with 50 to 249 complete procedures, and continued to differentiate well between group B and group C, with 250 to 500 complete procedures, suggesting that continued acquisition and improvement of surgical skill is required to perform these tasks. These criteria (with respective score changes and P values for groups A and B and groups B and C) were as follows: continuous curvilinear capsulorrhexis flap construction (score change, 2.15, P < .001; and score change, 1.6, P < .003), incision and paracentesis (score change, 1.8, P < .001; and score change, 1.2, P < .013), and speed and fluidity of procedure (score change, 1.67, P < .001; and score change, 1.2, P < .001). A further 3 criteria showed a high level of discrimination between groups A and B, but no significant score change between groups B and C, suggesting that these skills are acquired earlier in training. These criteria (with respective score changes and P values for groups A and B and groups B and C) were as follows: wound closure (score change, 2.08, P < .001; and score change, 0.26, P = .43), nucleus rotation (score change, 1.73, P < .001; and score change, 0.33, P = .33), irrigation and aspiration (score change, 1.15, P < .05; and score change, 0.16, P = .68).

The OSACSS has advantages over direct observation techniques such as the OSATS and its adaptations. One of the most important is that it is free of direct observational bias because the recording and, thus, the surgeon can be completely anonymized for the grader. Also, the original OSATS system was validated, but many of the subsequently adapted tools were not. This study demonstrated construct validity of the OSACSS system when applied to phacoemulsification surgery with the total scores as well as the global and task-specific scores.

Traditional surgical evaluations involve a single global rating made by a resident’s supervisors as to the adequacy of his or her technical proficiency. This type of rating may be inconsistent and unreliable, and cannot be considered an adequate assessment of technical skill on which to base formal feedback. Checklists and detailed global rating scales that assess operative skills have shown promise in other surgical disciplines, demonstrating feasibility, interassessor reliability, and construct validity. In 1997, the Accreditation Council for Graduate Medical Education endorsed the use of educational outcomes measures in ophthalmology, and competence in surgery was included. Similarly, in England, the Royal College of Ophthalmologists and its overseeing body have recommended work-based assessments to be implemented, one of which is an OSATS-derived tool. The OSATS-based systems of direct observation have begun to emerge in ophthalmology, such as Objective Assessment of Skills in Intraocular Surgery (OASIS) and Global Rating Assessment of Skills in Intraocular Surgery (GRASIS). These can be labor intensive and require a specific setup (the Objective Assessment of Skills in Intraocular Surgery instrument is based on a software and computer database). The OSACSS differs notably from these systems in its design by striving to be more simple in its setup and application.

Good assessment procedures ensure that program objectives are being met. Assessment is fundamental for promotion, certification, and licensure. No single method can adequately or comprehensively assess the surgical skills of residents in training. We believe the OSACSS is complementary to other methods and that each has advantages and disadvantages. As more research is carried out in this field, it is hoped that, in time, these systems will help to standardize surgical skill assessment, streamline surgical training, and improve overall surgical performance.

Submitted for Publication: May 2, 2006; final revision received July 8, 2006; accepted July 23, 2006.

Correspondence: George M. Saleh, MRCSed, MRCPs, Department of Ophthalmology, Frimley Park Hospital, Guildford Road, Frimley GU16 7UJ, Surrey, England (drgmsaleh@yahoo.co.uk).

Financial Disclosure: None reported.

REFERENCES