Introduction

Endoscopic sinus surgery (ESS) is an essential component of otolaryngology surgical training.1 Endoscopic surgery is a reliable choice for removing even large, posteriorly-located, intranasal lesions, e.g., intranasal pleomorphic adenoma can be approached using this technique, without open surgery.2 Competence has become an important issue in current surgical practice and training.2 There is also a pressing need for an intraoperative assessment tool that meets the high standards of reliability and validity as an outcome measure for different training strategies, and for recruiting foreign graduates into countries.2

The goal of our study was to determine the validity and reliability of an assessment tool for ESS.

Materials and methods

Data were collected prospectively in an observational study through evaluations at two tertiary academic institutions, i.e. St. Paul’s Sinus Centre, St. Paul’s Hospital, Vancouver, British Columbia, Canada, and King Fahd Medical City, Riyadh, Saudi Arabia, from December 2006 to December 2009. A 2-page evaluation form was developed in conjunction with the Objective Assessment of Technical Skills Surgery (OSATS) evaluation form developed by Reznick et al in Toronto to assess residents’ surgical skills. A Likert scale (1-5 where 5 = excellent) was used for evaluations. The Global Rating of Endoscopic Surgical Skills (GRESS) evaluation instrument was designed with input from academic otolaryngologists, fellowship-trained rhinologists, and experts in medical education. The experts’ comments were incorporated, establishing face and content validity. Residents from various levels of training were assessed objectively using this instrument. Internal consistency was evaluated using Cronbach’s α. Test-retest and inter-rater reliability was measured using intra-class correlation.

Results: A total of 31 assessments were completed by 15 residents. GRESS showed high reliability in the context of internal consistency (α = 0.99), test-retest (0.95, CI = 0.83-0.98), and inter-rater reliability (0.86, CI = 0.31-0.98).

Conclusions: This pilot study demonstrated that GRESS is a valid and reliable assessment tool for operating room performance.
comments were incorporated, establishing face and content validity. Residents from various levels of training were assessed objectively using this instrument intraoperatively, and were provided constructive feedback. Internal consistency was evaluated using Cronbach's $\alpha$. Test-retest and inter-rater reliability was measured using intra-class correlations. Data were analyzed using SPSS version 16 (SPSS Inc., Chicago, IL, USA).

**Ethical Issues**

The Institutional Review Boards of the institutions involved granted us permission to conduct this study. We declare that we have no financial or personal relationship(s) which may have inappropriately influenced us in writing this paper.

**Results**

A total of 31 assessments were completed for 15 residents who were evaluated by 5 faculty members as they performed ESS on patients over a period of 3 years from 2006-2009. Three residents were observed by 2 faculty members simultaneously.

Internal consistency, evaluated using Cronbach’s $\alpha$, showed high reliability ($\alpha = 0.99$). Test-retest reliability measured using intra-class correlation coefficient was also found to be high (ICC = 0.95, CI = 0.83-0.98). The intra-class correlation coefficient for the inter-rater reliability was ICC = 0.86 (CI = 0.31-0.98). In this study, both instruments showed construct validity, with an overall trend toward a higher score with a more advanced postgraduate year of training, and the faculty substantially outperformed residents (Figure 1).

**Discussion**

It seems obvious that direct observation of surgical skills in the operating room represents the ‘gold standard’ in terms of content and construct validity, knowing that surgical knowledge is probably best assessed by examinations, but surgical skills are best assessed in the workplace. Log-books form a useful record of experience gained, but experience does not necessarily equate with competence; therefore, developing such tools is vitally important. A good assess-

---

Table 1

<table>
<thead>
<tr>
<th>Tool 1: Likert Scale for Global Rating of Surgical Skills (Martin *et al.*4)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rating</strong></td>
</tr>
<tr>
<td><strong>Criterion</strong></td>
</tr>
<tr>
<td>Respect for tissue</td>
</tr>
<tr>
<td>Time &amp; Motion</td>
</tr>
<tr>
<td>Instrument handling</td>
</tr>
<tr>
<td>Knowledge of instruments</td>
</tr>
<tr>
<td>Flow of operation and forward planning</td>
</tr>
<tr>
<td>Knowledge of specific procedure</td>
</tr>
<tr>
<td>Use of assistants</td>
</tr>
</tbody>
</table>
ment must possess reliability, validity, educational impact, acceptability, and feasibility. Assessment can be defined as making a judgment against a defined reference, and it has two main purposes, which ideally should not interfere with each other. The first is to provide feedback to aid learning, i.e. a formative or low-stakes assessment, and the second is for examination/certification, i.e. a summative or high-stakes assessment. Competence-based assessments measure what a surgeon can do in a controlled representation of professional practice, e.g. when observed in the operating room or in a clinical skills laboratory. Miller’s triangle defines a simple hierarchy for the development and assessment of clinical skills. The GRESS is based on this model, and was developed by the modified Delphi technique, which uses email to gather information, provide feedback, and report conclusions. Simulations have been suggested to form the basis for technical skills training and assessment in the future. This is because of the decreasing opportunity to practice on real patients and the need for ‘deliberate practice’ in a non-threatening environment. Practice and assessment on simulations are no substitute for operative experience, but they enable surgeons to

### Table 2

<p>| Tool 2: Global Rating of Endoscopic Surgical Skills (RESS) |
|----------------------------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th><strong>Rating</strong></th>
<th><strong>Unsatisfactory</strong></th>
<th><strong>Adequate/satisfactory</strong></th>
<th><strong>Excellent/competent</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Not positioned properly.</td>
<td>Had to re-adjust during the procedure.</td>
<td>Positioned well, no need for re-adjustments.</td>
</tr>
<tr>
<td>2</td>
<td>Unaware of eye protection. Didn’t consider it.</td>
<td>Improper coverage of the eyes.</td>
<td>Lube in the eyes and taped properly.</td>
</tr>
<tr>
<td>3</td>
<td>Unaware of proper position. Didn’t consider it.</td>
<td>Improperly positioned, made some effort.</td>
<td>Appropriately positioned, appropriate rest, e.g egg foam crate.</td>
</tr>
</tbody>
</table>

**A Patient Setup**

1. Position of ETT
   - Not positioned properly.
   - Had to re-adjust during the procedure.
   - Positioned well, no need for re-adjustments.

2. Eye protection: Lube in the eyes. Uncovered
   - Unaware of eye protection. Didn’t consider it.
   - Improper coverage of the eyes.
   - Lube in the eyes and taped properly.

3. Head/patient positioning
   - Unaware of proper position. Didn’t consider it.
   - Improperly positioned, made some effort.
   - Appropriately positioned, appropriate rest, e.g egg foam crate.

4. Patient preparation:
   - Nasal decongestants -Throat pack
   - Unaware of patient preparation. Not considered.
   - Makes some effort to prepare patient. Placed improperly.
   - Ensures correct preparation of the patient: nasal decongestion, placement of throat pack.

**B Equipment Setup**

1. Proper setup of IGS/TV tower/endoscopes
   - Unaware of proper setup
   - Setup improperly done.
   - Setup properly done
   - Endoscope: correct image size, focusing of image, white balance.

2. Proper draping and placement of IGS tower
   - Unaware, did not attempt
   - Improperly placed and draped
   - Properly placed and draped

**C Use of endoscope**

1. Operative field
   - Operative field rarely central
   - Operative field usually central
   - Operative field always central

2. Perception of depth
   - Lack of perception of depth, causing trauma to structures
   - Incomplete perception of depth.
   - Complete perception of depth

3. Instrument visualization endoscopically
   - Rarely keeps tip of instrument in view.
   - Sometimes keeps instrument tip in view, but not consistently
   - Consistently keeps instrument in view

4. Communication with anesthesiologist when injecting local, moving head, BP control
   - No communication
   - Sometimes, not consistent
   - Consistent communication

---

become competent (and, therefore, confident) in key surgical skills before entering the complex operating room environment. A randomized trial showed that trainees who receive simulation training perform significantly better in the operating room. Others have developed tools, but only at the cadaveric level.

The faculty found the instrument to be easy to understand, complete, and practical and the residents felt that the instruments were helpful in providing immediate informative feedback on their performance.

In this study, we could not present data regarding complications that occurred with each resident during the evaluation while performing surgery. Actually, the GRESS assessment tool was not designed to measure the possible surgical complications that could occur with any senior surgeon, and every surgical procedure has its own risk of complications. What really matters is the rate of surgical complications, which could only be valid for clinical effectiveness measurements if calculated over a certain time period by repeating the same procedure a certain number of times. Because the assessment rate for each resident was limited, it not valid to take their complication rates into account to confirm the validity of testing.

Conclusions

This pilot study demonstrated GRESS to be a valid and reliable assessment tool of operating room performance to provide systematic and comprehensive feedback as part of the learning cycle. This tool was developed as an intra-operative instrument as opposed to other tools which used cadavers. Potential applications will be in tracking resident development throughout postgraduate training, and offering a structured means of certifying operative skills.

Limitations

Recommendations

Because of the limited number of subjects in this study, more data should be gathered by using our instrument on a larger scale, e.g. in other residency programs and over a longer period of time. Other assessment tools and checklists specific for otolaryngology are currently being developed for research and evaluation.

Acknowledgement

The authors are very thankful to Dr. Mian Usman Farooq, Senior Specialist, King Abdullah Medical City, Makkah, Saudi Arabia for his technical support.

References


Osama Marglani, Assistant Professor
Department of Otolaryngology
Head & Neck Surgery,
College of Medicine,
Umm-al-Qura University,
Consultant Otolaryngology
King Abdullah Medical City, Makkah, Saudi Arabia
Tel.: 00966-566500350
Fax: 00966-25664663
E-mail: osamamergalani@yahoo.com