Otology & Neurotology
31:759–765 © 2010, Otology & Neurotology, Inc.

Objective Assessment of Mastoidectomy Skills in the Operating Room

Howard W. Francis, Hamid Masood, Kashif N. Chaudhry, Kulsoom Laeeq, John P. Carey, Charles C. Della Santina, Charles J. Limb, John K. Niparko, and Nasir I. Bhatti

Department of Otolaryngology–Head and Neck Surgery, Johns Hopkins Hospital, Baltimore, Maryland, U.S.A.

Objective(s): To determine the feasibility and validity of an objective assessment tool designed to measure the development of mastoidectomy skills by resident trainees in the operating room.
Study Design: Prospective longitudinal validation study.
Setting: Tertiary referral center and residency training program.
Subjects: Otolaryngology residents.
Main Outcome Measure: Technical performance as measured over time using Task-Based Checklist (TBC) and Global Rating Scale (GRS) developed for assessment of mastoidectomy skills.
Results: Seventy pairs of evaluations were completed on 15 residents, showing strong correlation between both instruments ($r = 0.93$; $p < 0.0001$). Our instrument demonstrated construct validity for both TBC and GRS, showing higher scores with increasing surgical experience in otology. Both instruments showed high interitem reliability with Cronbach α coefficients of 0.98 and 0.95 for TBC and GRS, respectively. Regression analysis showed that thinning posterior external auditory canal ($p < 0.05$) and opening antrum to deepen dissection at sinodural angle ($p < 0.05$) were the strongest predictors of overall surgical performance.
Conclusion: Our assessment tool is a feasible and valid method of evaluating acquisition of mastoidectomy skills in the operating room. It can be integrated into surgical teaching in the operating room and yields information for direct formative feedback. **Key Words:** Competence—Evaluation—Mastoidectomy—Objective structured assessment of technical skills—Operating room—Otolaryngology—Surgical education.
Otol Neurotol **31:**759–765, 2010.

With the introduction of the Accreditation Council for Graduate Medical Education outcome project (1), residency programs are under increasing pressure to develop and implement objective tools for the evaluation of their trainees' competence. Existing skill assessment methods (2,3), however, use norm-referenced testing and involve significant subjectivity. Studies have shown such methods to be unreliable (4). Means of assessment based on recall such as In-Training Evaluation Reports have also been criticized for demonstrating poor validity and reliability (5,6). There is a need to develop more objective assessment tools that use criterion-referencing wherein trainees are measured against defined criteria and not relative to one another.

Objective structured assessment of technical skill is a performance-based assessment of technical skills developed by Martin et al. (7) that has been validated for almost a decade in other specialties and more recently in otolaryngology–head and neck surgery (8–10). Although a number of new tools have been developed to assess surgical technical performance, their use remains within the confines of surgical laboratories (2). With the exception of 1 study on tonsillectomy (11), all the objective evaluation tools developed in otolaryngology have been tested on cadavers, animal models, or simulation laboratories (12–14).

As a key indicator procedure for residency training in otolaryngology–head and neck surgery, mastoidectomy is critical to the surgical management of ear, temporal bone, and skull base diseases. Laeeq et al. (15) have pilot tested an assessment tool for competency in mastoidectomy on cadaveric temporal bone using both a global instrument and a Task-Based Checklist (TBC) instrument. At the end of a 3-year evaluation, the 2-part skills assessment tool was determined to be feasible, reliable, and valid in the dissection laboratory. These tools also provided insights that may be valuable in efforts to improve surgical instruction.

The aim of the present study is to document the feasibility and validity of a modified version of this 2-part mastoidectomy tool when used to evaluate residents'

Address correspondence and reprint requests to Nasir I. Bhatti, M.D., Department of Otolaryngology–Head and Neck Surgery, Johns Hopkins Outpatient Center, 601 N. Caroline Street, 6th Floor, Baltimore, MD; E- mail: nbhatti1@jhmi.edu

surgical skills in the operating room (OR). An additional objective of this study is to identify the individual skills that may serve as indicators of overall surgical performance.

MATERIALS AND METHODS

Study Design

This is a prospective study of mastoidectomy skill development and assessment in resident trainees over a 3-year period. Fifteen residents in the Johns Hopkins Department of Otolaryngology–Head and Neck Surgery training program were observed while performing mastoidectomy procedures in the OR and were evaluated at the end of each procedure by one of 6 faculty members in the division of Otology and Neurotology. Oral consent to participate in the survey was obtained at the start of each case. Failure to give consent did not affect residents' subsequent evaluation. The study was approved by the Institutional Review Board of the Johns Hopkins School of Medicine.

Components of the Assessment Instrument

Components of the assessment tools used for objective structured assessment of technical skills in our otology training program have previously been described as they were applied to use in temporal bone laboratory (15). The TBC evaluates the performance of a series of individual surgical steps required to safely accomplish the goals of the surgery (Table 1). Such procedure-specific checklists have also been validated by previous research (16). The Global Rating Scale (GRS) evaluates skill domains that include visual-motor and cognitive performance required for a successful mastoidectomy operation (Table 2). The mastoidectomy GRS was based on the tool previously developed by Winckle et al. (16) Reznick et al. (17), which has been validated for the assessment of technical skill both in the OR and the simulated environment. The design and creation of a TBC has been previously described (15).The items on the checklist part of the instrument are the critical steps of the mastoidectomy. These steps were identified by the Otology faculty using modified Delphi technique. By faculty input, a mastoidectomy TBC of 22 items was developed, with 5-point Likert scale with descriptors anchored at the middle and ends of the scale. These items were then clumped into 7 major tasks. To minimize the effect of prompting on the assessment, we added an instruction sheet to the assessment tool instructing the evaluators to weigh the need for significant prompting when scoring a resident's performance. Our faculty determined a minimal acceptable ("pass") level at the center (3) of the 5-point scale to allow assessment of residents at all levels of competency. This approach also allows residents to keep improving their competency level beyond this minimally acceptable level.

Process of Evaluation

Faculty evaluators observed residents performing mastoidectomy in the OR and evaluated their performance. Completed checklists were electronically submitted via the Internet-based E*Value Evaluation System (Advanced Informatics) (18). Residents participating in the study were in postgraduate years 2 to 6. Evaluations were performed regularly throughout the study period and were analyzed at the end of the study.

Statistical Analysis

The data were analyzed using STATView 5.0.1 (SAS Institute, Cary, NC, USA) and STATA 8.0 (StataCorp LP, College

TABLE 1. *Task-Based Checklist*

	Unable to perform		Performs with minimal prompting		Performs easily with good flow	
1. Initial bone cuts						
a. Placement of superior cut	1	2	3	4	5	N/A
b. Placement of canal cut	1	2	3	4	5	N/A
2. Defining anatomic limits						
a. Identification and definition of tegmen	1	2	3	4	5	N/A
b. Sharpen posterior EAC cortex	1	2	3	4	5	N/A
c. Define sigmoid sinus and sino-dural angle	1	2	3	4	5	N/A
3. Open antrum						
a. Deepen dissection at sinodural angle	1	2	3	4	5	N/A
b. Open antrum from posterior to anterior	1	2	3	4	5	N/A
c. Atraumatic exposure of short process of incus	1	2	3	4	5	N/A
4. Digastric dissection						
a. Define cephalic edge digastric muscle	1	2	3	4	5	N/A
b. Follow to SMF	1	2	3	4	5	N/A
5. Thin posterior EAC cortex (translucent)						
a. View posterior EAC en face	1	2	3	4	5	N/A
b. Use side/front of appropriate bur	1	2	3	4	5	N/A
c. Saucerization	1	2	3	4	5	N/A
6. Open facial recess						
a. Even removal of infralabyrinthine bone	1	2	3	4	5	N/A
b. Medial thinning of EAC cortex	1	2	3	4	5	N/A
c. Identify and preserve VII and chorda tympani	1	2	3	4	5	N/A
d. Identify ME anatomy	1	2	3	4	5	N/A
e. Decompress facial nerve to SMF	1	2	3	4	5	N/A
7. Posterior atticotomy						
a. Thin sup. EAC cortex	1	2	3	4	5	N/A
b. Thin anterior tegmen	1	2	3	4	5	N/A
c. Remove intervening bone	1	2	3	4	5	N/A
d. Identify epitympanic anatomyto supratubal recess	1	2	3	4	5	N/A

N/A indicates not applicable; SMF, stylomastoid foramen.

TABLE 2. *Global Rating Scale*

1. Understanding of indications/objectives for surgery					
1	2	3	4	5	N/A
Deficient understanding of indication/objectives of surgery		Understands most indications/objectives of surgery		Fully understands indications/objectives of surgery	
2. Interpretation of preoperative tests					
1	2	3	4	5	N/A
Unable to interpret and/or relate to surgical objectives		Able to interpret and/or relate to surgical objectives most of the times		Able to interpret and/or relate to surgical objectives easily	
3. Use of otologic drills					
1	2	3	4	5	N/A
Choose inappropriate bur and/or repeatedly awkward use of drill		Choose appropriate bur and occasionally awkward use of drill		Use appropriate bur and use drill effortlessly	
4. Knowledge of instruments					
1	2	3	4	5	N/A
Frequently asks for wrong instrument or used inappropriate instrument		Knew names of most instruments and used appropriate instrument		Obviously familiar with instruments and their names	
5. Use of microscope					
1	2	3	4	5	N/A
Repeatedly inappropriate position, magnification, or focus		Competent use of microscope but occasional inappropriate position and magnification		Optimal visualization and appropriate microscope use	
6. Respect for surgical limits					
1	2	3	4	5	N/A
Uses unnecessary force on tissue or caused damage by inappropriate use of instruments		Careful handling of tissue but occasional inadvertent damage to tissue		Consistently handled tissues appropriately with minimal damage	
7. Time and motion					
1	2	3	4	5	N/A
Many unnecessary moves		Efficient time/motion but some unnecessary moves		Clear economy of movement and maximum efficiency	
8. Knowledge of specific procedure					
1	2	3	4	5	N/A
Deficient knowledge and needed instruction at most steps		Knew all important steps of operation		Demonstrated familiarity with all aspects of operation	
9. Flow of operation					
1	2	3	4	5	N/A
Frequently stopped and unsure of next move		Some forward planning with reasonable progression		Obviously planned course of operation with effortless flow	
10. Overall surgical performance					
1	2	3	4	5	N/A
Poor		Performs majority of surgery acceptably		Outstanding	

Station, TX, USA) data analysis and statistical software packages. Internal consistency was assessed by measuring Cronbachs α as measure of interitem reliability for both checklist and global items. Construct validity to differentiate between an expert and a novice was determined by comparing scores across level of training in otology. We also identified the checklist items that had the strongest correlations to overall surgical performance, which is scored as a separate item on the GRS using simple regression analysis and further analyzed by multiple regression analysis. For all statistical purposes, a p value of less than 0.05 was considered significant.

RESULTS

Seventy pairs of TBC and GRS evaluations were completed for 15 residents by 6 evaluators over a 3-year period. Residents were evaluated by the faculty at the end of the each procedure. The instrument was found to be feasible based on the faculty feedback and time taken to complete the evaluation. Administering the tool in the OR did not prolong the OR time. The compliance rate approached 100% after making some modifications to make the evaluation process more feasible for faculty evaluators. These included the shortening of the instrument, addition of an instruction manual, and introduction

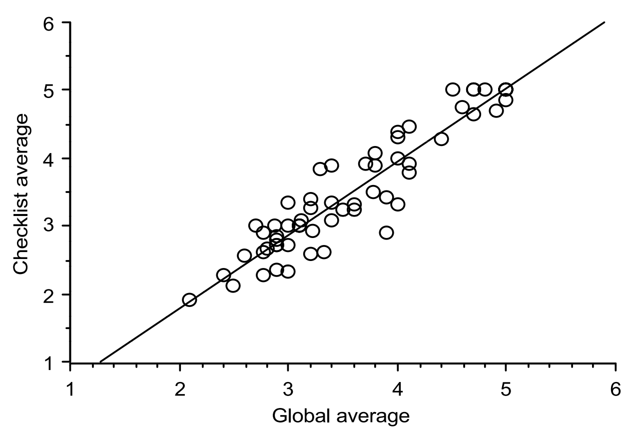

FIG. 1. Comparison of the TBC and GRS scores shows strong correlation ($r = 0.93$; $p < 0.0001$), suggesting that these assessments may be interchangeable.

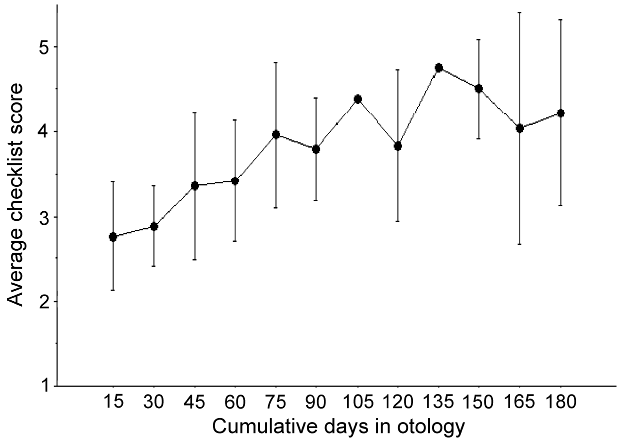

FIG. 2. Average TBC score plotted against cumulative days of concentrated otology experience showing improving performance with clinical experience ($r = 0.6$; $p < 0.0001$), which demonstrates construct validity.

of the paper evaluations in addition to computer-based evaluations.

There was strong agreement between the 2 (GRS and TBC) instruments as reflected by a significant correlation ($r = 0.93$; $p < 0.0001$) between mean scores (Fig. 1).

Both scales showed high internal consistency and construct validity. Internal consistency was determined using Cronbach α revealing reliability coefficients of 0.98 and 0.95 for TBC and GRS, respectively (Table 3). Three items on the TBC instrument were excluded from this analysis due to inadequate data (Items 6e, 7c, and 7d), whereas all 10 items on the GRS were included. Construct validity refers to the degree to which a test can differentiate between an expert and a novice (8). This instrument showed a significant correlation between cumulative otology experience and mean performance scores for both TBC ($r = 0.60$; $p < 0.0001$) and GRS ($r = 0.57$; $p < 0.0001$).

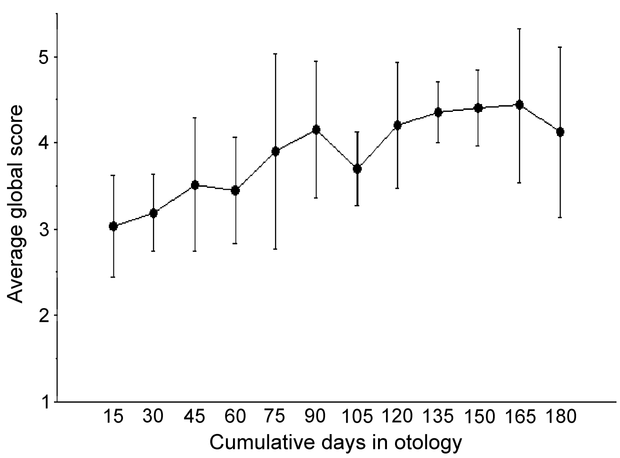

FIG. 3. Average GRS score plotted against cumulative days on the otology service showing improving performance with clinical experience ($r = 0.57$; $p < 0.0001$), highlighting construct validity of this tool.

TABLE 3. *Cronbach α as measure of interitem reliability (internal consistency)*

	Checklist scale	Global scale
No. items used	18	10
Reliability coefficient	0.98	0.95

Figures 2 and 3 show an upward trend in mean performance scores for both scales with increasing time on the otology service. When compared with residents in the first 60 days (3.1 ± 0.7) of otology experience, residents in their second 60 days (3.9 ± 0.7) and third 60 days (4.4 ± 0.8) had significantly higher performance scores on TBC (analysis of variance, $F = 17.1$; $p < 0.0001$). Performance measured using the GRS was similarly different between these levels of otology experience (analysis of variance, $F = 16.6$; $p < 0.0001$). Both measures demonstrated a ceiling effect whereby gains declined after the first 90 days.

"Overall surgical performance" is evaluated on the GRS as a summative assessment of surgical skills by the faculty evaluator. The items determined by univariate analysis to be most strongly correlated to overall surgical performance in descending order were 3a, "open antrum; deepen dissection at sinodural angle" ($r = 0.82$; $p < 0.0001$); 5a, "thin posterior external auditory canal (EAC) cortex; view posterior EAC en face" ($r = 0.80$; $p < 0.0001$); 2c, "defining anatomic limits; define sigmoid sinus and sino-dural angle" ($r = 0.79$; $p < 0.0001$); and 5b, "thin posterior EAC cortex; use side/front appropriate bur" ($r = 0.76$; $p < 0.0001$). Multivariate analysis that included all 4 measures as independent variables confirmed that 5a "thinning posterior EAC" ($p = 0.01$) and 3a "open antrum; deepen dissection at sinodural angle" ($r = 0.43$; $p = 0.02$) were the strongest predictors of overall surgical performance (Table 4).

DISCUSSION

Surgical competence is a multimodal function. Proficiency in the performance of an operative procedure is fundamental to a successful outcome (19). It has been

TABLE 4. *Task-Based Checklist items with strongest correlation to overall surgical performance*

TBC items: strongest predictors of overall surgical performance	Correlaton to overall surgical performance: univariate analysis	Correlaton to overall surgical performance: multivariate analysis
3a, Deepen dissection at sinodural angle	$r = 0.82$; $p < 0.0001$	$r = 0.43$; $p = 0.02$
5a, Thin posterior EAC cortex: view posterior EAC en face	$r = 0.80$; $p < 0.0001$	$r = 0.40$; $p = 0.01$
2c, Define sigmoid sinus and sino-dural angle	$r = 0.79$; $p < 0.0001$	$r = -0.03$; $p = 0.87$
5b, Thin posterior EAC cortex: use side/front appropriate bur	$r = 0.76$; $p < 0.0001$	$r = 0.08$; $p = 0.61$

EAC indicates external auditory cortex; TBC, Task-Based Checklist.

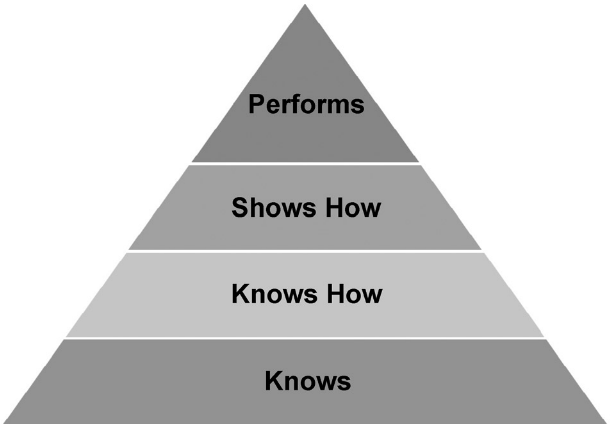

FIG. 4. Miller triangle. A competent surgical resident should be able to attain the shows how (the third) level in Miller triangle (modified from Bhatti N I, Cummings CW: Competency in surgical residency training: defining and raising the bar. *Acad Med* 2007; 82:569–73).

challenging for residency program directors to determine exactly how to go about assessment and documentation of surgical competency (20). Despite the development of a number of tools for assessment of technical skills, none or only a few have been incorporated into standard practice. This can be due to their complexity, poor validity, or lack of feasibility (21). Bhatti and Cummings (22) have suggested that a competent surgical resident should be able to attain the shows how (the third) level in Miller triangle (Fig. 4) during evaluations in real-life situations. Therefore, assessment within the OR is essential to evaluating individual surgeons and residency training programs.

The TBC part of our instrument consists of the critical steps of the mastoidectomy. These tasks were identified by input from the faculty using modified Delphi technique until the consensus was achieved on the items that were to be included in the TBC. Deconstructing the overall surgical performance into a checklist of the different tasks of the procedure allows a more objective and structured evaluation by turning the evaluators into observers rather than interpreters of behavior (2). Residents are explicitly aware of the items on the instrument, thus making the performance goals unambiguous and well defined. Evaluation on a smaller and a well-stated task reduces the potential of the bias by factors such as the relationship between the evaluator and the resident, resident's performance on other steps of the surgery, and competency in areas other than technical competence. Our instrument, although more systematic, quantitative, reproducible, and thorough than other methods of assessments, is still not 100% objective. This is because criteria used for achieving competence are still based on the judgment of the attending otologist. Recently studied more objective evaluation methods include dexterity analysis systems that consist of electromagnetic field generators and sensors that are attached to a surgeon's hand. Data generated by the sensors are used to assess dexterity measures such as number of movements, distance travelled by the hands, and time taken for the task (2). This form of assessment, despite its objectivity, still remains experimental and confined to the laboratory because of the lack of practicality and feasibility. The way forward in the implementation of the mastoidectomy assessment tool is the use of videotapes for the assessment and defining of the benchmarks to make the evaluation completely objective.

To transfer skill assessments from research laboratories onto real cases, there is an obvious need to assess these skills being practiced in real-life situations. An effective skill acquired in a training laboratory or on a simulator can only be validated if duplicated in real-life situations. This is in analogy with the aviation industry where use of simulators to accredit pilots to fly is a globally accepted policy. Although pilots are trained on simulators to avoid critical errors, they are not permitted to fly until they pass a flying test in the air in the presence of an instructor (23). Human beings possess more complicated and diverse anatomic details than airplanes, and therefore, the need to evaluate operative performance in real-life situations is even more critical. Additionally, the variability in comorbidities and disease processes further necessitates assessments in the OR.

For any instrument to be successfully implemented in the OR, it has to be practical. Sidhu et al. (3) define an ideal evaluation instrument to be feasible, comprehensive, and relevant. It is important to address the issue of feasibility before widespread adoption of an assessment mechanism. A feasible assessment should take less time than would be needed to assess the complete operation and should be doable at times convenient for the raters (24). Based on our experience in the OR, we made some modifications in the checklist item. The aim of these modifications was to make the instrument shorter so that compliance can be increased. We noticed that the residents were not being evaluated in items such as "digastric dissection" and "posterior atticotomy" because these steps are not routinely performed in the OR. These items were removed from the instrument for use in the OR. The other modification was sending paper evaluations to the faculty surgeons. Some preferred paper evaluations because they thought logging into their e-mail in the OR to access the evaluation was cumbersome. We noticed that the compliance rate increased from 84 to 100% after making these modifications.

This article describes the results of a 3-year deployment of a skills assessment program in operative otology in our residency program. The skills assessment tools used to measure technical performance in mastoidectomy have been adopted with minimal difficulty by faculty and residents because of the formative and timely feedback to trainees that our protocol affords. Because the skills assessment protocol was built on a foundation of faculty consensus and on ongoing professional development, faculty have found it to be a feasible and empowering addition to their teaching responsibilities, resulting in significant buy-in and compliance. By affording for the first time a quantitative view into skill development in otolaryngology residents, this and similar assessment programs hold the promise of guiding the maturation of surgical skills according to the abilities and needs of individual trainees. Systematic improvements in training goals and methods can be realistically planned, executed, and evaluated with such a program. Assessment tool can serve as a standard which, if adopted widely, could facilitate comparison of different studies on approaches to improve surgical training.

We have demonstrated the reliability and construct validity of the TBC and GRS tools as used by our faculty. These results are likely to reflect the favorable design of these tools. How these tools are used by faculty, however, is also an important determinant of program success. The value of efforts to achieve consensus on the items to be scored, the meaning of different scores, as well as the shared experience of using these tools in temporal bone courses cannot be overlooked (15).

Strong agreement between TBC and GRS is not surprising given that both were designed by consensus of the same group of faculty surgeons. Having worked with residents together for several years and having discussed their didactic needs, these surgeons have a shared view of common weaknesses and critical tasks that speak of surgical competence. Thus, it is expected that certain key positive observations ("thinning posterior EAC" and "open antrum; deepen dissection at sinodural angle") would be strong predictors of overall surgical performance (Table 4). However, the strong agreement between TBC and GRS scores belies their diverse uses in skill assessment. Whereas the TBC provides actionable appraisal of technical skill in mastoidectomy, the GRS offers a more general judgment of cognitive and visual-motor performance necessary for effective surgery. Although the GRS is insufficiently specific to guide technical development, it provides a generic assessment of competence that correlates well with the more specific TBC. We would therefore propose the use of TBC as a frequently used feedback tool and the GRS as a confirmatory evaluation of competency at transitions of professional training.

By deconstructing the mastoidectomy procedure, the TBC plays an important role in providing objective feedback to the resident (24). Assessing performance in each step of the procedure proved useful in focusing specifically on weak areas of a resident's skills for targeted remediation. It is through this formative feedback that the implementation of an assessment program has the most potential to benefit trainees, programs, and society (3). The identification of tasks that strongly correlate with overall surgical performance provides an opportunity to enhance efficiency of time and resources spent on the teaching of these skills. Such observations enable the postulating of testable hypotheses that could have major relevance to educational methods and outcomes.

A potential weakness of our study is the lack of standardization with variability and complexity of the cases. Standardization of the cases in the OR is extremely difficult if not impossible. Second, when residents operate with faculty, the amount of verbal and physical help and direction can affect how well a resident performs. This effect was minimized by adding an instruction sheet to the assessment tool. Third, residents' scores can potentially be affected by faculty bias. Evaluation of a resident by multiple evaluators may have reduced this bias to minimal. In the future, this bias can potentially be eliminated by using video-based assessment of surgical skills, hiding the identity of the resident from evaluators. In addition, long-term trials should be conducted to assess the degree to which "overall surgical performance" correlates with more objective outcomes such as operative times, blood loss, or rate of complications.

This study carries several advantages and potential implications for the assessment of technical skills at mastoidectomy and improvement of the learning curve. Operating room evaluations not only allow appropriate assessment of surgical skills but also provide opportunity for other aspects of a procedure such as communication and reaction to distractions to be rated. All 3 important factors that influence skill learning—contiguity (learning the proper sequence of motor skills), practice, and feedback (20)—are influenced. Assessments such as this tool can play a vital role in helping transfer skill acquisition and evaluation from laboratory to the OR, retaining and consolidating it as a permanent skill.

Our study demonstrates the validity and feasibility of an assessment tool to evaluate residents' mastoidectomy skills in the OR. The TBC scale we used has the potential to identify weak areas of a resident's surgical skills to help provide formative feedback, whereas the GRS may help evaluate adherence to surgical principles. We also demonstrated that feasibility of an instrument can be enhanced by identifying and weighing checklist items that strongly predict the overall surgical performance. To our knowledge, this is the first study that addresses assessment of mastoidectomy skills in the OR. This tool may require further research and refinement, but it already has the potential to help assess residents' skills on a much wider scale.

REFERENCES

1. Accreditation Council for Graduate Medical Education (ACGME) general competencies. Available at: www.acgme.org. Accessed June 10, 2009.
2. Moorthy K, Munz Y, Sarkar SK, et al. Objective assessment of technical skills in surgery. *BMJ* 2003;327:1032–7.
3. Sidhu RS, Grober ED, Musselman LJ, et al. Assessing competency in surgery: where to begin? *Surgery* 2004;135:6–20.
4. Streiner DL. Global rating scales. In: Neufeld VR, Norman GR, eds. *Assessing Clinical Competence*. New York, NY: Springer, 2009: 119–41.
5. Gray JD. Global rating scales in residency education. *Acad Med* 1996;71:55–63.
6. Turnbull J, Gray J, MacFayden J. Improving in-training evaluation programs. *J Gen Int Med* 1998;13:317–23.
7. Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg* 1997;84:273–8.
8. Lin S, Laeeq K, Bhatti NI, et al. Development and pilot testing

of a feasible, reliable, and valid operative competency assessment tool for the endoscopic sinus surgery. *Am J Rhinol* 2009;23:354–9.
9. Syme-Grant J, White PS, McAleer P. Measuring competence in endoscopic sinus surgery. *Surgeon* 2008;6:37–44.
10. Zirkle M, Atplin MA, Anthony R, et al. Objective assessment of temporal bone drilling skills. *Ann Otol Rhinol Laryngol* 2007;116:793–8.
11. Roberson DW, Kentala E, Forbes P. Development and validation of an objective instrument to measure surgical performance at tonsillectomy. *Laryngoscope* 2005;115:2127–37.
12. Arora H, Uribe J, Ralph W, et al. Assessment of construct validity of the endoscopic sinus surgery simulstor. *Arch Otolaryngol Head Neck Surg* 2005;131:217–21.
13. Leung RM, Leung J, Vescan A, et al. Construct validation of a low-fidelity endoscopic sinus surgery simulator. *Am J Rhinol* 2008; 2008:642–8.
14. Zirkle M, Roberson DW, Leuwer R, et al. Using a virtual reality temporal bone simlator to assess otolaryngology trainees. *Laryngoscope* 2007;117:258–63.
15. Laeeq K, Bhatti NI, Carey JP, et al. Pilot testing of an assessment tool for competency in mastoidectomy. *Laryngoscope* 2009;119:2402–10.
16. Winckle CP, Reznick R, Cohen R, et al. Reliability and construct validility of a structured technical skills assessment form. *Am J Surg* 1994;167:423–7.
17. Reznick R, Regehr G, Macrae H, et al. Testing technical skill via an innovative “bench station” examination. *Am J Surg* 1997;173:226–30.
18. Evalue healthcare education solutions. Minneapolis, MN: Advanced Informatics. Available at: www.e-vlue.net. Accessed November 2, 2009.
19. Hall JC, Ellis C, Hamdorf J. Surgeons and cognitive processes. *Br J Surg* 2003;90:10–6.
20. Goff B, Mandel L, Lentz G, et al. Assessment of resident surgical skills: is testing feasible? *Am J Obstet Gynecol* 2005;192:1331–40.
21. Aggarwal R, Grantcharov T, Moorthy K, et al. An evaluation of the feasibility, validity and reliability of laproscopic skills assessment in operating room. *Ann Surg* 2007;245:992–9.
22. Bhatti NI, Cummings CW. Competency in surgical residency training: defining and raising the bar. *Acad Med* 2007;82:569–73.
23. Ball CG, Kirkpatrick AW, Faliciano DV, et al. Surgeons and astronauts: so close yet so far. *Can J Surg* 2008;51:250.
24. Dath D, Regehr G, Birch D, et al. Toward reliable operative assessment. The reliability and feasibility of videotaped assessment of laparoscopic technical skills. *Surg Endosc* 2004;18:1800–4.