The Laryngoscope

Validation of an Objective Assessment Tool for Tonsillectomy in Otolaryngology Resident Training

Luke Jakubowski, MD[†]; Brittany A. Leader, MD[†]; Stacey L. Ishman, MD, MPH; Robert Chun, MD; Sergey S. Tarima, PhD; Sanjay Parikh, MD; Margaret Skinner, MD; Marc Thorne, MD; Robert Weatherly, MD; Gregory Wiet, MD; David J. Brown, MD

Objective: Create and validate an objective structured assessment of technical skills (OSATS) for otolaryngology residents learning how to perform a tonsillectomy.

Study Design: Multicenter prospective longitudinal validation study.

Methods: A multi-institutional study at six tertiary academic otolaryngology residency programs from July 2009 to May 2012. Using the modified Delphi technique, a panel of pediatric otolaryngologists created a tonsillectomy task-based checklist (TBC) for a tonsil OSATS using a 5-point Likert-type scale. Residents were assessed by pediatric otolaryngology staff at the time of surgery with the TBC and a global rating scale. Procedure time, patient age, number of previously performed tonsillectomies, and surgical technique were also collected.

Results: One hundred sixty-seven tonsil OSATS were completed for 38 residents, and competency was recorded for 99 (59.2%). Residents scored as competent had performed significantly more previous tonsillectomies than those deemed noncompetent, 44.4 $\pm$ 35.6 and 13.5 $\pm$ 11.6, respectively ($P < .001$). The mean overall score on the tonsil TBC was 4.0 $\pm$ 0.8 and 2.6 $\pm$ 1.0 for competent and noncompetent, respectively ($P < .001$). Higher number of tonsillectomies performed and mean tonsil TBC score significantly increased the likelihood of competency ($P < .001$). Each additional tonsillectomy performed increased the likelihood of achieving competency by 6.3% ($P = .006$, 95% confidence interval (CI): 1.330–1.110), and each 1.0 point increase in mean tonsil TBC score increased the likelihood of competency by a factor of 2.71 ($P = .006$, 95% CI:1.330–5.513). There is a 95% likelihood of competency at 48 tonsillectomies or a tonsil TBC score of 4.91.

Conclusion: The tonsil OSATS is a valid and feasible instrument to assess resident competency with tonsillectomy and provides timely objective feedback.

Key Words: Xxx.

Level of Evidence: 4.

***Laryngoscope*, 131:E359–E366, 2021**

INTRODUCTION

The training and surgical skills assessment models introduced by Halsted over 100 years ago were based upon subjective evaluation. In the 2001 Outcomes Project, the ACGME (Accreditation Council for Graduate Medical Education) mandated that all residency programs objectively assess resident surgical skills.[1] However, surgical residency programs often fall short of achieving this mandate.[2]

The ACGME developed and implemented the Outcomes Project with an overriding goal to improve resident education through integration of six core competencies and their objective assessment.[1] Since the ACGME mandate, a number of surgical specialties have developed and implemented objective assessment tools to measure competency such as the Objective Structured Assessment of Technical Skills (OSATS). OSATS is a framework developed by Reznick et al., which consists of both a task-based checklist (TBC) and a global rating scale (GRS).[3] An OSATS must be developed for each procedure needing assessment, and although a sound methodology, use of this approach does not in and of itself imply validity or reliability. Having said this, however, OSATS developed for procedures in other surgical fields have demonstrated proven validity, reliability, and feasibility.[4] Unfortunately, otolaryngology-specific OSATS

From the Department of Otolaryngology-Head & Neck Surgery (L.J.), University of Minnesota, Minneapolis, Minnesota, U.S.A.; Department of Otolaryngology-Head & Neck Surgery (B.A.L., S.L.I.), University of Cincinnati College of Medicine, Cincinnati, OH, U.S.A.; Division of Pediatric Otolaryngology (S.L.I.), Cincinnati Children's Hospital Medical Center, Cincinnati, OH, U.S.A.; Division of Pulmonary Medicine (S.L.I.), Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, U.S.A.; Department of Otolaryngology & Communication Sciences (R.C.), Medical College of Wisconsin, Milwaukee, WI, U.S.A.; Division of Biostatistics, Institute for Health and Equity (S.S.T.), Medical College of Wisconsin, Milwaukee, Wisconsin, U.S.A.; Department of Otolaryngology -Head and Neck Surgery (S.P.), University of Washington, Seattle, Washington, U.S.A.; Department of Otolaryngology -Head and Neck Surgery (M.S.), Johns Hopkins School of Medicine, Baltimore, Maryland, U.S.A.; Department of Otolaryngology -Head and Neck Surgery (M.T., D.J.B.), University of Michigan, Ann Arbor, Michigan, U.S.A.; Department of Otolaryngology -Head and Neck Surgery (R.W.), University of Kansas, Kansas City, Kansas, U.S.A.; and theDepartment of Otolaryngology -Head and Neck Surgery (G.W.), Nationwide Children's Hospital and The Ohio State University, Columbus, Ohio, U.S.A.

Editor's Note: This Manuscript was accepted for publication on April 23, 2020.

[†]These authors contributed equally to this work.

The authors have no funding, financial relationships, or conflicts of interest to disclose.

Send correspondence to Stacey Ishman, MD, Cincinnati Children's Hospital Medical Center, 3333 Burnet Ave. MLC# 2018, Cincinnati, OH 45229. E-mail: stacey.ishman@cchmc.org

DOI: 10.1002/lary.28739

are only used among 15.3% of otolaryngology residency programs.[2]

OSATS assessments have been praised for their proven track record; however, many of these studies were performed in a laboratory setting.[5–9] In 2010, van Hove et al. reviewed studies of objective assessment of surgical skills, and only 28 (26.9%) were performed in the operating room.[5] Similarly, although there has been increasing interest in OSATS in otolaryngology, it has been more frequently used in the laboratory setting. The use of simulation and cadaveric models are especially beneficial for learning emergency or more complex procedures such as pediatric airway foreign body removal,[6] transoral robotic surgery,[10] endoscopic sinus surgery,[7,11] or mastoidectomy.[8] However, it is important to develop and validate assessment tools to be used in the intraoperative setting as well. Therefore, single centers have begun creating intraoperative assessment tools.[12,13]

Currently, most otolaryngology training programs evaluate surgical performance with summative feedback provided at the end of the month or rotation, feedback that is intended to reflect performance over a period of time.[2] Limitations of this approach to feedback is that it is based on the memory and recall of the evaluating faculty member and does not provide detailed timely information. As an alternative, frequent formative feedback provides ongoing education during the learning experience, while providing the resident with real-time concrete constructive evaluations. Additionally, this immediate detailed feedback can identify specific deficiencies in a resident's surgical technique, and/or decision making that needs improvement. OSATS serve the dual purpose of providing objective assessment of technical skills, along with detailed and timely formative feedback to residents.

Tonsillectomy is one of the most common surgical procedures performed during otolaryngology residency training. U.S. otolaryngology residents perform an average of 114 tonsillectomies during their training.[14] However, the number of tonsillectomies performed to achieve competency has not been studied. Understanding when competency is objectively achieved could allow residents to redirect their time and efforts toward mastery of more complex procedures or those for which they have not yet achieved competency. Optimizing use of resident learning opportunities is particularly important since the onset of residency work-hour restrictions, which limit clinical exposure and necessitate greater efficiency in graduate medical education.

With the ACGME outcomes mandate in mind, the primary aim of this multi-institutional study was to develop and validate an objective assessment tool for tonsillectomy that could aid in formative feedback and prediction of competency, with the ultimate goal of improving allocation of resident time when learning surgical procedures. We hypothesized that residents who had performed more tonsillectomies were more likely to be competent, and this tool would allow for objective assessment of their skills. To our knowledge, this represents the first multicenter tonsillectomy OSATS study performed in the operating room.

METHODS

Using the modified Delphi technique, an OSATS evaluation tool for tonsillectomy (tonsil OSATS) was developed in 2009 by a panel of pediatric otolaryngologists from six participating institutions: Medical College of Wisconsin (Milwaukee, WI), University of Michigan School of Medicine (Ann Arbor, MI), Johns Hopkins School of Medicine (Baltimore, MD), The Ohio State University College of Medicine (Columbus, OH), University of Kansas Medical School (Kansas City, KS), and University of Washington School of Medicine (Seattle, WA). The Tonsil OSATS consists of a TBC along with a subjective rating of overall tonsillectomy competency, the GRS (Fig. 1).

The panel identified 10 pertinent steps in performing tonsillectomy and incorporated these into the tonsil TBC. Each step was evaluated on a five-point Likert-type scale with an additional "Not Performed" rating. "Not Performed" identified aspects of the procedure not performed by the resident or staff pediatric otolaryngologists and did not affect the overall mean score. At the time of each evaluation (carried out from July 2009 through May 2012), participating residents reported the number of tonsillectomies they had performed; this number was matched to the procedure number in their ACGME resident case log. Residents rotating on the pediatric otolaryngology service at the six participating institutions were eligible to participate.

Many studies regarding surgical education use the postgraduate year (PGY) to determine how advanced a resident should be in their training. Because the period when tonsillectomy is primarily taught can vary by program, we used the absolute number of cases performed as a predictor of competence in this study.

In addition, patient age (in years) and tonsillectomy method were recorded (coblation, electrocautery, or other). Immediately after the procedure, the staff pediatric otolaryngologist completed the tonsil TBC. Scores were based on the case just completed and did not include feedback on concomitant procedures such as adenoidectomy or ear tube placement. In addition to scoring the objective OSATS, the staff pediatric otolaryngologist subjectively evaluated the resident's competency with tonsillectomy GRS, scored as a yes or no, based solely on the case they just completed (Fig. 1). Because there can be challenging cases for any surgeon, the staff pediatric otolaryngologists were also asked to rate the case complexity as a standard case that could be performed by a general otolaryngologist (hereafter referred to as *standard*) or as a "challenging case for most general otolaryngologists" (hereafter referred to as *complex*). The final TBC score is reported as a mean of all 10 task scores with a maximum mean score of 5.0. Tasks rated as "Not Performed" or ratings that were missing were not factored into the mean score. Staff otolaryngologists were given the option of completing the tonsil OSATS on paper or through a password protected Web-based version.

All participating institutions obtained institutional review board approval for the study. All otolaryngology residents on the pediatric otolaryngology service and staff-level pediatric otolaryngology faculty at the six participating institutions were invited to participate in the study. Cases represent a convenience sample by participating staff pediatric otolaryngology faculty.

Inter-rater reliability, which was defined as a measure of agreement between staff pediatric otolaryngologists evaluating the same procedure, was assessed to further validate the tonsil OSATS. Five cases performed at the Medical College of Wisconsin, and the University of Michigan were selected (as a convenience sample) to evaluate inter-rater reliability when visiting faculty were available. For those selected, a tonsillectomy OSATS was completed by one pediatric otolaryngologist familiar with the resident and one visiting pediatric otolaryngologist who was not familiar with the resident, blinded to their PGY status,

Tonsillectomy OSATS Task Specific Rating Assessment

A faculty member is to complete this from after observing a resident perform a tonsillectomy. Participation is voluntary. By completing this form you are implicitly consenting to participate. Any reported or published data from this study will be presented without any identifying markers.

Faculty name:_____ Date:_____

Resident ID: _____ PGY Level: _____

Number of tonsillectomies performed: _____

Age of patient (years) _____

Technique:
- o Electrosurgical (i.e. Bovie)
- o Suction electrosurgical (i.e. suction Bovie)
- o Cold steel
- o Colbation-Total tonsillectomy
- o Colblation-Intracapsular
- o Microdebrider-Total tonsillectomy
- o Microdebrider- Intracapsular
- o Other _____

	Verbal instruction and demonstration	Requires instruction with errors	Independent with errors	Independent without errors	Independent and efficient	Not performed
1. Patient positioning and draping						NA
2. Atraumatic mouth gag placement						NA
3. Gag suspension						NA
4. Adequate exposure of the tonsils						NA
5. Grasps tonsils appropriately						NA
6. Finds the plane						NA
7. Dissects in the correct plane						NA
8. Obtains hemostasis						NA
9. Suctions stomach						NA
10. Removes gag safely while respecting ETT placement						NA

Attending rating of case complexity:
- ☐ Standard case that most general otolaryngologists would feel comfortable performing
- ☐ Challenging case for most general otolaryngologists

This was a challenging case because_____

Based on the overall performance on this case, is the resident capable of independently performing a tonsillectomy in a safe and competent manner?

Attending Response: Yes No

Fig. 1. Tonsillectomy OSATS form used to assess residents to determine whether they were competent on the procedure.OSATS = objective structured assessment of technical skills.

or the number of previous cases performed. These were selected based on the convenience and carried out by two authors (D.J.B. and R.C.).

The mean time for completion of the survey was determined using login and submission time data from the online survey.

Statistical Analysis

Analysis of the predictors of tonsillectomy competency, as well as results by tonsillectomy technique, were adjusted for resident effect using generalized estimating equations (GEE), with the exception of the variables with small cell counts for which the chi-squared test with Yates continuity correction was used.

The adjusted odds ratios are reported for overall subjective GRS ratings of competence for residents evaluated using the OSATS. In addition, Wald confidence intervals were reported from the multiple logistic regression model regarding the number of procedures, tonsillectomy TBC, and tonsillectomy technique fitted with GEE. Comparisons between tonsillectomy procedures performed by inexperienced (< 10) and experienced (> 50) residents were assessed with Chi-square test with Yates's continuity correction. The GEE was used to compare results with the tonsil TBC. Predictive properties of the model were evaluated with a receiver operating characteristic curve and area under the curve. Cohen's kappa was assessed on the small group (n = 5) of patients who underwent parallel assessments. P values < .05 were deemed significant. R version 3.3.2 was used for the analysis (www.r-project.org).

RESULTS

A total of 167 tonsil OSATS were submitted by 14 attendings from six institutions from the assessment of 38 trainees. The number of tonsillectomy procedures, TBC score, patient age, and tonsillectomy technique were summarized by competency for all 167 subjects in Table I. Patient age was reported for 162 evaluations; the mean was 5.7 ± 3.5 years. Difficulty ratings were completed for 163 evaluations: 154 were rated as standard, and nine were rated as complex.

All had complete data for subjective GRS of competency. Reported subjective ratings of competency rated 99 (59.2%) of the residents competent to perform a tonsillectomy independently, and 68 (40.7%) were deemed noncompetent (Table I). Those deemed competent had performed significantly more previous tonsillectomy procedures (44.4 ± 35.6) than those deemed noncompetent (13.5 ± 11.6, $P < .001$). The mean overall score on tonsil TBC was 4.0 ± 0.8 for those deemed competent and 2.6 ± 1.0 for those deemed noncompetent ($P < .001$).

There was no significant difference in patient age between cases rated as competent (5.6 ± 3.3 years) and noncompetent (5.8 ± 3.8 years, $P = .712$). More residents were deemed competent when using the radiofrequency tonsillectomy technique than electrosurgical cautery ($P = .004$) (Table I). This was supported in a further subset analysis of tonsillectomy technique, which revealed residents had performed significantly more tonsillectomies in the radiofrequency technique group (47.8) compared to the electrosurgical technique group (25.7) ($P < .018$).

There was no difference in number of procedures performed, mean TBC score, technique, or age of patient between complex and standard cases. The number of tonsil procedures performed and mean TBC score significantly affected the likelihood of a GRS rating of competency as seen in Table I ($P < .001$ for both). The use of radiofrequency technique was also positively associated with competency ($P < .001$). However, patient age did not significantly correlate with the likelihood of competency.

The adjusted odds ratios to predict competence by procedure number, surgical technique, and mean tonsil TBC score can be found in Table II. Each tonsillectomy procedure performed increased the likelihood of a GRS rating of competency by 6.3% ($P = .006$, 95% CI of 1.018–1.110). Moreover, with a 1.0 point increase in mean score on the tonsil TBC, the likelihood of a rating of

TABLE I.
Predictors of Tonsillectomy Competency.

Variables	Global Rating Scale			
	Total *N = 167*	Noncompetent N = 68	Competent N = 99	*P* Value
Procedure				*< .001*[G]
N	167	68	99	
Mean ± SD	31.8 ± 32.2	13.5 ± 11.6	44.4 ± 35.6	
Median (min–max)	24.0 (0.0–300.0)	10.5 (0.0–60.0)	39.0 (5.0–300.0)	
Tonsil TBC				*< .001*[G]
N	167	68	99	
Mean ± SD	3.5 ± 1.1	2.6 ± 1.0	4.0 ± 0.8	
Median (min–max)	3.7 (0.8–5.0)	2.8 (0.8–4.3)	4.2 (1.3–5.0)	
Age of patient				*.72*[G]
N	162	65	97	
Mean ± SD	5.7 ± 3.5	5.8 ± 3.8	5.6 ± 3.3	
Median (min–max)	5(1–17)	5(1–16)	5(1–17)	
Technique				*.004* [C+]
N	163	67	96	
Electrosurgical	117 (71.8%)	56 (83.6%)	61 (63.5%)	
Radiofrequency technique	41 (25.2%)	8 (11.9%)	33 (34.4%)	
Others	5 (3.1%)	3 (4.5%)	2 (2.1%)	

[C+]Chi-square test with Yates's continuity correction; [G]GEE controlling for resident effect

Procedure: number of tonsillectomies performed; age: age of patient in years; competent: overall rating of competent to perform a tonsillectomy independently; noncompetent: overall rating of not competent to perform a tonsillectomy independently.

SD = standard deviation; TBC = tonsillectomy task-based checklist.

TABLE II.
Adjusted Odds Ratios of Overall Subjective Rating of Competent for Residents Evaluated Using Tonsillectomy Objective Structured Assessment of Technical Skills (n = 163; 4 techniques were missing).

	Adjusted Odds Ratio	95% Wald Confidence Limits	
Procedure number	1.063	1.018	1.110
Tonsil TBC	2.71	1.330	5.513
Technique (radiofrequency vs. others)	2.647	0.610	11.488

For technique, 117 used electrosurgery; 41 used radiofrequency; and five were classified as other. The model included number of procedures and TBC score due to significance, whereas technique was included due to high-face validity.
Procedure: number of tonsillectomies performed; tonsil TBC: TBC mean score.
For procedure and tonsil TBC, odd ratios represents 1-unit increase.
TBC = tonsillectomy task-based checklist.

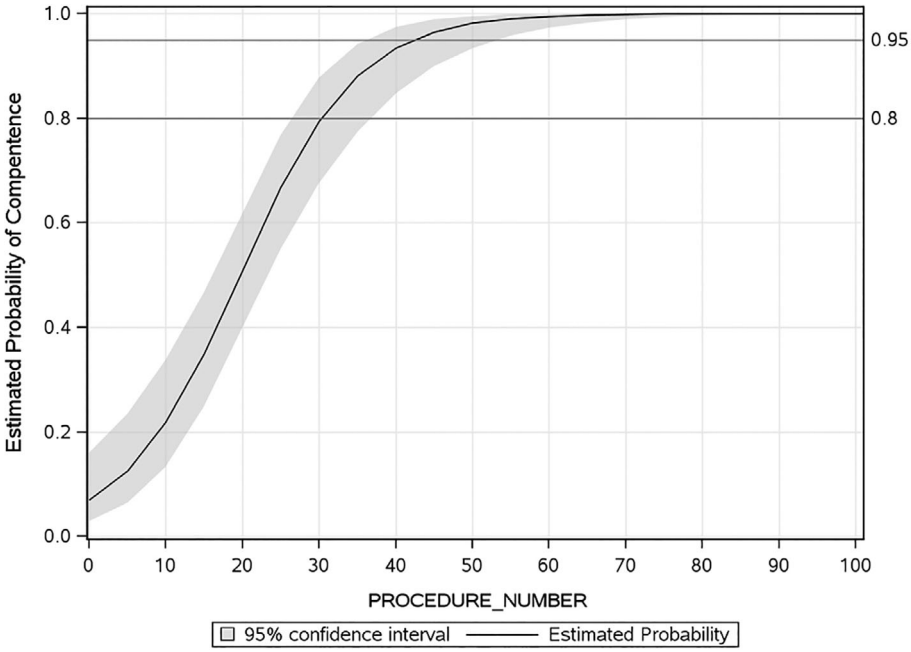

Fig. 2. Likelihood of competency by procedure number for residents performing tonsillectomy. Procedure (number of previous tonsillectomies performed).

competency increased by a factor 2.71 (P = .006, 95% CI of 1.330–5.513. After adjusting for the number of procedures performed and the mean TBC score, the tonsil removal technique was no longer associated with the likelihood of a rating of competency (P = .194, 95% CI of 0.610–11.488).

Using the procedure number to predict competency, there is an 80% likelihood of a subjective rating of competence at 31 standard tonsillectomies and a 95% likelihood at 48 standard tonsillectomies (Fig. 2). Using the mean score on the tonsil TBC, there is an 80% likelihood of a subjective rating of competence at a score of 3.99 and a 95% likelihood at a score of 4.91 (Fig. 3). The receiver operator characteristic (ROC) curve for this model demonstrated an area under the curve (AUC) of 0.9220 (Fig. 4).

To assess construct validity (i.e., increased procedure numbers correlate with an increased likelihood of competency on the GRS), we compared the mean tonsil TBC score and the likelihood of competency for residents performing ≤ l0 cases versus those ≥ 50 previous tonsillectomy cases. For those who had performed fewer than 10 procedures, the mean tonsil TBC score was 2.0 (standard deviation $\pm$ 0.8) compared to 4.3 (SD $\pm$ 0.5) for those who had performed more than 50 procedures (P < .001). The likelihood of being rated not competent for those who had performed fewer than 10 procedures was 88.2% versus 6.9% for those who had performed more than 50 procedures (P < .001) (Table III).

Five cases were evaluated by two otolaryngologists to assess inter-rater reliability. Of the 10 tasks included on the TBC, four had perfect correlation (κ = 1, P = .01); two had substantial agreement (κ = 0.64, P = .02); one had moderate agreement (κ = 0.54, P = .08); two had fair agreement (κ = 0.37, P = .01; κ = 0.28, P = .19); and one had slight agreement (κ = 0.16, P = .35). The GRS of competency also had perfect agreement (κ = 1, P = .01) (Table IV).

The mean time for completion of the survey was 87 $\pm$ 66 seconds. A survey of the faculty (not reported

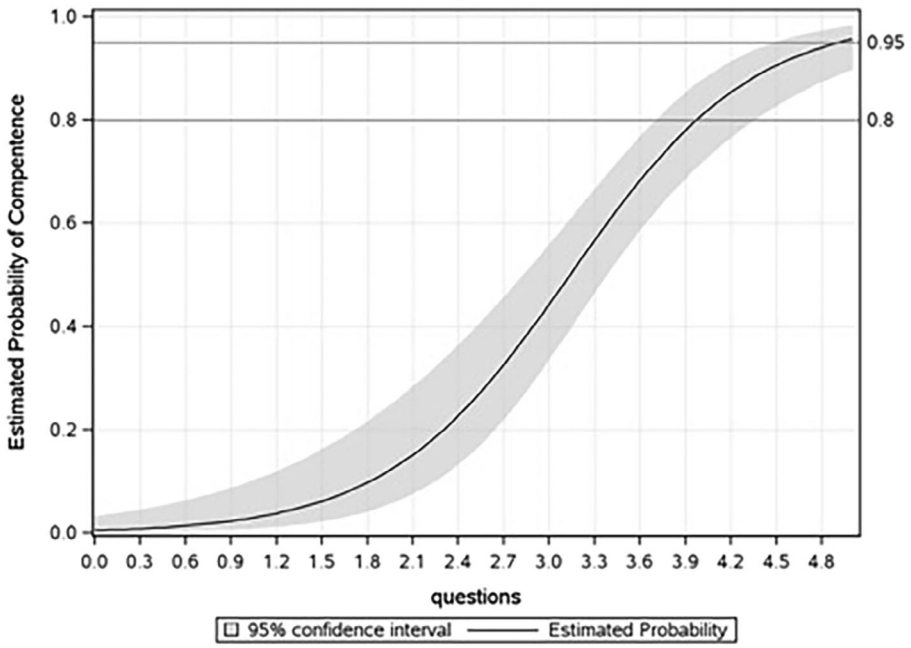

Fig. 3. Likelihood of competency by mean score of tonsil TBC for residents performing tonsillectomy. TBC = task-based checklist.

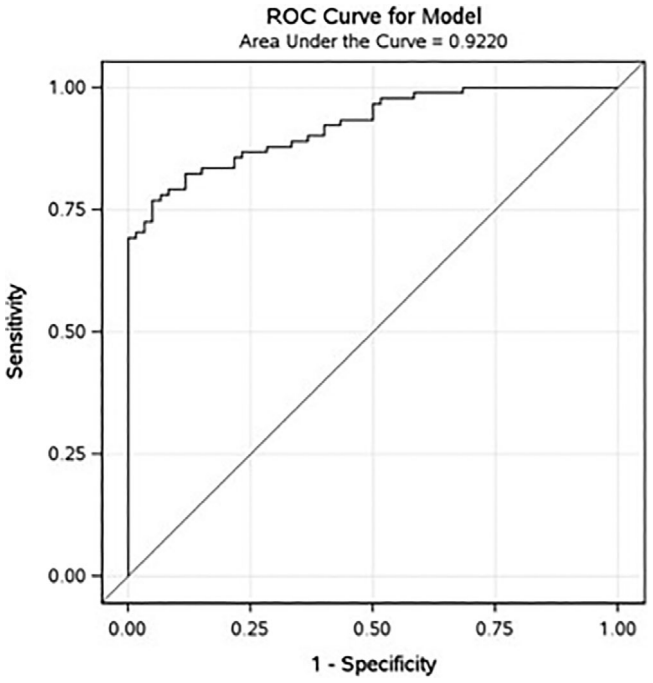

Fig. 4. ROC curve for the tonsil objective structured assessments of technical skills.ROC = receiver operating characteristic.

here) found the instrument to be easy to understand, comprehensive, and practical.

DISCUSSION

As hypothesized, residents who had performed more tonsillectomies were more likely to be assessed as competent. Residents deemed competent had also performed significantly more tonsillectomies prior to assessment than those rated not competent, and higher scores on the tonsil TBC correlated with a greater likelihood of being rated competent. In line with these findings, each tonsillectomy performed resulted in a 6.3% increase in the likelihood of being deemed competent, and each 1.0 increase in the mean tonsil TBC score increased the likelihood of competence by a factor 2.7. We found that there was a 95% likelihood of competency when residents had performed 48 tonsillectomies or had a tonsil TBC score of 4.91. Interrater reliability was perfect to moderate for seven of the 10 TBC procedure steps in a limited assessment of five residents.

These results are similar to previous OSATS studies, which have validated task-based checklists for the

TABLE III.
Mean Tonsil Task-Based Checklist Score and Global Rating Scale Competency Rating for Residents Who Had Previously Performed ≤ l0 Tonsillectomies Versus Those Who Had Performed ≥ 50 Tonsillectomies.

	Procedure Number		
	< 10 N = 34	*> 50 N = 29*	*P Value*
Tonsil TBC			*< .001*[G]
N	34	29	
Mean ± SD	2.0 ± 0.8	4.3 ± 0.5	
Median (min–max)	2.0 (0.8–3.5)	4.5 (3.2–5.0)	
Competence			*< .001* [C+]
No	30 (88.2%)	2 (6.9%)	
Yes	4 (11.8%)	27 (93.1%)	

[C+]Chi-square test with Yates's continuity correction; [G]GEE
Procedure: number of tonsillectomies performed; tonsil TBC: task-based checklist mean score.
SD = standard deviation; TBC = task-based checklist.

TABLE IV.
Interrater Reliability for Each Item Included in the Tonsillectomy Task-Based Checklist and Competency Evaluation (n = 5).

TBC	Kappa Coefficient
Patient positioning and draping	0.642
Atraumatic mouth gag placement	1.000
Gag suspension	1.000
Adequate exposure of the tonsils	0.375
Grasps tonsils appropriately	1.000
Finds the plane	0.642
Dissects in the correct plane	0.285
Obtains hemostasis	0.166
Suctions stomach	0.545
Removes gag safely while respecting ETT placement	1.000
Overall rating of competency	1.000

Tonsil TBC: Tonsillectomy task-based checklist mean score.

assessment of resident surgical skills.[4–11] TBC and GRS have been created for a variety of simulation models for otolaryngology surgical skills, which have similarly demonstrated construct validity and high interrater reliability. These include OSATs for airway foreign body management,[4] endoscopic sinus surgery,[5,6,10] mastoidectomy,[6,11] pediatric airway endoscopy,[7,8] and transoral robotic surgery training.[9] Similarly, OSATS are now being designed for the operating room, with a previous pilot study demonstrating feasibility and validity for a tonsillectomy-specific TBC and GRS.[8] The tonsil OSATS demonstrated high interrater reliability, which is comparable to reported findings for other OSATS such as global assessment of gastrointestinal endoscopic skills, for flexible gastrointestinal endoscopy skills,[12] which was > 0.95; and MISTELS, for laparoscopic surgery, which was 0.998.[13]

Interestingly, the use of the radiofrequency technique correlated with an increased likelihood of competency. This is likely due to the practice of some of the participating pediatric otolaryngology attendings, who required residents to learn tonsillectomy with electrocautery before having them attempt coblation tonsillectomy. This is further supported because the mean number of procedures performed by the radiofrequency technique group was almost twice as many (47.8) as those performed by those in the electrosurgical group (25.7).

Contemporary medical education, first introduced by Flexner, Halsted, and Osler at the beginning of the 20th century, remains the foundation of surgical education. The main tenants of surgical education include graduated responsibility, supervision, and mentorship; these factors remain integral to the training of future surgeons. However, as other professions move toward competency assessments, surgical training has lagged. In light of these factors, societal and professional pressures continue to push surgical education toward objective measures of competency,[1] and objective demonstration of surgical competency will likely be required in the future by regulatory agencies and credentialing authorities.

Additional challenges not envisioned by Flexner, Halsted, and Osler include the introduction of resident work-hour restrictions in the United States in 2003. Work hours were then further restricted in 2011 to limit the number of consecutive hours that residents can work.[1] These work-hour restrictions are particularly challenging for surgical educators because surgical trainees prioritize surgical case exposure in order to achieve competency in a wide array of procedures. For each surgical procedure, it is assumed that the total number performed as resident surgeon correlates with competency. However, the number needed to achieve competency likely varies by procedure and by surgeon.

We hypothesized that understanding the learning curve for a procedure and the ability to objectively assess surgical skills would allow surgical trainees and their supervising surgeon to optimally tailor resident's operative experience. These regular assessments of competence allowed for regular feedback to facilitate more efficient learning. We suggest that once competence is achieved, the resident could spend time working towards proficiency and expert status, as well as allowing residents and their instructors to apportion their time so that they can be sure to achieve competence in all areas of otolaryngology.

Our data suggests that after 48 tonsillectomies, there is a 95% likelihood of competency, well within the current mean of 114 tonsillectomies performed by residents prior to graduation. The performance of tonsillectomies beyond those needed to achieve competency should serve as an opportunity for residents to achieve proficiency or expert status; however, given restricted work hours, surgical educators may use this information to decide that this time might be better spent by residents learning additional procedures. Tonsil OSATS can be used to provide objective feedback and help assess surgical competency in conjunction with the judgement of the supervising surgeon. In addition to addressing the need for objective assessment tools in otolaryngology, this OSATS can be used to improve resident feedback by providing immediate detailed formative feedback.

The main limitation of this study is the sample size because the tonsil OSATS was utilized in only 167 cases. However, testing across multiple centers supports that these results are generalizable, and this tool can be used by residents of any PGY level. Secondly, only five cases were reviewed for interrater reliability; this data is limited (N = 5) due to the difficulty obtaining a second pediatric otolaryngologist who had no knowledge of the resident, PGY, or case number. However, our results demonstrated a high level of agreement between evaluators, which is consistent with interrater reliability findings for other specialty OSATS.[12,13] In addition, despite the limited number of cases, seven of 10 tasks on the tonsil PCS had moderate-to-perfect correlation.

Despite these limitations, we found the tonsil OSATS to be feasible and have construct validity. Future research is warranted across a larger sample size to confirm this level of interrater reliability.

CONCLUSION

Tonsil OSATS provide a valid and feasible measure of otolaryngology resident competence in performing tonsillectomy as well as provide a tool for formative feedback when used in conjunction with the judgement of the supervising surgeon. Tools such as these will be important as regulatory agencies and credentialing authorities start requiring objective demonstration of surgical competence.

BIBLIOGRAPHY

1. Swing SR. The ACGME outcome project: retrospective and prospective. *Med Teach* 2007;29(7):648–654.
2. Brown DJ, Thompson RE, Bhatti NI. Assessment of operative competency in otolaryngology residency: survey of US program directors. *Laryngoscope* 2008;118:1761–1764.
3. Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. Testing technical skill via an innovative "bench station" examination. *Am J Surg* 1997;173:226–230.
4. Goff B, Mandel L, Lentz G, et al. Assessment of resident surgical skills: is testing feasible? *Am J Obstet Gynecol* 2005;192:133–138.
5. van Hove PD, Tuijthof GJ, Verdaasdonk EG, Stassen LP, Dankelman J. Objective assessment of technical surgical skills. *Br J Surg* 2010;97:972–987.
6. Jabbour N, Reihsen T, Payne NR, Finkelstein M, Sweet RM, Sidman JD. Validated assessment tools for pediatric airway endoscopy simulation. *Otolaryngol Head Neck Surg* 2012;147:1131–1135.
7. Laeeq K, Infusino S, Lin SY, et al. Video-based assessment of operative competency in endoscopic sinus surgery. *Am J Rhinol Allergy* 2010;24: 234–237.
8. Laeeq K, Bhatti NI, Carey JP, et al. Pilot testing of an assessment tool for competency in mastoidectomy. *Laryngoscope* 2009;119:2402–2410.
9. Swift SE, Carter JF. Institution and validation of an observed structured assessment of technical skills (OSATS) for obstetrics and gynecology residents and faculty. *Am J Obstet Gynecol* 2006;195:617–621. discussion 621–623.
10. Sobel RH, Blanco R, Ha PK, Califano JA, Kumar R, Richmon JD. Implementation of a comprehensive competency-based transoral robotic surgery training curriculum with ex vivo dissection models. *Head Neck* 2016;38: 1553–1563.
11. Lin SY, Laeeq K, Ishii M, et al. Development and pilot-testing of a feasible, reliable, and valid operative competency assessment tool for endoscopic sinus surgery. *Am J Rhinol Allergy* 2009;23:354–359.
12. Ishman SL, Brown DJ, Boss EF, et al. Development and pilot testing of an operative competency assessment tool for pediatric direct laryngoscopy and rigid bronchoscopy. *Laryngoscope* 2010;120:2294–2300.
13. Ahmed A, Ishman SL, Laeeq K, Bhatti NI. Assessment of improvement of trainee surgical skills in the operating room for tonsillectomy. *Laryngoscope* 2013;123:1639–1644.
14. Accreditation Council for Graduate Medical Education. NATIONAL RESIDENT REPORT (comprehensive and key indicator tables) 2017–2018.