Systematic Review

Toolbox of Assessment Tools of Technical Skills in Otolaryngology–Head and Neck Surgery: A Systematic Review

Mathilde Labbé; Meredith Young, PhD; Lily H. P. Nguyen, MD, MSc, FRCSC

Objective: To support the development of programs of assessment of technical skills in the operating room (OR), we systematically reviewed the literature to identify assessment tools specific to otolaryngology–head and neck surgery (OTL-HNS) core procedures and summarized their characteristics.

Methods: We systematically searched Embase, MEDLINE, PubMed, and Cochrane to identify and report on assessment tools that can be used to assess residents’ technical surgical skills in the operating room for OTL-HNS core procedures.

Results: Of the 736 unique titles retrieved, 16 articles met inclusion criteria, covering 11 different procedures (in otology, rhinology, laryngology, head and neck, and general otolaryngology). The tools were composed of a task-specific checklist and/or global rating scale and were developed in the OR, on human cadavers, or in a simulation setting.

Conclusions: Our study reports on published tools for assessing technical skills for OTL-HNS residents during core procedures conducted in the OR. These assessment tools could facilitate the provision of timely feedback to trainees including specific goals for improvement. However, the paucity of publications suggests little agreement on how to best perform work-based direct-observation assessment for core surgical procedures in OTL-HNS. The sparsity of tools specific to OTL-HNS may become a barrier to a fluid transition to competency-based medical education.

Key Words: Assessment, toolbox, systematic review, competency, medical education, technical skills.

INTRODUCTION

Competency-based medical education (CBME) is in the midst of bringing about major changes to residency training and necessitates the use of defensible methods of assessment to appropriately judge residents’ competency. Judgments of competence are likely best made in the context of programmatic assessment, the careful combination of multiple types of assessment of trainee performance across time. Otolaryngology–head and neck surgery (OTL-HNS) residency programs are encouraged to move beyond written or oral examinations, and to emphasize the inclusion of work-based assessments of residents. Effective and frequent assessment directly in the workplace is a key component to the CBME implementation given the greater need for coaching and formative assessment on a continual basis and the decoupling of development of competence from time in training. Thus, to support judgments of competence, provide feedback, and document the trajectory of a resident toward competence, it is critical to have assessment tools that are of high quality and designed for use in direct-observation workplace-based assessment contexts.

Of the many roles expected of a competent physician, surgical technical skills are well represented in the learning competences expected of surgical residents. Other surgical specialties have extensive literature reporting on several validated assessment tools of technical skills, as well as toolboxes to help surgical educators select the instruments best suited for their context. In a recent large systematic review of all surgical specialties that presented 106 assessment tools, only one of the core procedures for OTL-HNS was represented. To our knowledge, there is no work reporting on the available OTL-HNS assessment tools for evaluating residents’ technical surgical skills in the operating room (OR). Here, we report on a systematic review process that has resulted in a toolbox that could be used by surgical educators to inform the development of programs of assessment to support CBME in OTL-HNS.

From the Faculty of Medicine (M.L.), Center for Medical Education (M.Y., L.H.P.N.), Department of Medicine (M.Y.), and Department of Otolaryngology (L.H.P.N.), McGill University, Montreal, Quebec, Canada

Additional Supporting Information may be found in the online version of this article.

Editor’s Note: This Manuscript was accepted for publication on September 5, 2017.

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

Send correspondence to Lily H. P. Nguyen, Department of Otolaryngology–Head and Neck Surgery, Montreal Children’s Hospital, McGill University, 1001 Boulevard Decarie, A02.3015, H4A 3J1 Montreal, Quebec, Canada. E-mail: phan.nguyen@mcgill.ca

DOI: 10.1002/lary.26943

Laryngoscope 128: July 2018 Labbé et al.: Toolbox for Assessment of Technical Skills

Laryngoscope 128: July 2018
METHODS

Search Strategy

A systematic search of the scientific literature was designed with the support of a medical librarian to identify the literature reporting on tools for assessing technical skills in OTL-HNS. The search strategy was adapted and executed in MEDLINE, PubMed, Embase, and the Cochrane Database of Systematic Reviews (see Supporting Appendix 1 in the online version of this article). The search keywords included a combination of words related to technical skills assessment such as “measurement,” “evaluation,” “competency,” and “proficiency,” as well as specific names of OTL-HNS–related procedures that were compiled by the authors. The complete set of search strategies can be obtained from the authors on request. The search included published articles from inception of the databases (as early as 1946) to 2016 in OTL-HNS.

Inclusion and Exclusion Criteria

Articles in the English or French language, that presented a novel assessment tool focused on assessing the technical skills of a trainee, that could be used in the operating room, and were developed for resident assessment were included. Rater-based assessments, defined as instruments designed for completing during or following direct observation in the OR by a clinical supervisor, were identified. Tools developed to assess performance of a trainee on specific procedures were included, whether the tool was developed and validated in the OR, in a simulation setting, or on cadavers. Assessment tools that were solely based on quantitative metrics (such as time to completion of a task) and publications where the main goal of the study was not to directly report on a new assessment tool were excluded.

Study Selection

Every title was reviewed by two independent authors for inclusion and exclusion following rater training (M.L. and J.L.). Discrepancies were discussed to reach consensus. The same process was subsequently performed for abstracts and full-text articles. Inter-rater reliability was not measured.

Data Extraction

For all articles that met inclusion criteria, title, author(s), year of publication, journal, country where the study was conducted, setting (e.g., OR, simulation, cadaver), sample size, and level of training of the participants was extracted. Information on the tool itself was also extracted, including rating-scale format (e.g., global rating scale [GRS], task-specific checklist [TSC]), number of items, and type of scale (e.g., binary, Likert scale).

RESULTS

Search Results

Results were reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Fig. 1). Seven hundred thirty-six unique articles were identified. Following review of titles and abstracts, 68 full-text articles were reviewed. Sixteen were found to meet the complete list of inclusion criteria and were included in our review.

Description of the Assessment Tools

Characteristics of the 16 assessment tools included in this review are detailed in Table I. Tools reflected a total of 11 different OTL-HNS procedures; four in otology, two in laryngology, two in rhinology, two in head and neck, and one in general otolaryngology. Eleven (68.8%) of the tools were composed of both a TSC and a GRS, three (18.6%) only used a GRS, and two (12.6%) only used a TSC. Five tools (31.3%) had a final component consisting of a global judgment on overall competency, and seven tools (43.8%) had a space for narrative feedback. Fourteen tools (87.5%) used a five-point Likert scale, and two (12.5%) used a binary scale (yes/no or complete/not complete).

DISCUSSION

To our knowledge, we present the first systematic review of direct-observation workplace-based assessment tools for technical skills in OTL-HNS residency training. To support the development of an assessment program within competency-based medical education, we offer a toolbox of 16 instruments covering a total of 11 OTL-HNS procedures that could be used to assess technical skills in the operating room. A task-specific checklist (TSC) component was present in the vast majority of our tools. This form of assessment focuses on directly observable actions and
### TABLE I.
Summary of Published Technical Skills Assessment Tools in Otolaryngology–Head and Neck Surgery Procedures.

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Origin</th>
<th>No. of Items</th>
<th>Evaluation</th>
<th>PGY Level</th>
<th>Setting</th>
<th>Global Judgment</th>
<th>Narrative Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cochlear implant surgery</td>
<td>Australia</td>
<td>TSC 8, GRS 7</td>
<td>5-point Likert scale</td>
<td>Unspecified + staff</td>
<td>Simulator</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct laryngoscopy and rigid bronchoscopy</td>
<td>USA</td>
<td>TSC 12, GRS 10</td>
<td>5-point Likert scale</td>
<td>2–5</td>
<td>OR (humans or animals)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Endoscopic sinus surgery</td>
<td>USA</td>
<td>TSC 21, GRS 10</td>
<td>5-point Likert scale</td>
<td>1–5</td>
<td>Human cadaver</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Canada and Saudi Arabia</td>
<td>TSC 10, GRS 7</td>
<td>5-point Likert scale</td>
<td>1–5 + staff</td>
<td>OR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hemi/thyroidectomy</td>
<td>USA</td>
<td>TSC 18, GRS N/A</td>
<td>5-point Likert scale</td>
<td>2–5</td>
<td>OR</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Myringotomy and tube insertion</td>
<td>UK</td>
<td>GRS 9</td>
<td>5-point Likert scale</td>
<td>Junior, senior, staff</td>
<td>OR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Septoplasty</td>
<td>Saudi Arabia</td>
<td>TSC 8, GRS 7</td>
<td>5-point Likert scale</td>
<td>2–5</td>
<td>OR</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Temporal bone dissection/</td>
<td>USA</td>
<td>TSC 35</td>
<td>Binary</td>
<td>2–5</td>
<td>Human cadaver</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Mastoidectomy</td>
<td>Canada</td>
<td>TSC 16, GRS 4</td>
<td>Binary, 4-point Likert scale</td>
<td>1–5</td>
<td>Cadaver bones</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tonsillectomy</td>
<td>USA</td>
<td>TSC 22, GRS 10</td>
<td>5-point Likert scale</td>
<td>1–5</td>
<td>Human cadaver</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Tracheotomy</td>
<td>Saudi Arabia</td>
<td>TSC 10</td>
<td>5-point Likert scale</td>
<td>1–5</td>
<td>OR</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

GRS = global rating scale; N/A = not available; OR = operating room; PGY = postgraduate year; SHO = senior house officer; SpR4 = specialist registrar 4; TSC = task-specific checklist.
thus provides a defensible mean to assess performance. TSCs are useful in providing timely feedback with specific improvement goals to trainees as well as defining areas of weaknesses that could require remedial activities.14,15 As they provide feedback that is specific to a technical task rather than overall competency, TSCs could be of most value when used in a formative setting, a low stake environment where the goal is to monitor the trainee’s learning. However, given the specificity of TSCs, their use necessitates careful consideration of what is included in a checklist for a given procedure, and the existence of a particular tool for each of the procedures of a surgical subspecialty.

A GRS component was found in a large majority of our tools. An obvious advantage of those nonspecific scales is the acceptability for surgical educators, who would need to become familiar with only one general assessment.16 A GRS offers an interesting opportunity to use a GRS instead of a TSC, the disadvantage of a GRS may be the necessity for more extensive rater training.6,17 The implementation of any new assessment tool should be accompanied by rater training and quality monitoring practices. Lack of rater training may lead to inconsistent ratings of performance, and may leave the assessment scores difficult to interpret.17 A GRS has, however, been reported to be more reliable and better able to differentiate across levels of expertise.18 With proper rater training, a GRS could probably be best situated for high-stakes summative settings where the goal is to evaluate the trainee’s skills at the end of a rotation, and could provide a better assessment of overall surgical competency.

Although some centers are working to develop the procedure-specific assessment tools similar to those compiled here, there is a clear trend in the literature outside of OTL-HNS toward a more centralized way of assessing technical skills of residents. The O-SCORE19,20 and the Zwisch model21,22 are examples of tools that offer what has been considered a more intuitive assessment approach, where the trainee is rated based on the involvement of the attending surgeon during the procedure. This concept of entrustability is directly linked to CBME goals.23 Tools like the O-SCORE or the Zwisch model could overcome known disadvantages of more traditional instruments, such as long delays to completion of ratings24,25 and could palliate concerns that CBME increases the assessment role of a clinical educator to the detriment of the quality of the learning environment.26–28 Further research is needed to evaluate the feasibility of more generic tools such as the O-SCORE in the context of OTL-HNS procedures.

Regardless of the structure of assessment tools identified in our review, we found that out of the 114 OTL-HNS procedures that are considered as core competencies to achieve during residency training,3 only 11 were covered by the assessment tools identified and summarized in this review. This is in contrast to other specialties such as general surgery, where certain procedures already have extensive literature published on assessment of residents’ technical skills.9,10 This paucity of literature may become a barrier to the successful implementation of CBME, and should raise concerns regarding the evidence base of assessment within OTL-HNS.26

The main limitation of our study is that we did not specifically evaluate or incorporate the sources of validity evidence nor the specific purpose (i.e., formative or summative) of the tools reported here. Although several modern validity frameworks have been integrated into the medical education literature (notably Kane via Cook et al.29 and Messick via Downing23), no formal evaluation of validity evidence was included in this review. However, we reported on the context (e.g., OR, simulation) in which each tool was originally developed, which suggests the circumstances where the instrument is best suited for use. Upon implementation of those instruments in a residency curriculum, attention should always be paid to the original reported validity evidence, and we strongly encourage careful quality monitoring for any assessment tool used. Another limitation of this study lies in the fact that we excluded metric-based tools to focus on rater-based assessments. However, given the emphasis on direct observation and workplace-based assessment within CBME, we feel that limiting our review to rater-based assessments was appropriate.

CONCLUSION

We created a toolbox of assessment tools reported for the rater-based assessment of resident’s technical skills for OTL-HNS operative procedures following a systematic review. These tools could be used to create a program of assessment to support formative and summative assessment within OTL-HNS residency training programs. More research is, however, required in OTL-HNS to expand available tools, expand evidence of validity supporting their use across a variety of contexts, and for consideration of adopting a global instrument to assess surgical competency.

Acknowledgments

The authors thank Joshua Lubov, a medical student at McGill University, for his help with data collection.

BIBLIOGRAPHY
