

Assessment of surgical competence for neck dissection: a pilot study

Érika Mercier, MD
 Louis Guertin, MD
 Eric Bissada, MD
 Apostolos Christopoulos, MD
 Marie-Jo Olivier, MD
 Jean-Claude Tabet, MD
 Nathan Yang, MD
 Tareck Ayad, MD

This work was presented at the 72nd Annual meeting of the Canadian Society of Otolaryngology — Head & Neck Surgery in Québec, Que., as a podium presentation. It was a winning submission at the annual Poliquin research competition.

Accepted January 11, 2021

Correspondence to:

T. Ayad
 Centre Hospitalier de l'Université de Montréal
 1051 Sanguinet St
 Montreal, QC H2X 0C1
 tareck.ayad@umontreal.ca

Cite as: *Can J Surg* 2022 March 9; 65(2). doi: 10.1503/cjs.018020

Background: Progressive implementation of the milestone competence-based curriculum has created a need for new objective and validated means to assess resident surgical proficiency. A previous systematic review of the literature by our group has highlighted a shortage of tools assessing surgical competence in oncologic procedures in otolaryngology — head and neck surgery.

Methods: We developed a procedure-specific assessment tool for neck dissection using a modified Delphi method. The 2-part design was modelled on the previously validated Objective Structured Assessment of Technical Skills checklist. The tool was then validated through a 1-year multicentric prospective study in collaboration with the residents and faculty from our academic centre. Additionally, we developed an online survey to assess the acceptability by residents and staff before and after the validation studies.

Results: A total of 29 evaluations were completed throughout the 2016–2017 academic year. Acceptability ranked high for both residents and staff, with a single discrepancy in responses regarding a potential formative as opposed to summative use of the tool. Validation study results showed significantly higher checklist scores among senior residents than junior residents, as well as a significant score progression over time ($p < 0.05$). Trends in scores on the task-specific tool correlated highly to results obtained on a validated global rating scale ($p < 0.05$).

Conclusion: The first tool assessing surgical competence in oncologic otolaryngology — head and neck surgery has been developed and shows promising validity.

Contexte : L'application progressive d'un programme de formation axé sur les compétences par jalons a forcé la mise en place d'une nouvelle façon objective et validée d'évaluer les compétences des résidents en chirurgie. Une précédente revue systématique de la littérature menée par notre groupe avait mis en lumière un manque d'outils pour évaluer les compétences en chirurgie oncologique de la tête et du cou.

Méthodes : Nous avons élaboré un outil d'évaluation spécifique pour la dissection du cou à l'aide d'une méthode Delphi modifiée. Le modèle en 2 parties s'inspire de la liste de vérification validée existante pour l'évaluation objective structurée des habiletés techniques (Objective Structured Assessment of Technical Skills). L'outil a ensuite été validé dans le cadre d'une étude multicentrique prospective d'une durée de 1 an, en collaboration avec les résidents et professeurs de notre centre universitaire. Nous avons aussi préparé un questionnaire en ligne pour en évaluer l'acceptabilité auprès des résidents et des professeurs avant et après les études de validation.

Résultats : En tout, 29 évaluations ont été complétées au cours de l'année de formation 2016–2017. L'acceptabilité a reçu un score élevé auprès des résidents et des professeurs, avec une seule disparité dans les réponses concernant une utilisation potentiellement formative plutôt que sommative de l'outil. Les résultats des études de validation ont montré des scores significativement plus élevés pour la liste de vérification parmi les résidents séniors par rapport aux résidents juniors, ainsi qu'une importante progression des scores avec le temps ($p < 0,05$). Les tendances des scores pour ce qui est de l'outil spécifique à la tâche ont été en étroite corrélation avec les résultats obtenus sur une échelle d'évaluation globale validée ($p < 0,05$).

Conclusion : Un premier outil d'évaluation des compétences en chirurgie oncologique de la tête et du cou a été créé et sa validité est prometteuse.

The future of Canadian otolaryngology — head and neck surgery (OTL-HNS) is shaped by the practices of the hundreds of surgeons throughout the country, who are in turn a product of the training they receive during residency. Adapting teaching methods to the challenges of modern-day medicine and ensuring graduates achieve a consistent and dependable level of competence is essential to maintaining the highest standards of care and expertise. In keeping with these values, the Competence by Design (CBD) curriculum was developed to ensure essential core competencies were consistently achieved, while tailoring residency to individual needs and progress.¹ This improved syllabus is gradually being implemented throughout Canadian residency programs, with OTL-HNS in the forefront as one of the first surgical specialties to bring it into effect.² These changes spawned a need for reliable and reproducible means of assessing the achievement of milestone competencies, particularly in the case of surgical skills where the norm has long consisted of informal and mostly subjective verbal case-by-case feedback.³ However, this approach lacks in structure, objectivity and traceability and hinders residents' ability to adequately pinpoint strengths and weaknesses in a systematized fashion, potentially limiting progress.

In the face of these new challenges and requirements, our group previously conducted a systematic review of existing literature to identify available methods of assessing surgical competence among OTL-HNS residents.⁴ We concluded that the most used, validated and applicable tool consists of a double checklist inspired by the Objective Structured Assessment of Technical Skills (OSATS) tool. The OSATS tool combines a global rating scale (GRS) and a task specific checklist (TSC), a model validated repeatedly through multiple studies.⁵⁻⁸ It has served as a template for the development of numerous contemporary instruments evaluating surgical competence.⁹⁻¹² Another point highlighted in the review was the absence of assessment tools developed to date for oncologic procedures, including neck dissection. As the CBD initiative is gradually implemented, it will remain important to develop a variety of validated tools to assess resident competence.

The aim of the present study was to contribute a means of assessing surgical competence for neck dissection that is readily applicable in an academic setting, that provides objective feedback to enable resident progression and provides a standardized and objective means of measuring competence. Specifically, the objectives of this study were to develop a TSC for neck dissection, then assess its validity, acceptability and feasibility.

METHODS

Developing an assessment tool for neck dissection

Two complementary checklists assessing different sets of surgical skills were created and combined to allow a comprehensive representation of surgical competence.

To build the GRS, the original questions and scoring system from the OSATS tool were retained (Figure 1). This checklist lists 7 criteria assessing residents' abilities to perform basic surgical tasks and is scored from 1 to 5 on a Likert scale, in increasing order of proficiency. Descriptive pointers are present as seen on Figure 1. To facilitate the use of this tool in a primarily French-speaking academic setting, the items were translated from English to French by a certified private translation service. To ensure clarity and preservation of meaning, the translated version was reviewed by the members of the oncology team of our centre until a consensus was reached on the formulation and interpretation. The final French version of the GRS tool is presented in Appendix 1, available at www.canjsurg.ca/lookup/doi/10.1503/cjs.018020/tab-related-content.

The TSC enumerates the fundamental steps to performing a standard neck dissection. We used a modified Delphi method to define the core elements of this surgery and produce a list of steps, which was reviewed and discussed among the 6 head and neck surgeons on service until a final consensus was reached. A total of 3 rounds were performed during which 41, 32 and finally 24 items were sequentially discussed. This list was developed in French, as it is the primary language used by the majority of the team. Dissection of levels IAB, IIAB, III and IV were selected as they are the most often addressed at our centre. The items were scored with a Likert-type scale ranging from 1 to 5. Descriptive pointers were developed to describe the skill level expected for each score in order to help standardize evaluator assessments. A column marked N/D (unavailable) was also added to score steps that were not performed by the resident. The final TSC tool is presented in Figure 2, and the French version can be found in Appendix 2, available at canjsurg.ca/lookup/doi/10.1503/cjs.018020/tab-related-content.

The tool was completed with spaces to insert the resident's attributed identification number, the resident's level, the name of the assessor, the date of surgery, the date of completion of the questionnaire, the levels addressed, previous neck radiation (yes/no), case difficulty (standard/difficult) and the time to complete the questionnaire. To maintain participant confidentiality, the authors randomly attributed identification numbers to each resident. A copy of the list of resident numbers was emailed to staff, but was undisclosed to residents.

Resident identification #: _____ Level of residency: _____
 Assessor(s): _____
 Date of surgery: _____
 Number of days between surgery and completion of this tool: _____
 Levels of neck dissection (IAB, IIAB, III, IV, V): _____
 Time to complete assessment tool: _____
 Neck radiation? Yes No Difficulty of case: Standard Difficult

Global Rating Scale (GRS)

Respect for Tissue					
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	
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	
Instrument handling					
1	2	3	4	5	N/a
Repeatedly makes tentative or awkward movements		Competent use of instruments but occasionally stiff or awkward		Fluid moves with instruments and no awkwardness	
Use of Assistants					
1	2	3	4	5	n/a
Consistently placed assistants poorly or failed to use assistants		Appropriate use of assistants most of the time		Strategically used assistants to the best advantage at all time	
Knowledge of specific procedure					
1	2	3	4	5	n/a
Deficient knowledge and needed instructions at most steps		Knew all important steps of operation		Demonstrated familiarity with all aspects of operation	
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	
Knowledge of instruments					
1	2	3	4	5	n/a
Frequently asked for wrong instrument or used inappropriate instrument		Knew names of most instruments and used appropriate instrument		Obviously familiar with the instruments and their names	
Did you provide formative feedback to the resident?				Yes	No
Do you think the resident is competent to perform the surgery independently?				Yes	No

Fig. 1. Global rating scale.

Validating the neck dissection competence assessment tool

Validity was assessed by means of a prospective multicentric validation study, involving the residents and staff of the Université de Montréal OTL-HNS program. Members of the oncology staff were asked to use the tool to evaluate surgical skill and competence of participating residents completing a neck dissection. A multicentric ethics committee approval was obtained at the Centre Hospitalier de l'Université de Montréal and Hôpital Maisonneuve-Rosemont (MP-02-2016-6252).

Participants and timeframe

Participants included Université de Montréal OTL-HNS residents in postgraduate years (PGY) 2–5. First-year resi-

dents were excluded from this study because of their limited involvement in neck dissection cases. All eligible residents were invited to be part of this project and received details about the study and the implications of their participation. All 6 head and neck program staff collaborated on the project. Detailed instructions on the use of the checklist and the grading system were provided to the assessors. For example, the supervising staff gave a score of 0 to a specific step when they had to perform the step in lieu of the resident. Therefore, a higher level of intervention by the assessor would translate to lower grades. The timeframe set to complete this study was the 2016–2017 academic year (July 2016 to June 2017).

Validity studies

Three different comparison groups were created in order to assess validity. To evaluate construct validity, a first

Task Specific Checklist (TSC) for Neck Dissection

	Unable to perform step	Performs step with much help	Performs step with some help	Performs step with minimal help	Performs step easily and with fluidity	
Positioning of patient	1	2	3	4	5	n/a
Positioning of skin incision	1	2	3	4	5	n/a
Incision of skin and platysma	1	2	3	4	5	n/a
Elevation of subplatysmal flap	1	2	3	4	5	n/a
Delimitation of level IA	1	2	3	4	5	n/a
Identification and preservation of marginal mandibular nerve	1	2	3	4	5	n/a
Identification +/- preservation of facial vessels	1	2	3	4	5	n/a
Inclusion of adipose tissue and lymph nodes up to the mandibular margin	1	2	3	4	5	n/a
Expose and elevate mylohyoid muscle	1	2	3	4	5	n/a
Identification and freeing of lingual nerve	1	2	3	4	5	n/a
Dissection of the deep plane of the submandibular gland	1	2	3	4	5	n/a
Unwrapping fascia from SCM muscle	1	2	3	4	5	n/a
Identification and dissection of accessory nerve (XI) up to digastric muscle	1	2	3	4	5	n/a
Dissection and delimitation of digastric muscle	1	2	3	4	5	n/a
Skeletonize accessory nerve (XI)	1	2	3	4	5	n/a
Dissection of the retrospinal level (IIB)	1	2	3	4	5	n/a
Dissection of the angle between the internal jugular vein and the accessory nerve (XI)	1	2	3	4	5	n/a
Identification of branches of the deep cervical plexus	1	2	3	4	5	n/a
Resection of the specimen from the deep plane	1	2	3	4	5	n/a
Delimitation of the inferior margin of the dissection	1	2	3	4	5	n/a
Unwrapping fascia from internal jugular vein	1	2	3	4	5	n/a
Identification and preservation of the vagus nerve (X)	1	2	3	4	5	n/a
Dissection of the thyro-lingual-facial trunk	1	2	3	4	5	n/a
Delimitation of the medial margin of the dissection	1	2	3	4	5	n/a

Fig. 2. Task Specific Checklist.

group dichotomized residents into junior (PGY2–3) and senior levels (PGY4–5) then compared score frequencies between them. Comparing score frequencies rather than overall score averages allowed for more granularity in understanding score distribution patterns within each group. In a second group, average scores on both the GRS and TSC were plotted and trended by level of residency in order to assess criterion validity. Finally, a third group compared residents' score frequencies during their first and third evaluations in order to determine individual progression over time.

Cumulated data were analyzed by a certified statistician affiliated with the research centre of the Université de Mon-

tréal hospital (CRCHUM). Data analysis was carried out using SPSS. We performed χ^2 tests to compare score frequency between junior and senior residents on the GRS and TSC tools as well as to assess resident score progression between their first and third evaluations. Average resident scores by level of residency were compared using the Student *t* test. Statistical significance was set at $p < 0.05$ for all tests.

Acceptability studies

Two online surveys were developed to explore resident and staff perception of the tool, including its impact on resident performance, the potential impact on staff

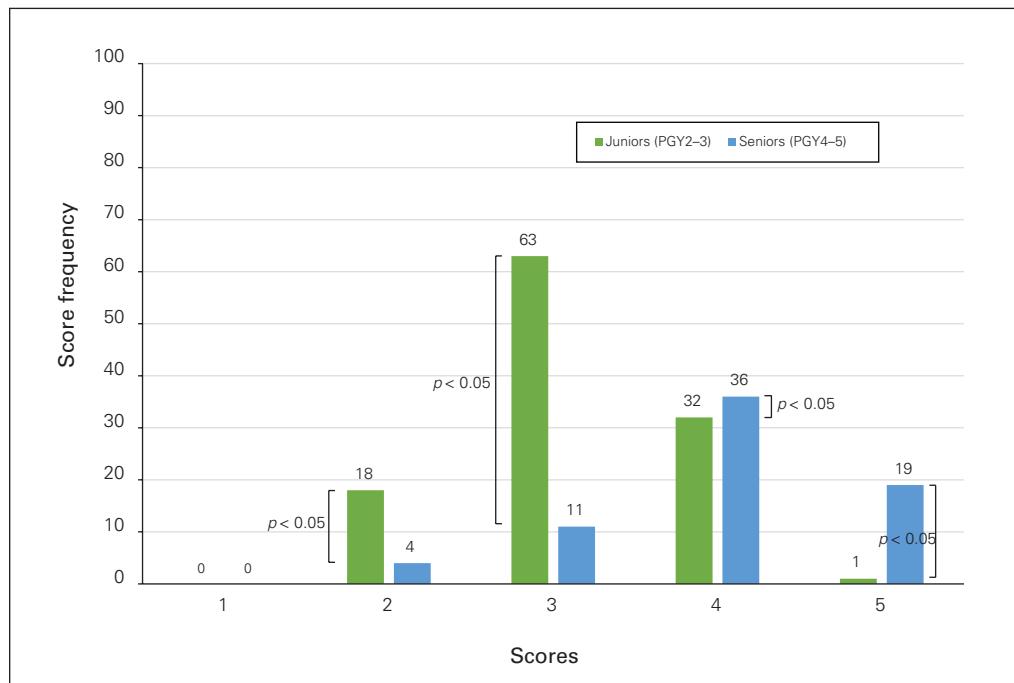


Fig. 3. Score distributions for junior and senior residents on the Global Rating Scale.

feedback and representation of competence. A Likert-style rating scale ranging from 1 to 10 allowed participants to score their level of agreement with each question, where 1 corresponded to “statement very incorrect” and 10 corresponded to “statement very accurate.” A score of 5 was considered to represent a neutral answer. The questionnaire was submitted to all participating residents at the beginning of the study, then again to residents and staff at the end of the study period. Mann–Whitney U tests were done to compare score changes, with $p < 0.05$ representing statistical significance.

Other steps used to assess acceptability included measuring the time required to complete the checklist, the number of days between the surgery and the completion of the evaluation tool, and the amount of feedback offered to residents.

RESULTS

Descriptive data

A total of 29 neck dissections were performed by 11 residents, 6 junior (PGY2–3) and 5 senior (PGY4–5), over the course of the 1-year study period. A slight predominance (62.1%) of cases were completed by junior residents. All residents were evaluated on the same total number of items on both the GRS and TSC, receiving a score of 0 on items they could not perform or for which they required assistance from an attending surgeon. The number of levels addressed during surgery varied; however, an average of 3 levels were dissected per case, with levels IAB, IIAB, IIA, III and IV being addressed simultaneously

in more than half of the surgeries included in this study. A total of 4 patients (13.8%) had previously received radiation therapy before surgery, and the same number of cases were marked as difficult by assessors.

Validity studies

Construct validity: score progression with seniority

Resident scores were dichotomized into junior (PGY2–3) and senior (PGY4–5) resident groups. The frequency at which each score (1 to 5) was obtained by the junior group on the GRS and TSC checklists was counted and compared with the same results for senior residents (Figure 3 and Figure 4). For both checklists, a score increase with seniority could be observed, with junior residents obtaining mainly scores of 3 and 4, and senior residents obtaining mainly scores of 4 and 5. The difference in score frequency was significantly different on scores 2 ($p = 0.041$), 3 ($p < 0.001$), 4 ($p = 0.0014$) and 5 ($p < 0.001$) for the GRS tool and on scores 2 ($p < 0.001$), 3 ($p < 0.001$), 4 ($p = 0.0037$) and 5 ($p < 0.001$) for the TSC tool. The difference was not statistically significant on score 1 for the TSC tool ($p = 0.62$). These results suggest good construct validity, in that the tool could distinguish between junior and senior residents.

Criterion validity: comparing results on TSC and GRS checklists

The average scores by level of residency were compared. Significant differences were found between the scores of PGY2 and PGY5, between PGY3 and PGY5 and between PGY4 and PGY5 residents on both the GRS

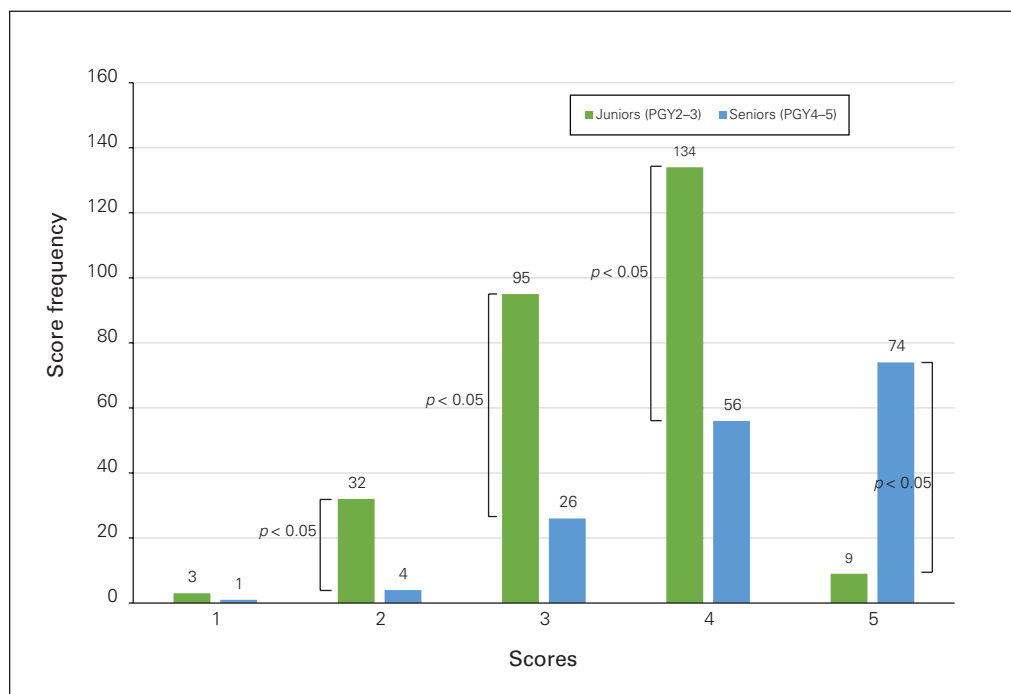


Fig. 4. Score distributions for junior and senior residents on the Task Specific Checklist.

and TSC (all $p < 0.001$). The average scores on the GRS and TSC were also plotted and compared by residency level. The resulting curves follow the same trend closely, and a paired Student t test showed no significant difference between the average scores on each checklist for the same level of residency. The results obtained with the TSC mirrored those obtained with the GRS; the latter being a recognized and validated tool, this implies good criterion validity for the novel tool ($p = 0.17$), hence implying there is no significant difference between both sets of data (Figure 5).

Score progression in time

Resident score progression was assessed by comparing scores obtained on their first evaluation with those obtained on their third evaluation, approximately 1 month later. For this study, junior and senior residents' results were pooled and score frequencies were compared over time (Figure 6). A clear progression can be observed with residents scoring mostly 2, 3 and 4 on their first evaluation and quickly progressing to scores of 3, 4 and 5 on their third assessment. Differences in score frequency were statistically significant for scores 2 ($p < 0.001$), 3 ($p = 0.0014$) and 5 ($p < 0.001$). Differences were not found to be significant for scores 1 ($p = 0.39$) and 4 ($p = 0.073$).

Autonomy

Residents were judged able to perform the procedure independently in 24.1% of cases, all of which were performed by PGY5 residents.

Acceptability studies

The questions and results obtained on the online survey sent to residents at the beginning and at the end of this study, as well as to attendings at the end of the study, are presented in Table 1. No significant differences in responses were identified for residents when comparing pre- and poststudy responses. We also compared post-study resident responses with staff responses. Residents and staff differed in opinion as to the purpose of the tool, with residents being in favour of a formative use, and attending surgeons preferring a summative use. Residents expressed disagreement with the statement that suggested using checklist tools for formal assessments, whereas attending surgeons were in favour of this ($p = 0.042$). There was no statistically significant differences on other questions.

Feasibility

Finally, the mean number of days between the surgery and completion of the checklists was 2.43 days, and the average time to fill out the tool was 4 minutes and 44 seconds. Feedback was given to residents for 96.6% (28 of 29 evaluations).

DISCUSSION

From a prior review of the literature, we found that the combined checklist tool was the most frequent method used to assess surgical competence for OTL-HNS

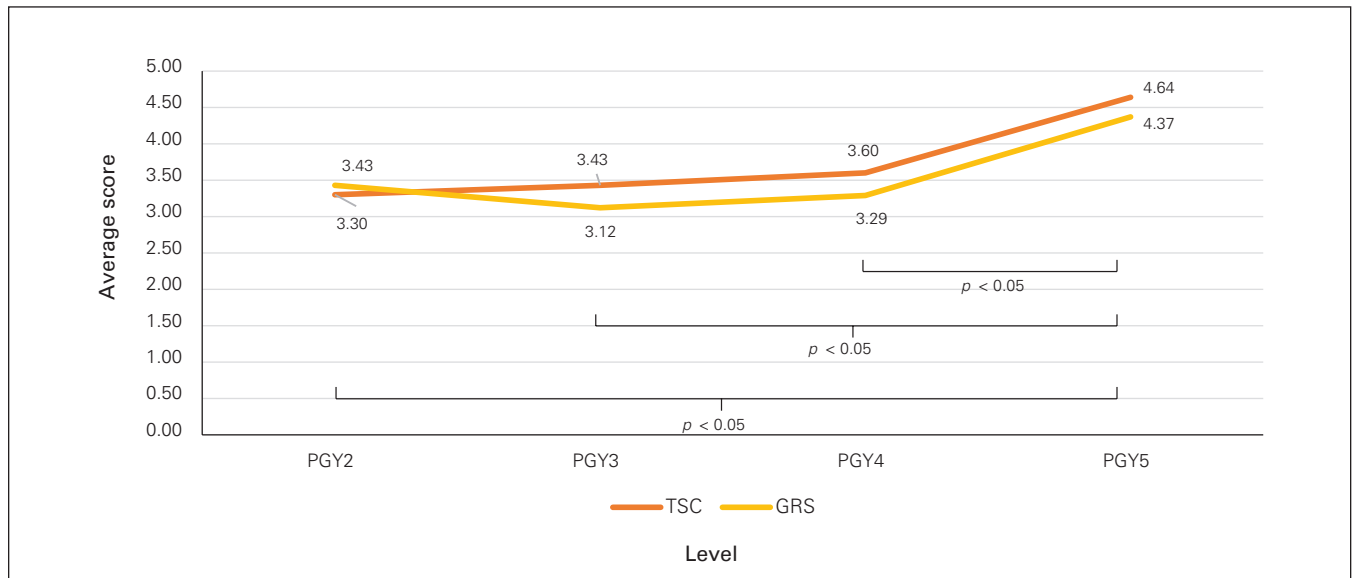


Fig. 5. Average scores on the Global Rating Scale (GRS) and Task Specific Checklist (TSC) by level of residency.

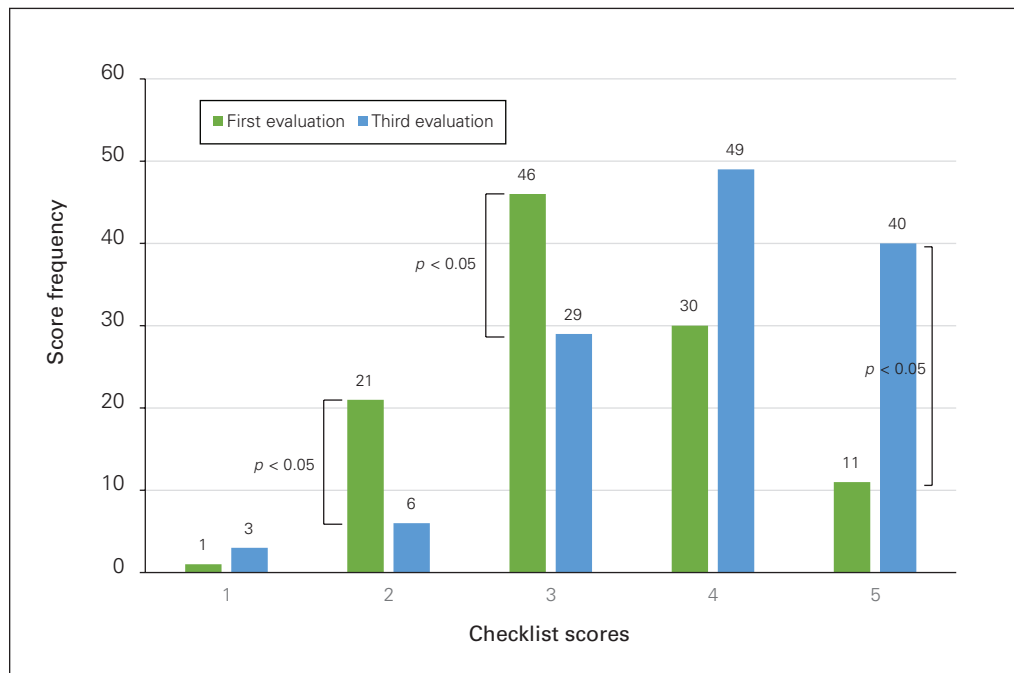


Fig. 6. Score distributions on first and third evaluations.

procedures. These OSATS-based instruments being easy to implement, low in cost and overall simple to use,⁷ we followed a similar methodology to develop the present tool for assessment of surgical competence in neck dissection.

One of the main goals of the present study was to create a tool that was easy to use and implement in a busy service, that would increase and improve the quality of surgical feedback residents receive, and that would allow an objective means for staff and program directors to keep track of resident progression and to offer focused assistance if required. Our results suggest that resident

progression can be measured in a reliable and standardized fashion using our tool. Our results show that it has the sensitivity to demonstrate significant progression in score results with increasing resident seniority, hence demonstrating the tool's construct validity. It must be noted, however, that resident scores did not appear to progress in a linear fashion with increasing level of training as would be expected. Figure 5 demonstrates that scores stagnate during PGY2–4, then improve significantly in PGY5. It is possible that there is some discreet level of progression that was not appreciated in this study as a result of the limited number of assessments.

Table 1. Average scores on the acceptability questionnaire

Survey questions	Residents			Attending surgeons	
	Pre-study	Post-study	<i>p</i> value*	Post-study	<i>p</i> value†
Q1: Will being evaluated increase your anxiety?	3.40	2.88	0.53	4.00	0.37
Q2: Will being evaluated negatively impact your performance?	2.65	2.38	0.82	1.75	0.28
Q3: Will there be a favorable impact on your global surgical skills?	6.80	6.25	0.47	7.50	0.15
Q4: Will there be a favorable impact on your skills for neck dissection?	7.60	7.25	0.86	7.00	> 0.99
Q5: You wish for these kinds of evaluations to remain formative?	7.15	7.75	0.60	4.75	0.15
Q6: You wish for these kinds of evaluations to be used in summative assessments?	4.70	4.00	0.43	7.25	0.042
Q7: You believe these kinds of evaluations are a good way to assess surgical skills in OTL-HNS?	7.25	7.50	0.71	9.00	0.15
Q8: You believe these evaluations will represent your level of competence?	6.45	7.38	0.17	9.00	0.07
Q9: You believe these evaluations will increase the amount of feedback you receive?	8.55	8.63	> 0.99	9.00	0.68
Q10: You believe these evaluations will help you identify your strengths and weaknesses?	8.70	7.75	0.24	8.25	0.68
Q11: You believe these evaluations will allow better reproducibility in your evaluations?	7.65	6.50	0.22	7.75	0.37
Q12: You believe these evaluations are applicable in residency?	7.90	7.50	0.78	9.25	0.15

OTL-HNS = otolaryngology – head and neck surgery.
 *Comparison of residents' pre- and post-study scores.
 †Comparison of resident and attending scores post-study.

Another explanation could be that resident exposure to neck dissection in our program is gradual, and that the steps performed are tailored to specific residency levels. Neck dissections are part of the PGY5 surgical objectives, and residents are given more frequent opportunities to participate in these cases; repetition in their last year may help them improve more quickly. It is also possible that surgical competence for neck dissection is acquired in an exponential fashion as experience and dexterity build.

In this study, the checklist tool was able to demonstrate rapid and significant resident score progression over a relatively short period of time. On their third assessment, residents had notably improved their scores, with the new tool demonstrating the ability to discern this progress. One potential reason for this rapid progress could be that with the detailed feedback residents perceived an incentive to improve their scores and sought out the necessary guidance to improve on any areas of weakness. Furthermore, the results on the TSC checklist correlate with the results obtained on the previously validated GRS checklist, supporting the accuracy of the newly developed TSC tool. These results lead us to believe that the neck dissection tool can be used reliably to assess resident progression, to help tailor feedback and assist in achieving competence.

Another objective of our study was to ensure that the tool displayed reasonable acceptability in our academic centres. We believe this to be the case, as anonymous surveys to residents and staff have shown that residents did not have an apprehensive attitude toward the evaluations and did not feel that use of the tool would have a negative impact on their performance. Both residents and staff felt that the tool would help increase feedback and help identify strengths and weaknesses. Resident responses remained fairly consistent between the pre-

and poststudy questionnaires. There was a small post-study decrease in scores on the questions addressing the ability of the tool to identify residents' strengths and weaknesses, on the reproducibility of the evaluations and on tool applicability throughout residency (questions 10–12). Although not statistically significant, this could be a result of users having a more nuanced opinion about the tool as they became aware of potential limitations, such as interrater variability, and adapted to the inconvenience of an added postoperative task. The only divergence in opinion between residents and staff involved the intended use of the tool, with staff preferring a summative, more formal use and residents favouring an exclusively formative use of this tool. Feasibility was assessed based on the average time to complete the tool, which was less than 5 minutes, and the percentage of feedback given to the resident by the attending, which was 96.6%. Use of the tool between operative cases is therefore feasible without adding an unreasonable burden on the attending surgeon. Most evaluations were completed within the first 2 days after the surgery, which leads to better internal consistency within results, as demonstrated by Ahmed and colleagues in a similar study.¹⁴

Feedback rates were excellent in our study, with residents receiving one-on-one time to discuss the case and receive individualized guidance on almost all cases. We believe this protected time to be of invaluable help to residents, who benefit from this insight and time to reflect on the case.

Limitations

There are some inherent limitations to this study. First, the small number of evaluations gathered and studied

might have provided insufficient statistical power on some of the analyses performed, such as resident progression scores. Additionally, the large number of assessors relative to the number of evaluations may have increased interrater variability and induced bias. Another limitation is the possibility of confounding factors affecting resident progression outside of them performing neck dissections. Residents may have different surgical exposure over the course of their residency, depending on specific rotations, hospitals and case availability. This might affect their surgical progression and disproportionally affect their scores when performing neck dissections.

Other sources of bias include the fact that throughout the study, attendings were not blinded to the identity of the participating residents, which might induce a bias related to preconceived ideas and impressions of the individual being evaluated, also known as the halo or horn effect. There is also the possibility of encountering the Hawthorne effect, where the behaviour of the participant is altered by their awareness of being observed. Additionally, different attending surgeons might offer different assistance or guidance through the cases, depending on their personal approach, despite our attempt to standardize their implication through detailed instructions. These limitations could have been prevented by having the procedures filmed and then scored anonymously afterwards, blinding the assessor. We considered the benefits and drawbacks of blind assessment with videotaping and opted against it for the following reasons. First, neck dissection is best assessed with 3-dimensional vision of the surgical field. Readily accessible videotaping devices in our institution do not offer this option. Moreover, a recent study has shown that there was a high interrater reliability among blinded (use of video recordings for assessment) and nonblinded (direct observation in the operating room), which facilitates the evaluation of residents and avoids the complexity of arranging video recordings and blinding raters.¹⁵ Hence, we felt that despite the potential imperfections of the direct observation method, it best reproduces the actual conditions in which the TSC for neck dissection will be used, and thus better reflects reality.

Delays between the surgery and the evaluation can induce a recall bias, where the surgeon might not remember some details of the surgery as clearly with passing time. We encouraged surgeons to complete the assessment as quickly as possible to prevent this.

To allow a broader use of the tool, a second similar study was subsequently performed to validate the tool in English (Figure 1 and Figure 2). Additional studies will be necessary to assess inter- and intra-rater reliability and to gain long-term insight on the use of the tool.

CONCLUSION

Within the scope of this study, a new means of assessing surgical competence for neck dissection was developed and validated within the limitations inherent to a small cohort with a small number of evaluations. The tool showed good construct validity, and was found to have good acceptability by residents and staff and to be implementable in an academic setting. Favouring the regular use of these types of tools for educational purposes promotes resident progression by encouraging case discussion and stimulating consistent individualised support. This is in line with the objectives of the CBD curriculum, where programs are tailored to individual needs and progression. Opening the discussion and encouraging a dialogue between mentors and learners could prove to be an important step in helping tomorrow's surgeons achieve competence. In order to increase the external validity of this tool, additional studies are ongoing to further validate this tool in English using larger cohorts and increased numbers of evaluations.

Acknowledgments: The authors thank Paule Bodson-Clermont and Mylène Baptista for their work and guidance on statistical analysis and interpretation.

Affiliations: From the Centre Hospitalier de l'Université de Montréal (CHUM), Montreal, Que. (Mercier, Guertin, Bissada, Christopoulos, Tabet, Yang, Ayad); the Centre de recherche du Centre Hospitalier de l'Université de Montréal, Montreal, Que. (Christopoulos, Ayad); and the Hôpital Maisonneuve-Rosemont, Montréal, Que. (Olivier, Ayad).

Funding: A \$2000 research bursary was allocated to the authors from the resident research fund of the ÔTL-HNS department of Université de Montréal. This sum served to cover translation and statistician expenses.

Competing interests: None declared.

Contributors: É. Mercier, A. Christopoulos, and T. Ayad designed the study. É. Mercier, L. Guertin, E. Bissada, A. Christopoulos, M.-J. Olivier, J.-C. Tabet and T. Ayad acquired the data, which É. Mercier, N. Yang and T. Ayad analyzed. É. Mercier wrote the article, which all authors reviewed. All authors gave final approval of the article to be published.

Availability of data and materials: The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Content licence: This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY-NC-ND 4.0) licence, which permits use, distribution and reproduction in any medium, provided that the original publication is properly cited, the use is noncommercial (i.e., research or educational use), and no modifications or adaptations are made. See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

References

- Swing SR. The ACGME Outcome Project: retrospective and prospective. *Med Teach* 2007;29:648-54.
- Bhatti NI, Cummings CW. Competency in surgical residency training: defining and raising the bar. *Acad Med* 2007;82:569-73.

3. Brown DJ, Thompson RE, Bhatti NI. Assessment of operative competency in otolaryngology residency: survey of US program directors. *Laryngoscope* 2008;118:1761-4.
4. Mercier E, Chagnon-Monarque S, Lavigne F et al. Objective Assessment of Technical Skills in otorhinolaryngology-head and neck surgery residents: a systematic review. *Otolaryngol Head Neck Surg* 2018;158:54 – 61.
5. Faulkner H, Regehr G, Martin J et al. Validation of an objective structured assessment of technical skill for surgical residents. *Acad Med* 1996;71:1363-5.
6. 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.
7. Niitsu H, Hirabayashi N, Yoshimitsu M et al. Using the Objective Structured Assessment of Technical Skills (OSATS) global rating scale to evaluate the skills of surgical trainees in the operating room. *Surg Today* 2013;43:271-5.
8. Hopmans CJ, Den Hoed PT, Van der Laan L et al. Assessment of surgery residents' operative skills in the operating theater using a modified Objective Structured Assessment of Technical Skills (OSATS): a prospective multicenter study. *Surgery* 2014;156: 1078-88.
9. 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;23:354-9.
10. 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;120:2294-300.
11. Marglani O, Alherabi A, Al-Andejani T et al. Development of a tool for Global Rating of Endoscopic Surgical Skills (GRESS) for assessment of otolaryngology residents. *B-ENT* 2012;8:191-5.
12. Obeid AA, Al-Qahtani KH, Ashraf M et al. Development and testing for an operative competency assessment tool for nasal septoplasty surgery. *Am J Rhinol Allergy* 2014;28:e163-7.
13. McAlister ED, Goldstein DP, Rotstein LE. Redefining classification of central neck dissection in differentiated thyroid cancer. *Head Neck* 2014;36:286-90.
14. Ahmed A, Ishman SL, Laeeq K et al. Assessment of improvement of trainee surgical skills in the operating room for tonsillectomy. *Laryngoscope* 2013;123:1639-44.
15. Schwartz J, Costescu A, Mascarella MA et al. Objective assessment of Myringotomy and tympanostomy tube insertion: a prospective single-blinded validation study. *Laryngoscope* 2016;126:2140-6.