

Université de Montréal

Développement et validation d'un outil d'évaluation de la compétence chirurgicale pour l'évidement cervical

Par
Érika Mercier

Université de Montréal
Faculté de médecine

Mémoire présenté
en vue de l'obtention du grade de maîtrise (M.Sc.)
en sciences biomédicales
option recherche clinique

Août 2018

© Érika Mercier, 2018

Résumé

Objectifs : Développer et valider un outil servant à évaluer les compétences chirurgicales des résidents en Oto-rhino-laryngologie et chirurgie cervico-faciale (ORL-CCF) pour l'évidement cervical.

Méthodes : Une revue systématique de la littérature a été réalisée afin de recenser les méthodes validées pour évaluer les compétences chirurgicales en ORL-CCF. Les procédures chirurgicales pour lesquelles elles ont été développées ont été répertoriées. La revue a permis d'identifier un déficit en outils d'évaluation spécifiques aux procédures oncologiques en ORL-CCF, incluant l'évidement cervical.

Une méthode de Delphi modifiée a été utilisée afin de développer une liste d'étapes essentielles à la complétion d'un évidement cervical. Cette liste a été combinée à la grille générale du *Objective Structured Assessment of Technical Skills* (OSATS) pour développer un outil spécifique à l'évidement cervical. Cet outil a été validé en salle d'opération auprès de l'équipe d'oncologie ORL-CCF de l'Université de Montréal et de ses résidents.

Résultats : Un total de vingt-neuf évaluations ont été complétées au cours de l'année académique 2016-2017. L'acceptabilité a été jugée élevée auprès des résidents et des chirurgiens d'ORL-CCF, avec pour seule disparité l'utilisation formative ou sommative de l'outil. Les études de validation ont démontré des scores significativement plus élevés chez les résidents séniors que chez les juniors, ainsi qu'une progression significative des scores au fil du temps ($p < 0,05$). La tendance des scores sur la grille spécifique corrélait avec les résultats obtenus sur la grille générale précédemment validée ($p < 0,05$).

Conclusions : Le premier outil évaluant spécifiquement les compétences chirurgicales des résidents en ORL pour l'évidement cervical a été développé et validé.

Mots-clés : Oncologie, évidement cervical, oto-rhino-laryngologie, pédagogie, résident, évaluation

Abstract

Objectives: To develop and validate a new tool assessing surgical competency for Otolaryngology – Head & Neck Surgery (OTL-HNS) residents learning neck dissection.

Methods: A systematic review of literature was done to list methods developed to assess surgical competency in OTL-HNS. The surgical procedures for which these tools were developed were catalogued. A lack of evaluation tools specific to oncologic procedures in OTL-HNS was identified, which includes neck dissection.

A modified Delphi method was used to develop a list of steps deemed essential to a group of experts in order to complete a neck dissection. This list was combined to the general list of the *Objective Structured Assessment of Technical Skills* (OSATS) tool to develop a tool specific to neck dissection. This tool was validated in the operating room with the collaboration of the OTL-HNS oncology team of *Université de Montréal* and of the residents of this program.

Results: A total of twenty-nine 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 demonstrated significantly higher checklist scores for senior residents as opposed to 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 competency in OTL-HNS residents for neck dissection was successfully developed and validated.

Key-words: Oncology, neck dissection, otolaryngology, pedagogy, resident, evaluation

Table des matières

Résumé.....	i
Abstract.....	iii
Table des matières.....	iv
Liste des tableaux.....	xi
Liste des figures.....	xii
Liste des sigles.....	xiii
Dédicaces.....	xiv
Remerciements.....	xv

Chapitre 1 : Introduction

1. Évolution du modèle américain des programmes de résidence chirurgicale dans le temps.....	1
1.1 Historique des modèles d'enseignement médical.....	1
1.2 La réforme « Next Accreditation System » aux États-Unis.....	6
1.3 La résidence et l'éducation médicale à l'international.....	9
1.3.1 Royaume-Uni.....	9
1.3.2 Australie.....	11
1.3.3 Pays-Bas.....	12
1.3.4 Allemagne.....	12

1.3.5	France.....	13
1.3.6	Suisse.....	14
2.	Implémentation du programme d'apprentissage par compétences au Canada.....	16
2.1	Prémises et implantation.....	16
2.2	Implications.....	17
3.	Élaboration du projet actuel.....	18
3.1	Revue de la littérature.....	19
3.2	Validation d'un outil d'évaluation pour l'évidement cervical.....	20

Chapitre 2: Revue de la littérature

1.	Article: Objective Assessment of Technical Skills in Otolaryngology Head & Neck Surgery Residents: A Systematic Review.....	22
1.1	Contribution des auteurs.....	22
1.2	Abstract.....	25
1.3	Introduction.....	26
1.4	Methods.....	28
1.4.1	Data base review.....	28
1.4.2	Full manuscript analysis.....	29
1.4.3	Data extraction.....	30
1.5	Results.....	32
1.5.1	Study of surgical competency evaluation methods.....	32

• Figure 2: Numeration and types of evaluation methods.....	33
1.5.2 Indexation of surgical procedures on which residents are evaluated.....	33
• Figure 3: Surgical procedures and fields of OTL-HNS represented.....	34
1.5.3 Assessment of the reliability and validity of the evaluation methods.....	34
• Figure 4: Reliability assessment.....	35
• Figure 5: Validity assessment.....	36
• Figure 6: Reliability assessment for the combined GRS and TSC.....	37
• Figure 7: Validity assessment for the combined GRS and TSC.....	38
1.6 Discussion.....	39
1.7 Conclusions.....	41
1.8 Acknowledgments.....	42

Chapitre 3 : Développer et valider un outil évaluant la compétence chirurgicale pour l'évidement cervical

1. Article: Assessment of Surgical Competency for Neck Dissection: A Prospective Validation Study.....	43
---	----

1.1	Contribution des auteurs.....	43
1.2	Abstract.....	48
1.3	Background.....	50
1.4	Methods.....	52
1.4.1	Developing an assessment tool for neck dissection.....	52
1.4.2	Validating the neck dissection competency assessment tool.....	53
1.4.2.1	Participants and timeframe.....	54
1.4.2.2	Validity studies.....	54
1.4.2.3	Acceptability studies.....	54
1.4.2.4	Additional studies.....	55
1.5	Results.....	56
1.5.1	Descriptive data.....	56
1.5.2	Validity studies.....	57
1.5.2.1	Score progression with seniority.....	57
•	Figure 11: Results of junior and senior residents on the GRS checklist.....	58
•	Figure 12: Results of junior and senior residents on the TSC checklist.....	58
1.5.2.2	Score progression by level of residency.....	59

• Figure 13: Average scores by year of residency for GRS and TSC checklists.....	60
1.5.2.3 Score progression in time.....	60
• Figure 14: Resident score progression.....	61
1.5.2.4 Autonomy.....	61
1.5.3 Acceptability studies.....	61
1.5.4 Feasibility.....	62
1.6 Discussion.....	63
1.7 Conclusions.....	67
1.8 Acknowledgments.....	68

Chapitre 4: Discussion

1. Limites du projet.....	69
1.1. Biais d'inclusion.....	69
1.2. Étude sans double insu.....	70
1.3. Biais de rappel.....	71
1.4. Traduction de la grille OSATS.....	71
1.5. Biais d'apprentissage.....	72
2. Suites du projet.....	72
2.1. Version anglophone.....	72

2.2. Participation d'autres centres.....	73
2.3. Étendre à d'autres procédures.....	73
Chapitre 5: Conclusion.....	82
Bibliographie.....	83
Annexe 1: Figures.....	I
1. Chapitre 2:	
• Figure 1: Search Flow Chart.....	I
2. Chapitre 3:	
• Figure 8: Objective Structured Assessment of Technical Skills (OSATS) tool.....	II
• Figure 9: French global assessment tool.....	III
• Figure 10: French task specific checklist tool for neck dissection.....	IV
Annexe 2: Tableaux.....	V
1 Chapitre 3:	
• Tableau I: Questionnaire d'acceptabilité.....	V
• Tableau II : Traduction anglophone du questionnaire d'acceptabilité.....	VII
Annexe 3 : Documents supplémentaires.....	IX
• Document 1: Validité selon les critères de Sullivan.....	IX

- Document 2: Consignes pour les patrons d'ORL.....X
- Document 3 : Consignes pour les résidents d'ORL.....XI

Liste des tableaux

Chapitre 3

Table I : Questionnaire d'acceptabilité

Table II : Traduction anglophone du questionnaire d'acceptabilité

Liste des figures

Chapitre 2

Figure 1: Search flow chart

Figure 2: Numeration and types of evaluation methods

Figure 3: Surgical procedures and fields of OTL-HNS represented

Figure 4: Reliability assessment

Figure 5: Validity assessment

Figure 6: Reliability assessment for the combined GRS and TSC tool

Figure 7: Validity assessment for the combined GRS and TSC

Chapitre 3

Figure 8: Objective structured assessment of technical skills (OSATS) tool

Figure 9: French global assessment tool

Figure 10: French task specific checklist tool for neck dissection

Figure 11: Results of junior and senior residents on the GRS checklist

Figure 12: Results of juniors and senior residents on the TSC checklist

Figure 13: Average score by year of residency for GRS and TSC checklists

Figure 14: Resident score progression

Liste des sigles

ABMS: American Board of Medical Specialities

AMA: American Medical Association

APC: Activités Professionnelles Confiables

CBD: Competency By Design

CHUM: Centre Hospitalier de l'Université de Montréal

CPC : Compétence Par Conception

CRCHUM: Centre de Recherche du Centre Hospitalier de l'Université de Montréal

GRS: Global Rating Scale

HMR: Hôpital Maisonneuve-Rosemont

N/A: Non Available

NAS: Next Accreditation System

OTL-HNS: Otolaryngology and Head & Neck Surgery

ORL: Otorhinolaryngology

OSATS: Objective Structured Assessment of Technical Skills

TSC: Task Specific Checklist

À ma mère Elizabeth, qui a toujours été une source de support, de force et de détermination. À mon père Pierre, qui sera toujours mon modèle et mon inspiration, dans mes projets et dans ma vie.

Remerciements

J'aimerais remercier tout d'abord l'équipe d'oncologie tête et cou de l'Université de Montréal qui a dédié son temps et son énergie à la participation et au bon fonctionnement de ce projet. Sans votre collaboration et votre dévouement, ce projet aurait été impossible.

J'aimerais ensuite remercier l'équipe des résidents en oto-rhino-laryngologie et chirurgie cervico-faciale de l'Université de Montréal, qui ont eux aussi généreusement accepté de m'appuyer par leur implication dans le projet. Je vous remercie par la même occasion de votre camaraderie tout au long des dernières années. Un merci spécial au Dr Nathan Yang, qui a accepté de réviser et corriger les étapes statistiques.

Un énorme merci au Dr Tareck Ayad pour votre patience infinie. Vous avez été pour moi un mentor et un modèle tout au long de ma résidence. Votre aide m'a été inestimable et je partage chaque petite et grande victoire avec vous.

Enfin, merci à ma famille, à mes amis et à toi Paul, qui m'avez aidé à remettre les choses en perspective durant les moments de doute et de difficulté. Merci d'être présents pour moi en toute circonstance, et de toujours me pousser à réaliser mes rêves. Je suis éternellement reconnaissante de vous avoir et je vous aime plus que je ne peux vous le dire.

Chapitre 1: Introduction

1. Évolution du modèle américain des programmes de résidence chirurgicale dans le temps

1.1 Historique des modèles d'enseignement médical

Infrastructure de la pédagogie médicale en Amérique du Nord

Le système d'éducation médicale nord-américain d'aujourd'hui est le produit d'une longue évolution, intimement reliée à l'histoire de la médecine elle-même. Comprendre les racines du système d'éducation médicale actuel et son évolution permet de mieux saisir la situation actuelle et les enjeux propres à notre époque. La médecine est une discipline ancrée dans l'histoire et la tradition, et la compréhension du passé donne la perspective nécessaire pour saisir et mettre en contexte les changements qui façonnent le futur de la médecine¹. Cependant, les nombreux changements et progrès vus au fil des siècles ont été les précurseurs des réformes actuelles dans les systèmes d'enseignement médicaux nord-américains, tel que le *Next Accreditation System* aux États-Unis et la Compétence par Conception au Canada, reflétant un désir constant de perfectionner la formation du corps médical.

Les débuts du système d'éducation médicale en Amérique du Nord sont fortement influencés par le système d'apprentis déjà en vigueur en Angleterre et en Hollande au XVIII^e siècle^{2,3}. Les conditions sont telles dans les colonies que peu de médecins européens à l'époque voient un intérêt à venir s'y établir². L'éducation des médecins nord-américains dépend donc de quelques « maîtres » ayant obtenu leur éducation médicale avant leur arrivée en Amérique ou encore de membres du clergé². Les apprentis de l'époque passent entre cinq et huit ans à acquérir

les connaissances de leur maître pour toute formation avant de s'établir². Leur apprentissage consiste alors à faire des courses pour le maître, à manipuler les agents pharmaceutiques et à mélanger les plâtres, avant de progresser vers les saignées, arracher des dents et répondre aux appels médicaux nocturnes². Ce système étant bien souvent imparfait, quelques apprentis nord-américains plus ambitieux choisissent de compléter leur formation en Europe où des institutions reconnues existent³. Même si leur nombre est petit, l'influence de ces voyageurs est considérable dans les colonies². La fin du XVIII^e siècle voit apparaître les premières écoles de médecine nord-américaines. Dr John Morgan du Collège de Philadelphie prononce son discours *Discours upon the Institution of Medical Schools in America*, qui déplore le système d'apprentis, et propose un curriculum médical et suggère que les étudiants de médecine acquièrent une éducation pré-médicale plus complète, incluant les mathématiques, la philosophie et les langues².

Au début du XIX^e siècle, quelques écoles de médecine de qualité sont établies aux États-Unis, dont celles qui sont connues aujourd'hui comme Yale College, Columbia University, Harvard et Dartmouth³. Cependant, au fil des années, les standards associés à ces écoles sont perdus suite à l'apparition d'une multitude de nouvelles écoles de médecine sans affiliations universitaires, résultant en une perte de la qualité de l'éducation³. De plus, à cette période, aucun système de licence officielle n'est encore instauré comme c'était le cas en Angleterre, cette dernière préférant laissant chaque état réguler son licenciement², ce qui contribue aussi à la dégradation de l'enseignement. Quelques écoles attribuent le titre de Docteur en Médecine ou de Bachelier à leurs étudiants, généralement après une période de deux ans d'études, mais plusieurs apprentis se proclament praticiens sans posséder de qualification formelle³. En 1821, l'état du Connecticut met sur pied un comité d'examen obligeant les praticiens non licenciés à passer les qualifications de l'état, une procédure favorable qui s'est ensuite répandue aux autres états³.

En 1847 l'*American Medical Association* (AMA) est fondée sous l'influence du docteur New Yorkais Nathan Davis, dans le but d'élever et uniformiser les standards nationaux d'éducation médicale^{2,3,4}. En 1889, l'hôpital John Hopkins ouvre ses portes et offre le premier programme de résidence aux États-Unis⁴. En 1910, le rapport *Medical Education in the United States and Canada*^{4,5} aussi connu comme le *Flexner report*, est rédigé par Abraham Flexner, un éducateur travaillant pour la fondation Carnegie, une organisation visant l'avancement de l'éducation⁶. Celui-ci fait un recensement sévère de la qualité de l'éducation médicale auprès de 148 écoles de médecine américaines et de sept autres au Canada². Dans son rapport, Flexner propose des standards d'admission élevés, une rigueur académique sévère et des programmes basés sur une approche scientifique^{4,5}. Suite à la publication de ce document, près de la moitié des écoles de médecine de l'époque se sont vu attribuer le statut « inférieur aux normes » et ont vu fermer leurs portes dans les 15 années suivant la publication⁵.

En 1917, Milton C. Winternitz, affilié à l'école de médecine de l'Université Yale, lance un nouveau curriculum d'éducation médicale baptisé le *Yale Plan*³. Ce nouveau système confère une grande autonomie à l'étudiant, qui est autorisé à administrer son temps et à diriger ses lectures selon son jugement et ses intérêts personnels. Des examens périodiques sont administrés en cours de parcours, au terme desquels l'étudiant se présente aux examens de qualification nationaux³. Le début du XXe siècle voit aussi apparaître le concept d'« internat », consistant en une année d'éducation supplémentaire après l'école de médecine classique durant lequel l'étudiant était sur le terrain et confronté à des niveaux de responsabilité croissants⁷. Ce format devient rapidement le standard de formation. Dans les institutions d'élite, il est attendu des étudiants qu'ils s'immergent complètement dans le monde médical, donnant lieu au terme « résidents », vu leur résidence permanente dans l'hôpital⁷. Au cours des décades suivantes, la résidence devient la norme et les résidents deviennent graduellement essentiels au

fonctionnement de l'institution⁷. La signature du projet de loi qui mène à la création du *Medicare* et *Medicaid* le 30 juillet 1965⁸ amène un financement considérable pour l'éducation médicale et l'expansion des résidences se poursuit. La résidence n'est alors plus réservée à une élite, mais devient le parcours commun pour tous les gradués de médecine⁷.

Durant les années 1920 à 1940, des associations de spécialisation médicale sont mises sur pied et assurent les normes de qualification de leurs membres⁴. La première association de ce type fut le *American Board of Ophthalmology* en 1916, suivie par le *American Board of Otolaryngology* en 1924⁴. En 1928, l'AMA publie le *Essentials of Approved Residencies and Fellowships* qui établit des standards les programmes de résidence⁴. En 1937, le *American College of Surgeons* publie son propre recueil de standards d'éducation chirurgicale intitulé *Fundamental Requirements for Graduate Training in Surgery*⁴. Un nouveau comité, le *Accreditation Council for Graduate Medical Education (ACGME)* est créé en 1981, afin de faire face à la variabilité de la qualité de l'éducation dont bénéficiaient les résidents et encadrer l'émergence du phénomène la surspécialisation⁹. En réponse à ces demandes, l'organisme introduit un modèle de structure pédagogique, augmente la quantité d'enseignement formel et encourage l'évaluation des résidents ainsi que la rétroaction⁹. Par exemple, l'ACGM développe le *Common Program Requirements* qui établit une série de standards qui guident la préparation et l'entraînement des résidents, en définissant entre autres les rôles et qualifications du personnel (directeur de programme, faculté), les prérequis pour l'éligibilité en résidence et en fellowship, les détails du curriculum de chaque programme d'éducation (sessions didactiques, responsabilités lors de la prise en charge de patients et la supervision adéquate, apprentissage des connaissances médicales et application en pratique), ainsi que les méthodes d'évaluation sommatives et formatives^{10,11}.

Évolution des approches à l'éducation médicale au Canada

Le curriculum des écoles de médecine canadiennes de la période précédant la Deuxième Guerre mondiale est grandement influencé par celui originellement proposé à l'Université McGill, la première école de médecine au pays⁵. Le programme est influencé par le modèle d'Edinburgh⁵, développé au XVIIIe siècle, qui consiste en deux ans de cours de sciences théoriques et en laboratoire, suivi de deux ans d'apprentissage sur les étages de l'hôpital¹². Sir William Osler, un médecin réputé enseignant à l'Université de médecine de McGill, y introduit à son époque une forte emphase sur l'apprentissage au chevet de patients⁵.

Au cours des années 1950 à 1960, les facultés de médecine canadiennes sont submergées par la quantité toujours croissante de connaissances devant être intégrées au curriculum⁵. Les étudiants en médecine deviennent également mécontents par rapport à la quantité d'information à absorber, dont la pertinence ne leur apparaît pas toujours claire. Plusieurs écoles de médecine commencent alors à mettre en contact les étudiants avec les patients dès la première année, plutôt qu'à la troisième, dans le but de fournir un cadre aux sciences de base enseignées⁵.

Plusieurs écoles commencent à intégrer le modèle d'apprentissage par « systèmes », où les sujets sont intégrés en fonction d'un organe ou d'un système⁵. Ainsi, les sciences de base, la clinique, les diagnostics et les prises en charge sont appris dans une séquence coordonnée. Les cours progressent vers un format de séminaires et tutoriels dans le but d'encourager la prise de décision et favoriser l'autonomie des étudiants sur leur propre apprentissage⁵.

L'Université McMaster est la première à adopter l'approche par problème, qui a attiré l'attention du monde entier par la suite⁵. Dans ce modèle, les étudiants sont divisés en petits groupes et collaborent entre eux en utilisant les ressources à leur disposition afin de résoudre des problèmes médicaux soumis par la faculté, sous la supervision d'un instructeur⁵. Quelques

avantages d'un tel curriculum sont d'encourager le partenariat entre les étudiants et la faculté, en plus d'enseigner la résolution de problèmes⁴. Une étude publiée dans le *British Medical Journal* a cependant relevé certaines limitations à ce type d'enseignement, soit la possibilité d'une frustration chez les enseignants qui apprécient transmettre leur savoir et leurs expériences, la nécessité d'avoir plus d'enseignant pour superviser les groupes, la possibilité qu'un étudiant n'ait pas accès à un enseignant inspirant qui aurait pu enseigner à un grand groupe et enfin, la surcharge d'information, où un étudiant pourrait se perdre dans la manière de diriger son étude et dans le tri d'information pertinente ou non¹³.

1.2 La réforme “Next accreditation system” aux États-Unis

Les dernières décennies ont vu émerger plusieurs progrès et une réforme importante en éducation médicale. En tant qu'accréditeur de l'éducation médicale graduée, l'ACGME assume la responsabilité d'ériger et d'appliquer les standards qui régissent la formation de la prochaine génération de médecins. Dans le cadre de sa mission, l'ACGME introduit au début des années 2000 six domaines de compétence clinique, soit l'apprentissage et le perfectionnement basé sur la pratique, les soins aux patients et les compétences techniques, la pratique basée sur les systèmes, les connaissances médicales, les compétences interpersonnelles et de communication ainsi que le professionnalisme¹⁴. En 2009, l'ACGME développe un programme restructurant son système d'accréditation autour de ces compétences, menant à la création du *Next Accreditation System* (NAS)¹⁵, qui fut récemment implémenté dans les différentes spécialités à travers les États-Unis au cours des années 2013-2014. Les objectifs du NAS sont de préparer les médecins d'aujourd'hui à une pratique qui s'insère dans la mentalité de « revue par les pairs » du XXI^e siècle, d'accélérer le mouvement vers l'accréditation basée sur des issues éducationnelles concrètes et de réduire le fardeau relié aux imperfections de la structure du

système en place¹⁵. Le système précédant, appelé le *process-based system* présentait certaines limitations, entre autres son aspect « taille unique », où chaque étudiant devait se conformer au format du modèle éducationnel en place, l'absence de place à l'innovation pour les différents programmes d'enseignements, et une approche où on pénalise l'erreur plutôt que de récompenser le succès, comme le fait l'approche par compétences en encourageant la progression par l'acquisition d'aptitudes¹⁶. Le NAS repose sur un système de jalons (*milestones*) englobant un ensemble de compétences fondamentales spécifiques à chaque programme de résidence, et dont les étudiants doivent démontrer l'acquisition à des intervalles préétablis au cours de leur formation¹⁵. Le développement des différents jalons résulte d'une collaboration entre l'*American Board of Medical Specialities* (ABMS), des comités de révision, les différentes organisations de spécialisation médicale, les associations de directeurs de programme et des groupes de résidents¹⁵, et ce pour chaque spécialité. Il est important de comprendre cependant que les jalons ne représentent pas l'ensemble du curriculum d'un programme, mais plutôt des points de contrôle permettant de mesurer le progrès des résidents et d'assurer une uniformité de progression¹⁷. Ils devraient donc être jumelés à d'autres méthodes d'évaluation et de formation. De plus, le système de jalons se veut à visée formative afin d'aider les étudiants à passer avec succès leurs examens d'accréditation finaux et ne représente pas en lui-même une méthode d'accréditation¹⁷. Il n'est pas recommandé d'utiliser le système de jalons comme moyen d'évaluation en fin de stage, mais plutôt comme un guide de progression biannuel encourageant une discussion entre patrons évaluateurs et résidents¹⁷. Le NAS a pour objectif d'encourager cinq caractéristiques fondamentales d'une rétroaction efficace. Tout d'abord, la rétroaction devrait avoir lieu à l'intérieur de délais raisonnables. Deuxièmement, la rétroaction devrait être spécifique et ciblée sur des compétences spécifiques. Troisièmement, une rétroaction efficace devrait viser une balance entre la rétroaction positive et négative.

Quatrièmement, il faut permettre une période de réflexion et réaction afin d'encourager la mise en place d'un plan d'action. Enfin, l'objectif est de progresser au-delà de la discussion, et la rétroaction ne devrait être considérée complète que lorsqu'une action ou un changement n'ait été entrepris¹⁷.

Le NAS vise à créer une trajectoire logique de développement professionnel en ce qui a trait aux compétences jugées essentielles, tout en permettant des évaluations efficaces de ces dernières¹⁵. Les données en lien avec les évaluations permettront aussi à chaque centre de soumettre des tendances reflétant les performances de leurs résidents, et ainsi encourager une discussion pour promouvoir le perfectionnement du programme. Le NAS assure une éducation ciblée sur les compétences jugées pertinentes pour la santé d'individus et de populations tout en permettant la flexibilité d'intégrer aisément de nouvelles compétences émergentes au curriculum¹⁵. Quelques avantages du programme basé sur les compétences sont d'encourager l'application des connaissances plutôt que seulement l'acquisition des connaissances, d'être spécifiques à une spécialité, d'encourager un apprentissage individualisé et centré sur l'étudiant, de promouvoir la rétroaction et l'ajustement des plans d'apprentissages et enfin, de permettre multiples évaluations sur le terrain¹⁷. L'intégration de ce programme tend à s'éloigner des méthodes d'évaluation locales ou propres à chaque centre, qui présentent beaucoup de variabilité et peu d'uniformité, et à plutôt se tourner vers un ensemble d'outils d'évaluation pouvant être utilisés à travers les différents programmes et pour une même une région ou à travers le pays¹⁸.

1.3 La résidence et l'éducation médicale à l'international

La section suivante décrit la structure générale des programmes d'éducation médicale de six pays en dehors de l'Amérique. Les quatre premiers pays, soit le Royaume-Uni, l'Australie,

les Pays-Bas et l'Allemagne, se distinguent par leur avant-garde dans le développement du domaine de l'éducation médicale et par leur contribution à la littérature médicale internationale sur l'éducation^{19,20,21,22}. Ceci permet de faire une comparaison entre le modèle nord-américain et ce qui est fait actuellement à l'international dans d'autres programmes nationaux investis dans l'avancement de l'éducation médicale. Les deux derniers pays, la France et la Suisse, ne sont pas actuellement impliqués dans l'implémentation de programmes d'apprentissage par compétence et font office de contre-exemple à cette méthode, en plus de d'offrir une perspective francophone.

1.3.1 Royaume-Uni

Au Royaume-Uni, les études médicales pré-graduées durent en moyenne 5 à 6 ans, suite à quoi les étudiants complètent un internat obligatoire, durant lequel ils participent à un programme de fondements médicaux s'étendant sur deux ans¹⁹. La première année consiste en une période de transition du statut d'étudiant à celui praticien, à la fin de laquelle les étudiants reçoivent leur certification par le *General Medical Council*¹⁹. Au cours de la deuxième année, ils continuent à développer leurs compétences de base, reçoivent davantage de responsabilités et commencent à faire de la prise en charge de patients. Lors de la complétion de la deuxième année, les étudiants reçoivent une certification de réussite des compétence fondamentales¹⁹. Les gradués peuvent ensuite choisir d'entrer en spécialité ou de se diriger vers une pratique générale. Une autre alternative est d'aller obtenir de l'expérience clinique en travaillant comme médecin enregistré avant de commencer les études post-graduées. Dans certaines circonstances, des positions de médecin suppléant deviennent disponibles lorsqu'une place s'ouvre temporairement dans un programme de formation en spécialité¹⁹. Ces positions peuvent être remplies par tout médecin qui a terminé son programme de fondements. Un étudiant post-gradué

peut choisir d'y appliquer pour se donner un peu plus de temps pour choisir une spécialité, mais la place peut aussi être remplie par des spécialistes, des consultants ou des médecins généralistes¹⁹.

Le *General Medical Council* est l'organisme responsable de l'éducation médicale post-graduée. Les résidences peuvent durer entre trois et huit ans et consistent en deux types de formation possibles¹⁹. La première est la formation directe, où l'étudiant progresse automatiquement à l'étape suivante de sa formation après avoir atteint les compétences de base¹⁹. La deuxième option est la formation décalquée, et consiste en deux à trois années centrées sur la formation de base, à la fin desquelles le résident doit s'inscrire à une seconde ronde de compétition ouverte pour appliquer à un poste de formation en spécialité¹⁹. La formation est basée sur l'apprentissage par compétences, où chaque programme définit ses standards en termes de connaissances, d'habiletés techniques et de comportements, le tout en accordance avec les standards du *Good Medical Practice* établi par le *General Medical Council*¹⁹. Celui-ci décrit quatre domaines fondamentaux, soit le domaine des connaissances, des habiletés et de la performance, le domaine de la sécurité et de la qualité, le domaine de la communication et de la collaboration et enfin, le domaine de maintien de la confiance¹⁹. La supervision de la progression des compétences est faite par l'évaluation formative des *Supervised Learning Events*, soit l'équivalent des APC canadiens, à travers des évaluations sommatives de performance, des examens et le jugement des superviseurs. La complétion du programme donne accès à la certification, nommée *Certificate of Completion of Training*¹⁹.

1.3.2 Australie

En Australie, la formation pré-graduée s'étend sur cinq à six ans, suivie d'un internat obligatoire d'une année, durant laquelle les étudiants font des stages dans différents milieux de

la santé¹⁹. Cette année est nécessaire pour obtenir la certification par le *Medical Board of Australia*¹⁹. Il est possible de choisir d'obtenir de l'expérience clinique en travaillant comme médecin enregistré avant de commencer ses études post-graduées. La majorité des médecins juniors ayant complété leur première année d'études post-graduées continuent de travailler au moins une autre année dans des hôpitaux publics et ruraux, dans des centres de santé communautaires ou dans des centres de pratique générale¹⁹. Ce phénomène est causé entre autres par les pénuries de places dans la majorité des spécialités, et les admissions très compétitives. Cette période procure une exposition variée aux étudiants qui combent par la même occasion un besoin en services de santé¹⁹.

Les programmes de résidence durent entre trois et sept ans et certains programmes incluent des composantes d'apprentissage par compétences, mais ceci n'est pas encore universel en Australie¹⁹. Les standards éducationnels sont déterminés par le *Australian Medical Council* qui se charge des normes d'accréditation, mais la mise en place des curricula est établie par chaque état ou territoire¹⁹. La certification est obtenue suite à un examen divisé en deux parties, une écrite (le *AMC Computer Adaptive Test*) et une orale (le *AMC Clinical Examination*), toutes deux administrées par le AMC²³.

1.3.3 Pays-Bas

Aux Pays-Bas, les études pré-graduées durent six ans, suite à quoi les étudiants se présentent pour obtenir leur certification¹⁹. La dernière année en est une de transition où les responsabilités des étudiants croissent et ceux-ci atteignent le niveau de résident débutant. Une fois certifié, il est possible de choisir d'obtenir davantage d'expérience clinique en travaillant comme un résident hors-étude avant de commencer ses études post-graduées, généralement dans

l'hôpital où l'étudiant désire travailler¹⁹. Ils ne reçoivent alors pas de formation formelle, mais peuvent assister aux activités scientifiques.

Le *College of Medical Specialisms* régule la formation une fois en spécialité¹⁹. Les résidences durent entre quatre et six ans, et la formation en pratique générale est de trois ans¹⁹. L'évaluation de la progression aux Pays-Bas est modelée sur l'approche par compétences¹⁹. Tous les résidents débutants leur formation reçoivent un « plan d'éducation individuel » déterminé avec le directeur de programme, basé sur les compétences déjà acquises avant le début de la spécialité, en particulier si le résident a précédemment complété des années d'expérience clinique¹⁹. Au cours de la résidence, les résidents complètent un portfolio de leur progression¹⁹. Celui-ci est à la base de leurs évaluations et se veut un reflet de la progression de chaque résident.

1.3.4 Allemagne

Les étudiants en médecine allemands complètent six années d'études pré-graduées, puis entre cinq à six ans d'études post-graduées en résidence¹⁹. La formation post-graduée n'est pas rattachée à des centres académique et l'apprentissage repose plutôt sur l'expérience de travail, sans cours formels¹⁹. Le curriculum n'est pas encore centré sur l'acquisition de compétences, mais le système est en réforme ce format pourrait potentiellement être adopté au cours de prochaines années¹⁹. Les étudiants doivent passer trois rondes d'examens de certification d'état pour obtenir leur licence, appelé le *Approbatonsordnung für Ärzte*, sous la tutelle du German Medical Association. Le premier examen (M1 ou Physicum), a lieu après deux ans d'études pré-graduées, et consiste en deux jours d'examen écrit et un jour d'évaluation orale sur l'anatomie, la biochimie et la physiologie. Le deuxième examen (M2) a lieu avant la sixième et dernière année d'étude et consiste en un examen écrit extensif de trois jours basé sur des cas cliniques.

Enfin, le troisième examen (M3) a lieu après la dernière année de formation et combine des portions cliniques et écrites^{19,24,25}.

1.3.5 France

En France, le parcours d'éducation médicale est divisé en trois cycles qui s'étendent sur neuf à onze années d'étude. Au premier cycle, la première année est commune aux différents domaines d'études de la santé et se divise en deux semestres composés d'enseignement théorique et dirigé²⁶. L'année se solde avec quatre concours indépendants selon le parcours, soit médecine, pharmacie, odontologie et sage-femme²⁶. Seuls les étudiants bien classés au concours peuvent poursuivre leurs études dans le champ d'intérêt. Durant la deuxième année, les étudiants complètent un stage d'un mois en milieu hospitalier, sous l'autorité d'un cadre infirmier. La reste de l'année est passée à l'étude de matières fondamentales, ainsi qu'à l'apprentissage de l'anglais médical²⁶. Au deuxième cycle, la première année en est une d'apprentissage théorique et d'observation en milieu hospitalier. Les trois années suivantes forment la base de l'apprentissage clinique, et sont divisées en onze modules transversaux organisés par étapes de l'existence humaine plutôt que par biosystème, dans le but de favoriser une approche globale à la personne²⁶. Ces modules sont les suivants : Apprentissage de l'exercice médical, De la conception à la naissance, Maturation et vulnérabilité, Handicap, incapacité, dépendance, Vieillesse, Douleur, soins palliatifs, accompagnement, Santé et environnement, maladies transmissibles, Immuno-pathologie, maladies inflammatoires, Athérosclérose, hypertension, thrombose, Cancérologie, onco-hématologie, Synthèse clinique et thérapeutique, de la plainte du patient à la décision thérapeutique, urgences²⁶. À ceci s'ajoute des stages rémunérés de deux à quatre mois en milieu hospitalier, au cours desquels l'étudiant devient « étudiant hospitalier » et obtient ses premières responsabilités, approche des patients et assure la couverture de

gardes²⁶. L'étudiant doit passer avec succès un examen à la fin de chacune de ces années et présenter ses évaluations de stage. À la fin du deuxième cycle, il se présente au « certificat de synthèse clinique et thérapeutique », une évaluation pratique²⁶. Au troisième cycle, l'étudiant choisit un Diplôme d'Études de Spécialités en fonction de sa réussite au concours. La spécialité dure entre quatre et cinq ans auxquels s'ajoutent une ou deux années supplémentaires pour obtenir le diplôme d'Études de Spécialités Complémentaires pour l'acquisition de compétences additionnelles²⁶. Enfin, la complétion d'une thèse est obligatoire durant le troisième cycle et porte habituellement sur le domaine d'étude choisi. Le titre de docteur en médecine est obtenu après dépôt de la thèse et la satisfaction des exigences du cycle²⁶. Les différents examens sont régis par le Conseil national de l'Ordre des médecins²⁶.

1.3.6 La Suisse

En Suisse, la formation médicale débute avec la formation prégrade qui se donne aux universités de Genève, Lausanne, Zurich, Bâle et Berne, qui donne accès au diplôme fédéral de médecin²⁷. Les études médicales sont divisées en trois cycles, le premier cycle correspond à celui de bachelier et dure trois ans. Le deuxième cycle, aussi de trois ans, est l'équivalent d'une maîtrise, au terme duquel les étudiants passent le *Federal medical examination in human medicine*, encadré par le *Federal office of Public Health*. Enfin, le troisième cycle correspond à la spécialisation²⁸. La formation postgraduée s'étend sur cinq ans. Deux ans au minimum doivent être passés en formation clinique dans un établissement de catégorie A, soit un centre hospitalier qui offre une exposition et une formation touchant à l'ensemble des objectifs de la spécialité, tant au niveau du nombre de patients, des chirurgies offertes²⁹. Le centre dispose aussi de d'experts en une variété de sous-spécialités et sont engagés dans le développement de leur discipline, entre autres par la recherche. Au plus, trois ans peuvent être passés dans un centre

catégorie B, soit un centre hospitalier couvrant une grande partie des objectifs de formation de spécialité²⁹. Pour assurer une formation complète, ces centres doivent établir un contrat de collaboration écrit avec un centre de catégorie A, pour créer un réseau ou un regroupement de formation postgraduée²⁹. Une année au plus peut être passée dans un centre de catégorie C, offrant donc certaines parties de objectifs de formation de spécialiste, et ceux-ci doivent aussi s'associer avec des centres de catégorie A²⁹. Au moins une année de formation en centre hospitalier doit être passé dans un deuxième établissement hospitalier²⁹. De plus, les étudiants peuvent consacrer au plus six mois de leur formation comme assistant dans un cabinet médical reconnu ou dans un domaine ambulatoire, comme une polyclinique²⁹. Si désiré, un maximum de six mois peuvent être crédités pour la recherche scientifique dans le domaine sur demande préalable²⁹. Enfin, une année au plus de formation MD/PhD peut être validée en lieu d'une activité de recherche²⁹.

Au cours de leur dernière année de formation, les étudiants passent l'examen de spécialiste, qui est divisé en une première partie écrite donnée en français ou en allemand, composée d'une centaine de questions à choix multiples et s'étendant sur quatre heures, et une partie orale pratique²⁹. La partie orale est faite devant deux évaluateurs, dont un doit faire partie de la commission d'examen. Cette commission est nommée pour deux ans par le Société suisse d'ORL-CCF, et doit être formée d'au moins huit membres avec représentation adéquate des différentes régions linguistiques²⁹. La partie orale est individuelle et dure soixante à quatre-vingt-dix minutes²⁹. Le candidat doit préalablement déposer trois dossiers anonymes de patients qu'il a traité de manière autonome. Il est ensuite questionné sur un de ces cas dans la première partie de l'examen²⁹. Au cours de la deuxième partie, les examinateurs questionnent le candidat sur un à trois cas qu'ils ont préparés²⁹. La troisième partie évalue la relation médecin-patient et

le contact avec le patient et des questions d'éthique et d'économie de la santé sont aussi soumis au candidat²⁹. Les trois parties doivent être réussies pour passer avec succès l'examen.

2 Implémentation du programme d'apprentissage par compétences au Canada

2.1 Prémisses et implantation

Au Canada, un programme similaire au NAS américain a été lancé et encadré par le Collège Royal des médecins et chirurgiens du Canada, intitulé la compétence par conception (CPC). Ce nouveau programme a été lancé en juillet 2017 avec la mise en place de la CPC dans les programmes d'oto-rhino-laryngologie et chirurgie cervico-faciale et d'anesthésiologie³⁰ qui ont été choisis sur la base de la culture préexistante d'observation directe et de rétroaction dans ces programmes³¹. Cinq programmes supplémentaires se sont ajoutés en juillet 2018 et une douzaine de programmes additionnels seront intégrés à ce modèle en juillet 2019³⁰. L'objectif actuel est d'intégrer tous les programmes canadiens d'ici 2022³⁰.

Ce programme vise à remédier à certaines lacunes du système actuel et à mettre l'emphase sur l'apprentissage plutôt que sur la durée de formation³⁰. Il encourage entre autres la responsabilisation des étudiants face à leur apprentissage et favorise la transparence dans la formation. La CPC amène aussi une solution au renoncement à faire échouer les étudiants dans le système actuel, souvent attribué au manque de preuves objectives d'inaptitude³⁰. Similairement au NAS, la CPC vise à favoriser une rétroaction utile et rapide, à établir clairement des objectifs propres à chaque programme et à créer une ligne du temps commune pour l'acquisition de compétences en résidence³⁰.

2.2 Implications

Un nouveau vocabulaire s'insère dans la mise en place de la CPC. Les activités professionnelles fiables (APC) sont définies comme des tâches en lien avec la discipline qui peuvent être confiées à des résidents et dont la complétion peut être examinée et évaluée par un superviseur³⁰. Multiples observations sériées d'une APC dessinent un portrait global du rendement d'un apprenant et déterminent son passage à l'étape subséquente, en suivant une pente de difficulté croissante. Chaque spécialité définit le nombre et le type d'APC pour assurer la bonne progression de leurs résidents³⁰. Ceux-ci permettent aux résidents de cibler les attentes qu'ont les enseignants envers eux, selon leur niveau. Les APC comprennent en général plusieurs jalons, soit des compétences individuelles nécessaires à la réalisation de la tâche. Les jalons englobent une variété d'habiletés en lien avec l'expertise, la communication et la compétence technique et sont reliés aux différents rôles CanMEDS³⁰. Ils permettent de cerner les endroits où un apprenant doit s'améliorer en vue de réussir une APC. La signification du référentiel CanMEDS sera discuté plus bas.

Ensemble, les APC et les jalons servent de guide et de référence aux évaluateurs ainsi qu'aux résidents afin d'orienter le parcours d'apprentissage, lui-même divisé en quatre étapes, soit la progression vers la discipline, l'acquisition des fondements de la discipline, maîtrise de la discipline et transition vers la pratique³⁰. Par la même occasion, l'objectif du Collège Royal implique de placer les examens du Collège Royal à la fin de l'étape de maîtrise de la discipline, afin de permettre aux finissants de se concentrer sur l'apprentissage de l'autonomie et peaufiner leurs compétences³⁰. Pour obtenir le droit de pratique indépendant, les résidents auront tout de même l'obligation de terminer l'étape transition vers la pratique pour compléter leur certification³⁰.

Un projet précurseur à cette réforme du système d'éducation médical fut le développement du référentiel CanMEDS en 1990, lui aussi encadré par le Collège Royal des médecins et chirurgiens du Canada. Le référentiel CanMEDS a pour objectif de définir les compétences essentielles à l'exercice de la médecine et établir les bases de l'enseignement médical au Canada³². Il s'articule autour de sept rôles que doit incorporer un médecin compétent, soit expert médical, communicateur, collaborateur, leader, promoteur de la santé, érudit et professionnel³². Le référentiel est le fondement des certifications et de l'agrément au Canada et ses principes sont au cœur du programme de CPC³².

3 Élaboration du projet actuel

Les changements prenant place dans le système d'éducation médical canadien décrits précédemment auront des impacts directs sur la structure des programmes de résidence et sur la manière dont les résidents sont évalués. Un des bénéfices sera la création d'un parcours de résidence taillé sur mesure, adapté aux besoins de chaque individu, permettant aux étudiants de consacrer plus de temps aux compétences plus difficiles à assimiler, et de passer rapidement sur les notions déjà acquises. Ce modèle permettra la mise en place de résidences de durées variables, selon le temps que met chaque résident à acquérir l'ensemble des compétences. La manière dont les évaluateurs détermineront le succès à performer un acte médical devra évoluer pour répondre aux nouvelles exigences. Dans les cas de programmes chirurgicaux, ceci implique de créer un accès à des méthodes objectives et reproductibles d'évaluer la compétence des résidents à performer de manière sécuritaire, organisée et autonome les différents gestes chirurgicaux faisant partie de leur curriculum. La création et le développement de tels outils avaient été peu explorés dans le passé étant donné qu'historiquement la compétence chirurgicale

était déterminée par le jugement et les observations des chirurgiens enseignants en salle d'opération. Ces évaluations sont généralement subjectives, ne sont pas basées sur des critères clairement définis et manquent de fiabilité en raison du biais de rappel et de la variabilité inter-observateur³⁴. L'Université de Montréal et le programme d'ORL-CCF se trouvant à l'avant-garde de ces changements, nous nous sommes penchés sur ce besoin croissant de moyens accessibles, économiques et fiables de documenter la progression des résidents au bloc opératoire. Nous avons jugé que la meilleure manière de peindre un tableau précis de la situation et des ressources disponibles était de procéder à une revue de la littérature répertoriant les méthodes d'évaluation des compétences chirurgicales en ORL-CCF.

3.1 Revue de littérature

Une réforme majeure du curriculum de formation des futurs médecins est en cours actuellement au Canada avec le développement et l'implémentation du programme de « Compétences par conception »³¹. Cette réforme s'insère dans le cadre d'une restructuration internationale de l'enseignement médical, avec un virage de plus en plus généralisé vers des méthodes d'apprentissage par acquisition de compétences³¹. L'évaluation objective de compétences fondamentales sont au cœur de tels programmes et dans le cas de spécialités chirurgicales, ceci implique l'évaluation de plusieurs compétences techniques. L'appréciation des compétences chirurgicales était généralement faite de manière informelle et subjective dans le cadre du curriculum classique, offrant peu de reproductibilité et de moyens de documentation standardisés³⁴. L'implémentation de la CPC nécessite donc la création de nouvelles méthodes d'évaluation adaptées à l'appréciation de compétences techniques de manière uniforme. Dans ce contexte, la première étape de notre projet a consisté à déterminer quelles sont les méthodes disponibles et validées actuellement pour évaluer les compétences chirurgicales des résidents

en ORL-CCF. Nous voulions aussi déterminer pour quels types de chirurgies ces méthodes avaient été développées et pour quels champs de la spécialité. Pour ce faire, nous avons procédé à une revue de la littérature extensive. Nous avons fait ressortir que la méthode d'évaluation la plus utilisée actuellement pour démontrer la compétence chirurgicale des résidents en ORL-CCF est la grille de type liste double, évaluant à la fois les compétences chirurgicales générales et les compétences spécifiques à une procédure. Ce type de méthode d'évaluation a été validé à multiples reprises dans la littérature, entre autres avec la grille OSATS³⁴. Nous avons aussi fait ressortir que des outils avaient déjà été développés et validés pour la grande majorité des chirurgies de base de l'ORL-CCF, englobant presque tous les champs de la spécialité. Nous avons cependant identifié une sous-représentation de procédures oncologiques, malgré que ces chirurgies soient courantes et que leur apprentissage fassent partie des objectifs de résidence en ORL-CCF.

3.2 Validation de l'outil d'évaluation

La deuxième partie du projet a consisté à développer et valider notre propre outil d'évaluation dans le but de combler les besoins identifiés en évaluation de chirurgies oncologiques, en choisissant de nous concentrer sur l'évidement cervical. Cette procédure est courante, fait partie des objectifs de la résidence, comporte des étapes bien définies et est suffisamment complexe pour générer une courbe d'apprentissage sur plusieurs années. Nos études de validation ont démontré que l'outil est suffisamment sensible pour identifier une progression chez un même résident, ainsi qu'entre des résidents de différents niveaux, ce qui est désirable pour évaluer le progrès d'un résident, ou au contraire identifier un retard. De plus, notre outil est applicable dans un environnement académique, de par sa rapidité de complétion, sa simplicité et son accessibilité. L'outil est peu coûteux et peut être utilisé dans une visée

formative comme formelle. Il produit un rapport standardisé, quantitatif et objectif, qui permet de faire un suivi chronologique de l'évolution du résident. Afin de compléter les études de validation, des analyses de données inter et intra-observateur auraient été intéressantes afin de déterminer respectivement l'impact d'avoir multiples évaluateurs sur les résultats des évaluations et pour évaluer la tendance des résultats chez un même évaluateur. Cependant, nous n'avons pas un nombre suffisamment élevé d'évaluations après une année d'étude pour permettre la réalisation de telles analyses. Le projet se poursuit au-delà du spectre de cette étude et ces analyses pourront être réalisées dans une phase ultérieure.

Chapitre 2: Revue de la littérature

1. Article: Objective Assessment of Technical Skills in Otolaryngology Head & Neck Surgery Residents: A Systematic Review

1.1 Contribution des auteurs

1 Érika Mercier

- a) Conception, préparation et supervision du projet
- b) Triage et étude des articles retrouvés
- c) Compilation, analyse et interprétation des données recueillies
- d) Rédaction et correction du manuscrit

2 Ségolène Chagnon-Monarque

- a) Triage et étude des articles retrouvés
- b) Compilation, analyse et interprétation des données recueillies

3 François Lavigne

- a) Révision et correction du manuscrit

4 Tareck Ayad

- a) Conception, préparation et supervision du projet
- b) Révision et correction du manuscrit

Title:

Objective Assessment of Technical Skills in Otolaryngology Head & Neck Surgery Residents:
A Systematic Review

Authors:

Érika Mercier, M.D.¹, Ségolène Chagnon-Monarque², François Lavigne M.D., FRCSC¹,
Tareck Ayad, M.D, FRCSC¹.

1. Centre Hospitalier Universitaire de Montréal, Montréal, Canada
2. Medical student at Université de Montréal, Montréal, Canada

Corresponding author:

Tareck Ayad

Centre Hospitalier Universitaire de Montréal

1560, Sherbrooke Est

Montréal (Québec), Canada, H2L 4M1

Presented at the Quebec Otolaryngology Annual Meeting of 2016, in Montréal, Canada

Presented at the Canadian Society of Otolaryngology Annual Meeting of 2017, in Saskatoon,
Canada

Presented at the ENT World Congress of 2017 (IFOS), in Paris, France

Published in the American Academy of Otolaryngology Journal in 2017

Keywords:

Resident; residency; assessment tool; assessment method; evaluation; evaluation tool;
evaluation method; reliability; validity; surgical competency; competency

1.2 Abstract

Objectives: The primary goal is the indexation of validated methods used to assess surgical competency in Otorhinolaryngology – Head & Neck Surgery (OTL-HNS) residents. Secondary goals include assessment of the reliability and validity of these tools, as well as the documentation of specific procedures in OTL-HNS involved.

Data Sources: MEDBASE, OVID, Medline, CINAHL and EBM, as well as the printed references available through the *Université de Montréal* library.

Review Methods: The PRISMA method was used to review digital and printed databases. Publications were reviewed by two independent reviewers and selected articles were fully analyzed in order to classify evaluation methods and categorize them by procedure and subspecialty of OTL-HNS involved. Reliability and validity were assessed and scored for each assessment tool.

Results: Through the review of 30 studies, five different evaluation methods were described and validated to assess the surgical competency of OTL-HNS residents. The evaluation method most often described was the combined global rating scale (GRS) and task-specific checklist (TSC) tool. Reliability and validity for this tool were overall high, however considerable data was unavailable. Eleven distinctive surgical procedures were studied, encompassing many subspecialties of OTL-HNS: facial plastics, general ENT, laryngology, otology, pediatrics and rhinology.

Conclusions: Although assessment tools have been developed for a wide array of surgical procedures, involving most OTL-HNS subspecialties, the use of combined checklists has been validated repeatedly in the literature and shown to be easily applicable in practice. It has been applied to many OTL-HNS procedures, but not in oncologic surgeries to date.

1.3 Introduction

Over the past few years, the conventional framework of residency where competency is established through time spent rotating on different surgical services has been gradually replaced with a curriculum based on the objective acquisition of program specific competencies^{35,36}. This change is happening in various countries across the world including the United States where the Milestones³ program was implemented by the Accreditation Council for Graduate Medical Education (ACGME). In Canada, the Royal College of Physicians and Surgeons (RCPS) is gradually implementing the Competence by Design program³⁸. The Intercollegiate Surgical Curriculum Program (ISCP) in the United Kingdom has already integrated the Procedure Based Assessment (PBA) as part of their surgical programs³⁹. These programs have in common the role of providing assessment of trainee's operative technical skills for routine surgical procedures until certification. For surgical residency programs such as Otolaryngology, these new curriculums imply both the acquisition of clinical and technical skills. In the current system, it is customary for the assessment of a resident's technical skills to be evaluated informally by attending surgeons during operating time or retrospectively at the end of a resident's rotation. However, this method does not ensure that residents receive reproducible, consistent and standardized feedback to help them identify their strengths and weaknesses⁴⁰. For program directors, this method also does not provide a means of objectively and accurately ensuring the acquisition of varied surgical skills⁴¹. To palliate this problem, various assessment tools have been progressively developed and validated in order to provide a traceable and dependable way to evaluate surgical skills in otolaryngology residents⁴². Despite this progress, many standard Otolaryngology procedures are devoid of specific evaluation methods⁴³. The main objective is to review the available literature in order to identify the methods currently available and validated to assess surgical competency in Otolaryngology residents. We also aim

at reporting the reliability and validity of these assessment tools, as well as identifying specific procedures or areas of sub-specialties in Otolaryngology for which the development of surgical assessment tools might still be needed.

1.4 Methods

1.4.1 Data base review

In order to answer the aforementioned questions, we conducted a systematic review following the guidelines of the PRISMA statement. An extensive research was conducted on January 18th 2016, with the help of a professional health librarian affiliated to our institution. All relevant published studies were extracted from five different databases including MEDBASE, OVID, Medline, CINAHL and EBM, as well as the printed literature available through the *Université de Montréal* library. The terminology used for the search were: Aptitude Tests OR operative competence OR clinical competence OR comprehensive assessment OR surgical skill OR surgical aptitude OR aptitude test OR dexterity test OR general competence OR surgical performance OR Measuring competence OR drilling skill OR Clinical Competence OR Educational Measurement OR Competency-Based Education AND Otolaryngology OR otolaryngology OR Sinusitis/Surgery OR Rhinitis/Surgery OR Otorhinolaryngologic Surgical Procedures OR Bronchoscopy/standards OR Laryngoscopy/standards OR Otologic Surgical Procedures OR Mastoid OR Nasal Septum/surgery OR Paranasal Sinuses/surgery OR Rhinoplasty/standards OR Thyroidectomy/standards OR Tonsillectomy/standards OR Ear/surgery OR Nose/surgery OR Larynx/surgery AND Internship and Residency OR Residency OR Internship OR resident OR Education, Medical, Graduate OR trainee OR training OR graduate.

This search strategy resulted in the extraction of a total of 1453 studies from the available digital literature and of 14 articles from printed literature. Two reviewers checked the references for duplicates and a total of 1416 original articles were retained for analysis. All article titles and abstracts were reviewed independently by É.M and S.C.M, and were retrieved if both

reviewers considered the study to be relevant. Any disagreement between reviewers was solved by discussions among the authors to reach consensus or by a third party (T.A), if necessary. Criteria used for inclusion of studies were: research written in English or French, published after 1990, containing information on assessment tools for resident evaluation in Otolaryngology Head & Neck Surgery (OTL-HNS). Assessment tools from other surgical specialties were not considered in this review. Exclusion criteria were: not medical education, not ENT-specific, not specific to assessment of surgical skills, not a study/not the desired type of research paper (editorial, opinion essay, open letter), not related to or applied to residents and finally not available in full text (article could not be retrieved in its entirety).

1.4.2 Full manuscript analysis

Application of these criteria resulted in 43 abstracts identified as possible relevant studies. These 43 full-text articles were then assessed for eligibility. To identify relevant publications, the articles were evaluated independently by the same two reviewers (E.M, S.C.M.) and the following aspects were assessed and analyzed: the type of article, the nature of the tool described in the paper, the otolaryngology subspecialty involved, the surgical procedure for which the tool was developed, the overall risk of bias and the main conclusions of the authors. In cases where only abstracts of the original articles were available through regular research in databases and search engines, an advanced search was conducted by the health studies librarian to gain access to the missing article. Through this process, thirteen studies did not meet the inclusion criteria, and were excluded for the following reasons: one study assessed surgeon's accreditation, one study determined the surgical aptitudes of otolaryngology applicants, one study did not validate an assessment tool, one study was not applied to residents, one study was only available in Japanese even though the abstract was in English, four studies were not specific to OTL-HNS

procedures, three studies were only abstracts of oral presentations and finally one study focused on evaluating the components that contribute to measurement error but was not a true analysis of the scale or its usefulness as an evaluation tool. In the end, 30 publications met our inclusion criteria and were included in our final analysis. The flow chart covering the search and analysis of available literature is described in Figure 1 (Annex 1, p.I).

1.4.3 Data extraction

For each of the 30 articles retained, the following information was obtained: year and location of publication, type of article, evaluation method, number and type of participants, subspecialty of OTL-HNS, surgical procedure involved as well as general conclusions. Subspecialties of OTL-HNS included facial plastics, general otolaryngology, laryngology, otology, pediatrics, rhinology, skull base/neurotology and oncology. Studies were then organized by method of evaluation used and by type of procedure involved.

Among the studies included in the qualitative synthesis, an additional assessment of the reliability and validity of the evaluation tools was performed by the same two reviewers (É.M, S.C.M.), based on the evaluation elements highlighted by Gail M. Sullivan, in an editorial on the validity of assessment instruments⁴⁴ (Annex 3, Document 1, p.IX). Dr Sullivan's criteria were chosen as a reference because the editorial in which they are described was frequently referred to in literature for survey validations and in the methodology of studies similar to our own⁴⁵. Evaluation of reliability was based on internal consistency, inter-rater consistency and intra-rater consistency, as determined respectively by the Cronbach's alpha value or Pearson's correlation coefficient, the Kappa value and the intra-class correlation coefficient (ICC). Analysis of validity consisted of appraisal of the instrument's content, of the response process as defined by the relation between the subject's thoughts and the test results, of the relationship

of the instrument to other variables, such as a previous study or a predefined gold standard and finally, of the consequences defined as the pass or fail performance in the instance of a cut-off performance score. For analysis of reliability, each article was scored according to each of these criteria as either low, medium or high by the two reviewers (É.M., S.C.M.) according to Cronbach alpha and Kappa values. Values below 0,49 were marked low, values between 0,50 and 0,75 were medium and values of 0,75 and above were marked high. When information about these criteria was unavailable, the item was marked N/A. The measure of the components of validity as described by Sullivan et al. remain subjective. As such, general guidelines were used to classify the articles into high, medium and low categories. High scores were attributed to articles with the presence of objective values confirming elements of high validity scores, the use of validated methods to develop tools (e.g. Delphi method), extensive description of a quality methodology or evidence in the manuscript that the component of validity described was fully in compliance with the description made by Sullivan et al. Low scores were attributed if no validated or objective method were used to develop the new tool, in presence of objective evidence of low validity scores or in the absence of elements in compliance with the descriptions by Sullivan et al for each component. Finally, medium scores were attributed if a mix of elements from both high and low score were present, or if elements did not satisfactorily fit the criteria to qualify into one or the other category.

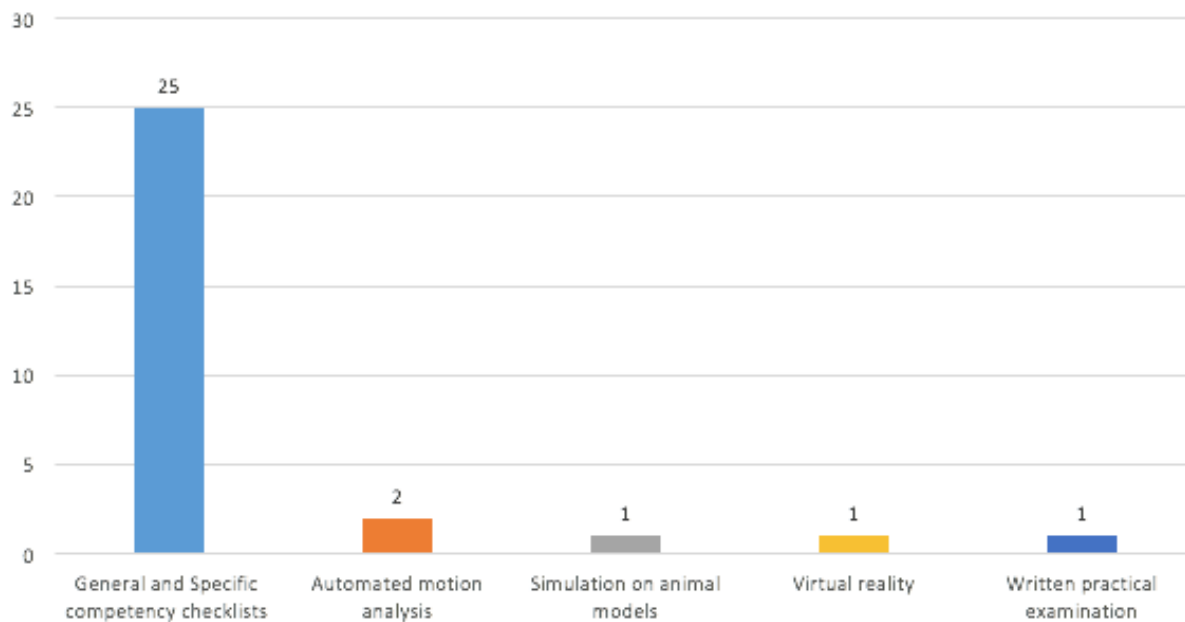
1.5 Results

1.5.1 Study of surgical competency evaluation methods

Five different evaluation methods were described throughout the 30 articles reviewed. They are the combined global rating scale (GRS) and task-specific checklist (TSC), automated analysis of the surgeon's motions, simulations on animal models, virtual reality technology and finally written practical examinations. The combined checklist tool was by far the most used evaluation method in the literature, described in 25 out of 30 articles reviewed. It is composed of two complimentary checklists, the first being the global rating scale (GRS), originally based on the OSATS tool, developed by Martin et al.⁴⁶, that has been repeatedly used and validated in different specialties^{47,48,49,50}. It comprises a set of general skills applicable to any surgical procedure from whichever specialty, that must be mastered to become a proficient surgeon⁵¹. The second scale is the task-specific checklist (TSC), which comprises a list of skills and steps specific to a surgical procedure that are necessary to the safe and autonomous completion of the procedure⁵². These tools can be used in either live surgery cases or in a simulation setting. The checklists listed in this review were used to assess surgical competency in residents in both situations. The second most prevalent evaluation method, the automated analysis of motion, comprises devices that record and analyses micro-movements of the surgeon's hand during a procedure, and examines patterns of motion associated with skill and experience⁵³. Simulation on animal models include completing procedures on cadavers and/or live anesthetized animals, to assess the surgical skills of the primary surgeon⁵⁴. Virtual reality has been used to simulate surgical situations and cases, enabling the resident to explore techniques safely and redo procedures as needed, while allowing automated or direct staff evaluation⁵⁵. Finally, written examinations imply having residents complete a standardized set of printed or computed-based

questions, ranging from drawing a step of the procedure, short answers and descriptions⁵⁶. A graph analysis of these evaluation methods as well as their representation throughout available literature is presented in Figure 2 (p.33).

Figure 2: Numeration and types of evaluation methods

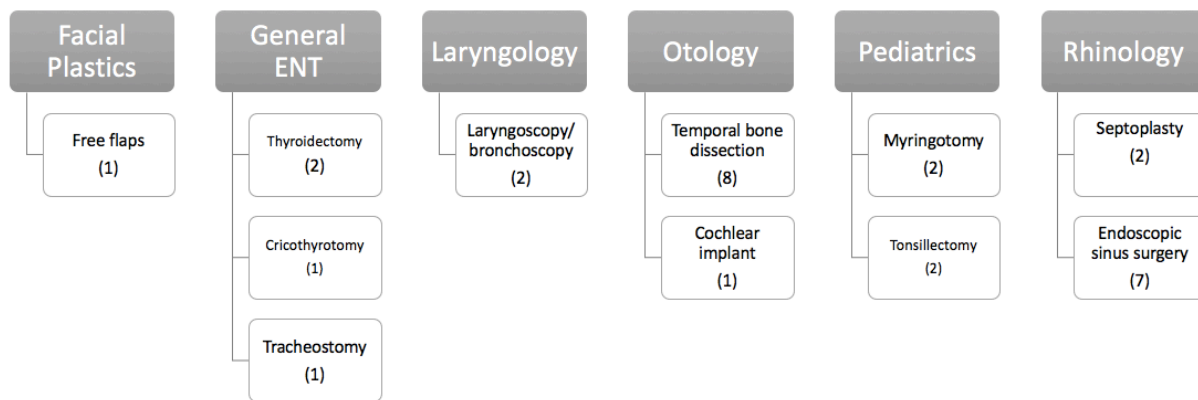


1.5.2 Indexation of surgical procedures on which residents are evaluated

As mentioned previously, the type of surgical procedure on which residents were evaluated and the area of OTL-HNS involved for each assessment tool was indexed during the full article analysis. This data was used to put forth procedures for which validated assessment tools have been developed, and in the same way, highlight any gaps in the existing literature on the subject. The complete indexation of represented procedures and corresponding areas of OTL-HNS are shown in Figure 3 (p.34). Eleven different procedures were found to already have

at least one specific evaluation tool, from six areas of OTL-HNS: facial plastics, general ENT, laryngology, otology, pediatrics and rhinology. The procedures most often used to assess resident's levels of surgical skill were temporal bone dissection and endoscopic sinus surgery, with respectively 8 and 7 studies focusing on these procedures.

Figure 3: Surgical procedures and fields of OTL-HNS represented



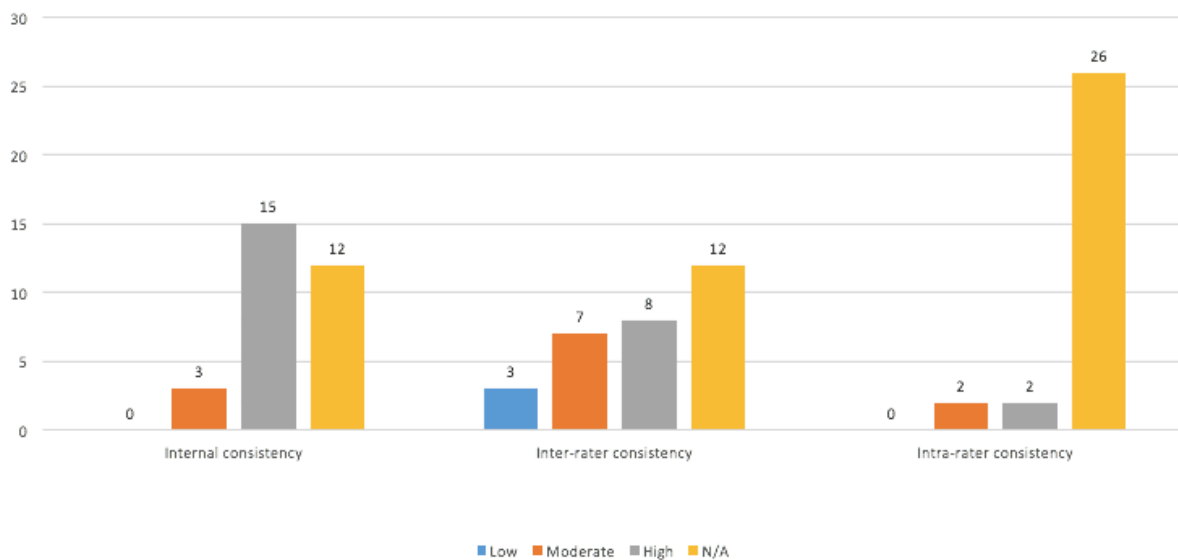
*Numbers in brackets represent the number of publications studying this procedure

1.5.3 Assessment of the reliability and validity of the evaluation methods

As stated by Gail M. Sullivan's criteria (Annex 3, Document 1, p.IX), the assessment of reliability was based on internal consistency, inter and intra-rater consistency for each evaluation tool⁹. Internal consistency serves to confirm that the tool tests for the concept of interest, while intra and inter-rater consistency test respectively the outcomes of having only one versus many raters use the tool, to determine the reproducibility of the results⁵⁷. Each of these criteria was ranked as either high, moderate or low. Internal consistency was overall high, with 15 tools ranking high, three ranking moderate, and none scoring low. However, 12 papers did not present sufficient data to draw useful conclusions in the internal consistency criteria. Inter-rater consistency was ranked high in seven articles, moderate in six and low in three.

Fourteen articles have insufficient information to permit useful analysis. More importantly, very little information was available to assess intra-rater consistency, as 26 articles out of 30 had insufficient data. Two articles were rated high for this item, two more rated moderate, while none rated low. Aforementioned data is depicted in Figure 4 (p.35).

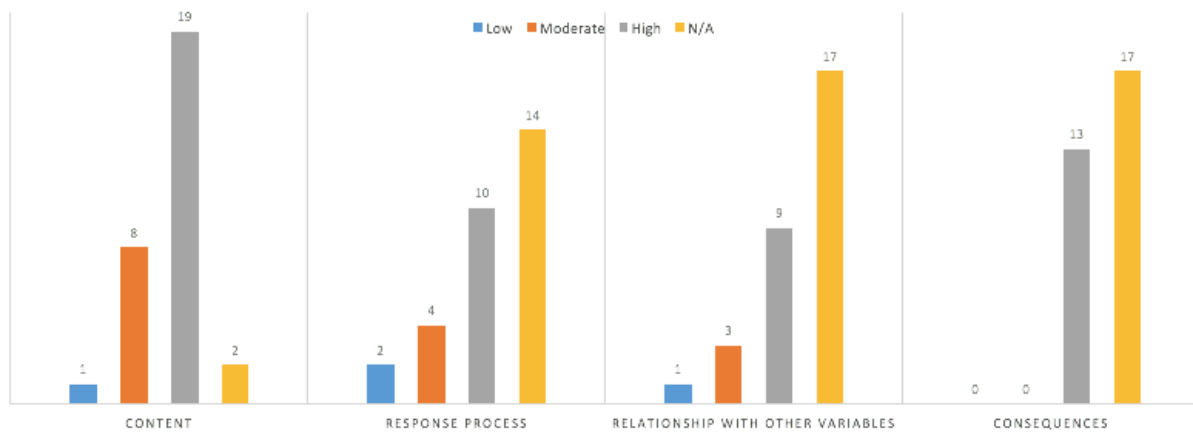
Figure 4: Reliability assessment



Assessment of validity was done in the same manner, using the following criteria: content, response process, relationship with other variables and consequences. “Content” encompasses the quality of the process and steps leading to the creation of the tool, including people and concepts involved. The “response process” evaluates if the subject’s test results reflect his actions. It also encompasses the quality of the rater’s training as well as the instructions given to the subjects prior to evaluation. The “relationship with other variables” criterion relates to the way the evaluation tool compares to other similar tools or to a gold-standard. Lastly, the “consequences” criterion refers to the capacity of the evaluation tool to reflect better scores after additional training. The majority of articles scored high on the content criterion, with 19 ranking high, 8 moderate and 1 low. Data was unavailable for 2 studies. On

the response process criterion, 10 evaluations tools were rated high, 4 as moderate and 2 low. However, 14 studies did not have sufficient data for useful analysis. In regards to relationship with other variables, as many as 17 studies had insufficient data for analysis. Within the ones assessed, 9 tools scored high, 3 moderate and 1 low. Finally, the consequences criterion was rated high on 13 studies. No studies were rated moderate or low. Lastly, data was insufficient for 17 studies. This information is summarized in Figure 5 (p.36).

Figure 5: Validity assessment



Similar analyses were conducted including this time only with the studies in which the primary assessment method was the combined GRS and TSC tools, in order to gain perspective on the inherent reliability and validity of this widely used assessment tool. The GRS and TSC checklists are the most reviewed and possibly the most promising prospective resident evaluation methods. They are straightforward, inexpensive, readily available and easily applicable to an academic practice, where time is often a limiting factor^{58,59,60}. These analyses put forth the results of the validity and reliability extracted from their corresponding studies. They show favorable outcomes overall, with 67,7% of studies displaying high reliability and 74,5% of studies exhibiting high validity, within the studies displaying sufficient data for analysis. More specifically, the reliability scores were predominantly high on all three

predefined criteria (internal consistency, inter-rater consistency and intra-rater consistency), with respectively 14, 7 and 1 studies rating high. Comparatively, only 3, 5 and 1 studies were respectively rated moderate, while a mere 2 studies ranked as low in the inter-rater consistency measure. However, 7, 10 and 22 studies had insufficient data in each one of the three categories. These results are summarized in graph form in Figure 6 (p.37). In the same manner, validity scores were predominantly high with content, response process, relationship to other variables and consequences criterion scoring high in respectively 16, 8, 6 and 11 studies. In the same order, 6, 8, 6 and 0 scored moderate, while only 1 study scored low on each content and response process criteria. Finally, 1, 12, 15 and 13 studies respectively had insufficient data for analyses. Figure 7 (p.37) summarizes this data.

Figure 6: Reliability assessment for the combined GRS and TSC

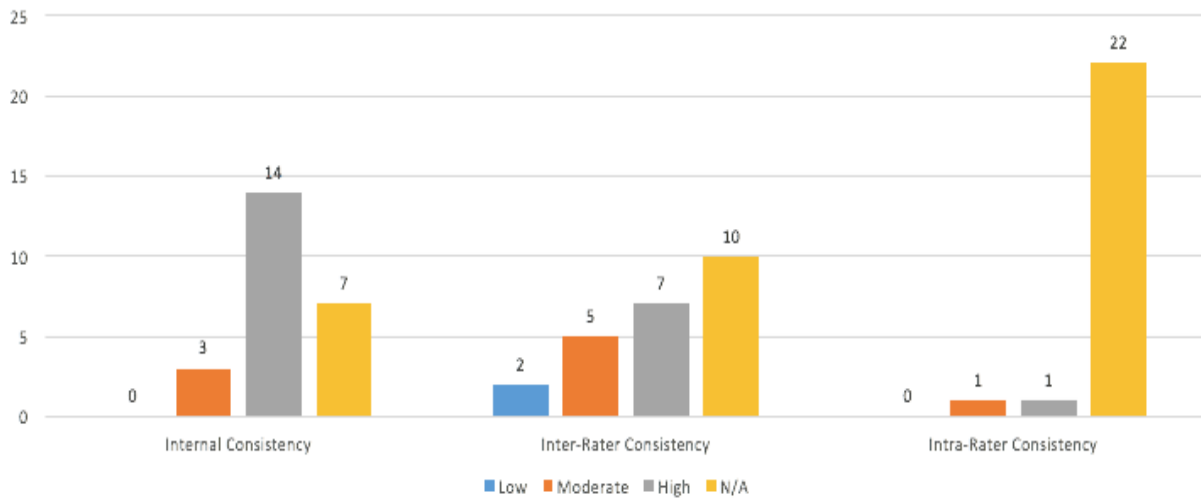
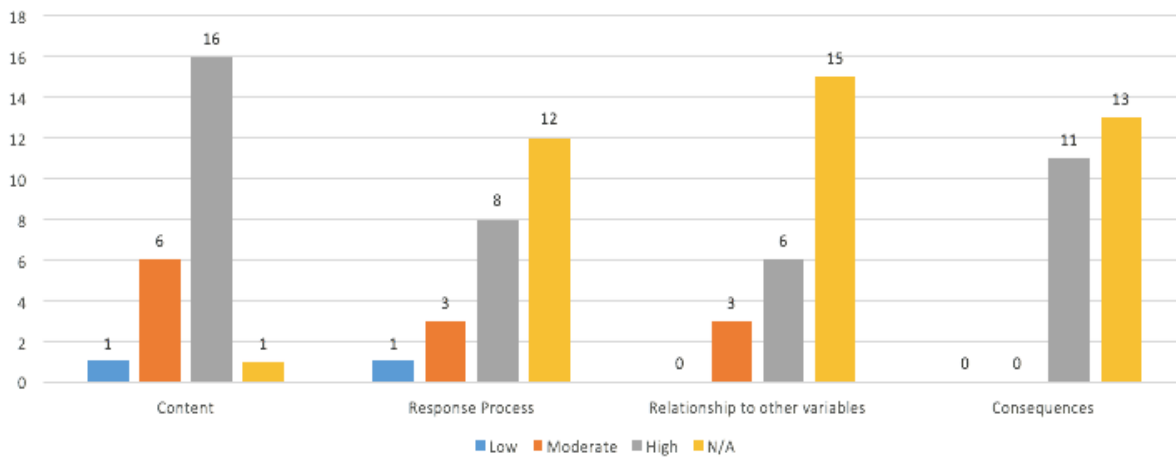


Figure 7: Validity assessment for the combined GRS and TSC



1.6 Discussion

The array of tools existing to assess surgical competency in residents is diverse and explores a variety of approaches. Newly developed tools, such as the automated analysis of motion tool and the virtual reality tool, integrate innovative technology to standardize, quantify and objectify assessments⁶¹. However, these methods are often costly, require specific equipment and might get quickly outdated⁶². They might not yet be suited for a widespread use in an everyday academic setting, because of lack of applicability. While a tool like simulation on animal models can model reality very closely, it also has the limitations of high costs, need for dedicated infrastructure, inevitable anatomical differences with human specimens and potential ethical issues⁶³. It might remain an interesting choice for teaching purposes, but might not be appropriate for sequential monitoring of resident proficiency. The written practical examination has the advantage of being very applicable, but tests better for knowledge than skill and thus, might be a good complementary tool to another form of technical evaluation. These potential limitations might explain why few studies address these various methods. Finally, the GRS/TSC tool can allow for the evaluation of surgical technical competencies in an operating room setting, is procedure-specific, applicable in an academic setting and does not require any special equipment⁶⁴. These characteristics combined with the relative simplicity of development might explain why it is overrepresented in our review. It should be noted, however, that one limitation of this study is the potential presence of a publication bias, which would favor an overrepresentation of assessment tools showing favorable results in terms of acceptability, validity and reliability. This could lead to overly optimistic conclusions about these assessment tool's performances and give a skewed appreciation of their performance. Also, many studies in our review used the checklists in simulation situations rather than actual operating room cases.

This is another potential limitation of this review as the results could introduce confounders. A simulation context does not include some of the specificities and challenges of live surgery.

As stated previously, the most prevalent techniques being the temporal bone dissection and the endoscopic sinus surgery. A possible explanation for this is that these procedures are relatively straightforward, with a commonly agreed upon set of steps to successfully complete the procedure, making them easier to standardize. Also, these procedures are both easily recorded on digital devices for delayed double blinded viewing by an independent assessor, rendering them amenable to serve in validity tests⁶⁴. However, it is interesting to note that none of the resident skill assessments were done on cases involving oncological procedures. It might be argued that such cases are often lengthy, intricate and delicate because of the aggressive nature of the disease, making them less amenable to straightforward skill assessments for procedures such as myringotomies or tonsillectomies. However, some basic oncologic surgical techniques, such as neck dissection, are within the scope of procedures OTL-HNS residents should master during their training.

1.7 Conclusions

Medical education, like the practice of medicine itself, is a field of constant evolution and reinvention. The recent changes in the format of residency give residents the opportunity to tailor their residency in order to meet their individual learning needs and skills. An array of assessment tools has been developed over the last years to meet the new evaluation needs, and to enable staff to monitor residents' progress. Five major means of assessing surgical competency have emerged in this review: the combined global rating scale (GRS) and task-specific checklist (TSC), the automated analysis of the surgeon's hand motions, simulations on animals. These tests have overall demonstrated high reliability and validity scores when analyzed according to Gail M. Sullivan's criteria. On its own, the combined GRS and TSC tool, which was used in over 83% of all assessment studies, presented validity and reliability results ranking predominantly in the high range, with publication biases being a possible limitation. It should be noted that while this tool has been frequently adapted for many fundamental surgeries of OTL-HNS, it has not yet been used for training in oncologic surgeries. Nevertheless, in the light of this study, many validated assessment methods are now available and may serve as templates on which to tailor new ways to assess resident progress and success in the future.

1.7 Acknowledgments

To Mrs Daniela Ziegler, librarian, healthcare reference and information specialist at Centre Hospitalier de l'Université de Montréal (CHUM) library and reference center, for her help in planning and executing a methodical and thorough review of paper and database references.

Chapitre 3 : Développer et valider un outil évaluant la compétence chirurgicale pour l'évidement cervical

1. Article: Assessment of Surgical Competency for Neck

Dissection: A Prospective Validation Study

1.1 Contribution des auteurs

1 Érika Mercier

- a) Conception et préparation du projet
- b) Développement de la grille d'évaluation
- c) Recrutement des participants
- d) Supervision du projet sur le terrain
- e) Compilation, analyse et interprétation des données recueillies
- f) Participation à la réalisation des analyses statistiques
- g) Rédaction et correction du manuscrit

2 Louis Guertin

- a) Développement de la grille d'évaluation
- b) Participation au recueil de données
- c) Révision, correction et approbation du manuscrit

3 Éric Bissada

- a) Développement de la grille d'évaluation
- b) Participation au recueil de données
- c) Révision, correction et approbation du manuscrit

4 Apostolos Christopoulos

- a) Développement de la grille d'évaluation
- b) Participation au recueil de données
- c) Révision, correction et approbation du manuscrit

5 Marie-Jo Olivier

- a) Développement de la grille d'évaluation
- b) Participation au recueil de données
- c) Révision, correction et approbation du manuscrit

6 Jean-Claude Tabet

- a) Développement de la grille d'évaluation
- b) Participation au recueil de données
- c) Révision, correction et approbation du manuscrit

7 Nathan Yang

- a) Participation à la réalisation des analyses statistiques
- b) Révision, correction et approbation du manuscrit

8 Tareck Ayad

- a) Conception et préparation du projet
- b) Développement de la grille d'évaluation
- c) Supervision du projet sur le terrain
- d) Révision, correction et approbation du manuscrit

Title:

Assessment of Surgical Competency for Neck Dissection: A Prospective Validation Study

Authors:

Érika Mercier¹, M.D.

Louis Guertin¹, MD, FRCSC

Eric Bissada¹, MD, FRCSC

Apostolos Christopoulos^{1,2}, MD, MsC, FRCSC

Marie-Jo Olivier³, MD, FRCSC

Jean-Claude Tabet¹, MD, FRCSC

Nathan Yang¹, M.D.

Tareck Ayad^{1,2,3} M.D. FRCSC

1. Centre Hospitalier de l'Université de Montréal (CHUM), 1051, Sanguinet Street, Montreal, Quebec, Canada, H2X 3E4
2. Centre de recherche du Centre Hospitalier de l'Université de Montréal, 900, Saint-Denis Street, Montreal, Quebec, Canada, H2X 0A9
3. Hôpital Maisonneuve-Rosemont, 5415, Assumption Boulevard, Montréal, Quebec, Canada, H1T 2M4

Corresponding author:

Dr Tareck Ayad

Centre Hospitalier de l'Université de Montréal

1051, Sanguinet St, Montreal, Quebec, Canada, H2X 0C1

Presented at the Canadian Society of Otolaryngology Annual Meeting 2018 as a podium presentation, in Quebec City, Canada

Winner of the national Poliquin Award for resident research 2018, Quebec City, Canada

Submitted for publication in the Journal of Otolaryngology-Head & Neck Surgery, Canada

1.2 Abstract

Background: Progressive implementation of the milestone competency-based curriculum has created a need for new objective and validated means to assess resident surgical proficiency. A previous systematic review of literature by the authors has highlighted a shortage of tools assessing surgical competency for oncologic procedures in Otolaryngology Head & Neck Surgery.

Methods: A procedure-specific assessment tool for neck dissection was developed using a modified Delphi method. The two-part design was modeled on the previously validated OSATS checklist. The tool was then validated through a one-year multicentric prospective study in collaboration with the residents and faculty from our academic center. Additionally, an online survey was developed to assess the acceptability by residents and staff before and after the validation studies.

Results: A total of twenty-nine 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 demonstrated significantly higher checklist scores for senior residents as opposed to 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 ($p < 0,05$).

Conclusions: The first tool assessing surgical competency in oncologic Otolaryngology Head & Neck Surgery has been successfully developed and validated.

No trial registration necessary.

Keywords: Resident, assessment tool, assessment method, evaluation, competency, neck dissection, applicability, reliability, validity.

1.3 Background

The future of Canadian Otolaryngology Head and Neck Surgery (OTL-HNS) will be shaped by the practice of the hundreds of surgeons throughout the country, who are in turn a product of the rigorous training they have received during residency. Adapting teaching methods to the challenges of modern day medicine and ensuring graduates achieve a consistent and dependable level of competency is essential to maintaining the highest standards of care and expertise. In keeping with these values, the Competency By Design (CBD) based curriculum was developed over the course of the last decades to ensure essential core competencies were consistently achieved, while tailoring residency to individual needs and progress⁶⁶. This improved cursus 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 effect⁶⁷. 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 feedback⁶⁸. 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 optimal progress.

In the face of these new challenges and requirements, the authors previously conducted a systematic review of existing literature to identify available methods of assessing surgical competency in OTL-HNS residents⁶⁹. Conclusions of this study have shown 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^{70,71,72}. Over the last decades, it has served as a template for the development of numerous contemporary surgical competency evaluation instruments^{74,75,76,77}. Another point highlighted

in the review was the absence of assessment tools developed to date for oncologic procedures, including neck dissection. In this era of implementation of the CBD initiative, it is of utmost importance to develop a variety validated of tools to assess resident competency.

This study's aim is to contribute a means of assessing surgical competency 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 competency. Specifically, the objectives of this study are to develop a task-specific checklist for neck dissection, then assess its validity, acceptability and feasibility.

1.4 Methods

1.4.1 Developing an assessment tool for neck dissection

Two complimentary checklists assessing different sets of surgical skills were created and combined to allow a comprehensive, detailed and systematic representation of surgical competency.

To build the global skill checklist, or GRS, the original questions and scoring system from the OSATS tool were retained (Annexe 1, Figure 8, p.II). This checklist lists seven criteria assessing residents' abilities to perform basic surgical tasks and is scored from one to five on a Likert scale, in increasing order of proficiency. Descriptive pointers are present on figures one, three and five. 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 center until a consensus was reached on the formulation and interpretation. The final GRS tool is presented in Figure 9 (Annexe 1, Figure 9, p.III).

The specific checklist, or TSC, enumerates the fundamental steps to performing a standard neck dissection. The authors employed 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 six subspecialized oncologic head and neck surgeons on service until a final consensus was reached. 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 center. The items were scored with a Likert type scale ranging from one to five. Descriptive pointers were developed to describe the skill level expected for each score, in order to help standardise 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 10 (Annexe 1, Figure 10, p. IV).

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. Confidentiality was required by the ethics committee and preferred by our team, as to encourage resident participation in the knowledge that their scores would not be divulged accidentally if a sheet was misplaced, and to avoid comparison in score performance between residents. We wanted participating residents to feel comfortable participating in the study and ensure it would not reflect negatively on them. For the same reason, the resident identification numbers were only disclosed to staff who were on a need-to-know basis in order to fill in the sheets. We did not want residents to be able to retrace scores if data was disclosed or found.

1.4.2 Validating the neck dissection competency 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 competency of participating residents completing a neck dissection in the operating room. A multicentric ethics committee approval was obtained at the two hospitals involved: Centre Hospitalier de l'Université de Montréal (CHUM) and Hôpital Maisonneuve-Rosemont.

1.4.2.1 Participants and timeframe

Participants included Université de Montréal OTL-HNS residents PGY-2 to PGY-5. First year residents were excluded from this study because of their limited involvement in neck dissection cases. All eligible residents were approached to be part of this project and received details about the study and the implications of their participation. They were made aware of the tool being validated and its use prior to the beginning of the study. The goal of this was to simulate real-life conditions in which the tool would be implemented. Residents would be made aware of the criteria on which they are being evaluated and encouraged to study and refer to the steps detailed the tool in order to progress and optimize learning opportunities before operating room cases. All six head and neck staff of the program collaborated on the project. The timeframe set to complete this study was the academic year 2016-2017, from July 2016 to June 2017.

1.4.2.2 Validity studies

Cumulated data was analysed by a certified statistician affiliated with the research center of the *Université de Montréal* hospital center (CRCHUM). Data analysis was carried out using SPSS v21. Chi-square tests were done to compare score frequency between junior and senior residents on the GRS and TSC tools, as well as on score progression over time. Comparisons of average resident scores by level of residency were done with Student's T-test. Statistical significance was set at $p < 0,05$ for all tests.

1.4.2.3 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 feedback and representation of competency. A Likert-style rating scale ranging from 1 to 10 allowed participants to score their level of agreement with each question, where 1 corresponding to

“statement very incorrect” and 10 “statement very accurate”. 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 to compare response progression. Mann-Whitney U tests were done to compare score progression, with $p < 0,05$ statistical significance.

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

1.4.2.4 Additional studies

Evaluators were asked to detail which levels of the dissection were addressed, to specify if the neck has previously received radiation therapy and to grade the case as standard or difficult.

1.5 Results

1.5.1 Descriptive data

A total of 29 neck dissections were performed by 11 residents over the course of one year. A slight predominance (62,1%) of cases were completed by junior residents (PGY-2 and PGY-3). Levels IA, IB, IIA, III and IV were simultaneously dissected addressed in more than half of the surgeries included in this study. An average of three levels were dissected per case. 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.

The usual sequence of events was that when a neck dissection case was scheduled, the surgeon and resident determined prior to starting the case if there would be an assessment. If they agreed to go forward with the evaluation, the resident was allowed to review the steps and the checklist if desired. During the case, the resident was encouraged to complete as much as possible of the case autonomously, as described earlier. In most cases, at the end of the procedure the surgeon would fill out the checklist immediately, then meet privately with the resident to proceed with feedback. In a few cases, the checklist was filled out in the following days, and the resident was met soon after in a similar way. The sequence was repeated for every neck dissection case.

Residents received direct feedback on their performance in almost 100% of cases. Feedback consisted of a face to face discussion between staff and resident for between five to ten minutes, where each item on the GRS and TSC checklists was enumerated with a short explanation of the given score. This gave the opportunity for the assessor to objectively stress the resident's strengths and weaknesses and allowed the latter to ask questions and solicit tips

and pointers. The encounter was usually very positive and allowed for individualised and oriented guidance.

1.5.2 Validity studies

1.5.2.1 Score progression with seniority

Resident scores were dichotomised into juniors (PGY-2 and PGY-3) and seniors (PGY-4 and PGY-5). 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 11 p.58 and Figure 12 p.58). For both checklists, a score progression with seniority can be observed with junior residents obtaining mainly scores of 3 and 4, while senior residents score mostly 4 and 5. In both the GRS and TSC tools, the difference in score frequency was significantly different on scores 2, 3, 4 and 5 ($p < 0,05$).

Figure 11: Results of junior and senior residents on the GRS checklist

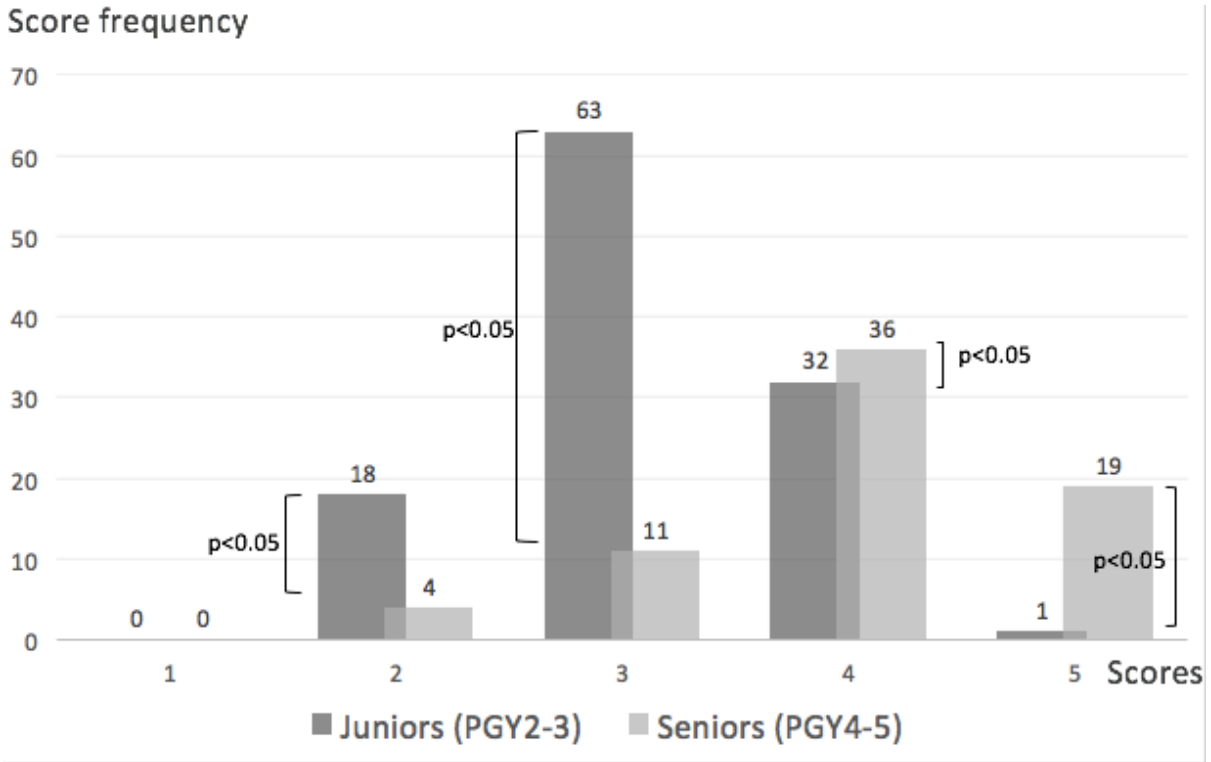
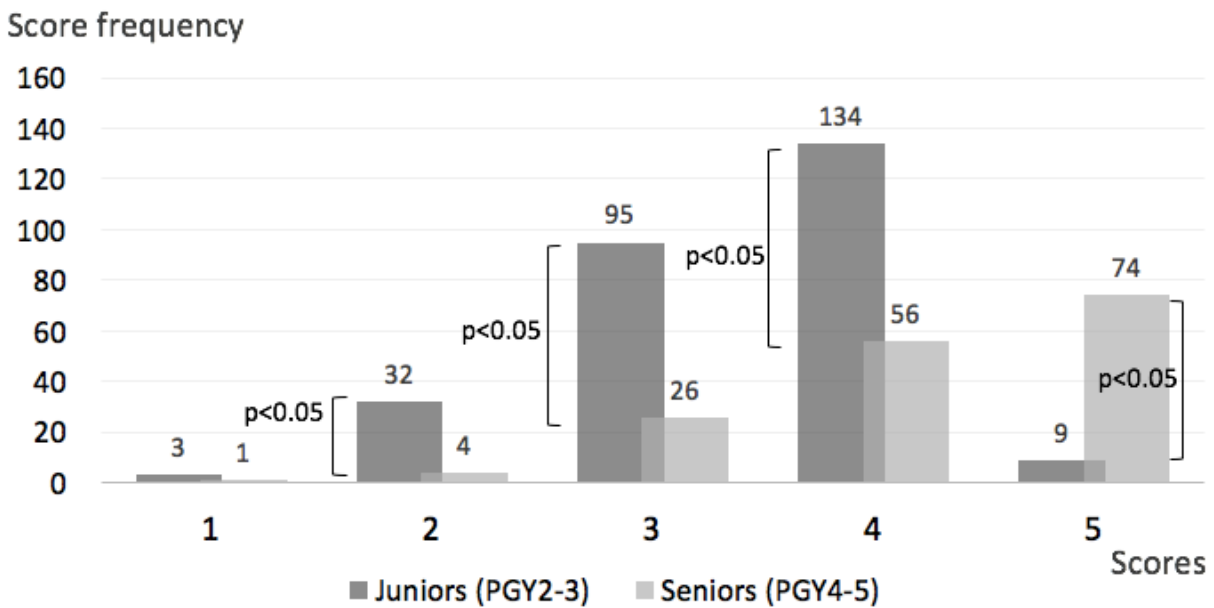


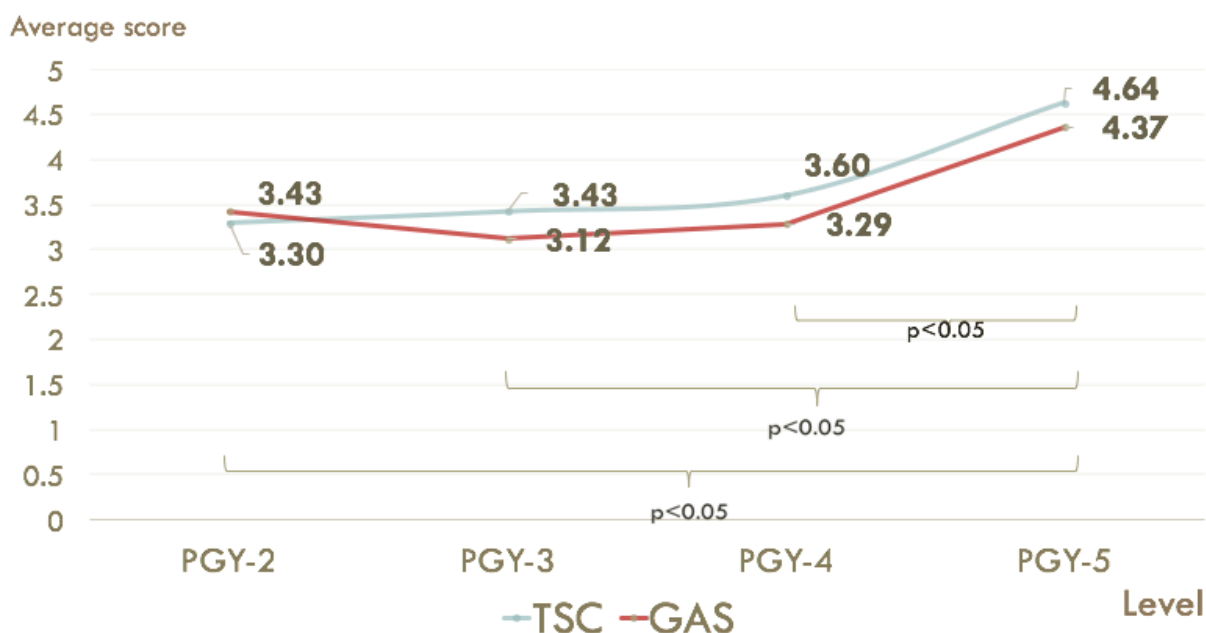
Figure 12: Results of junior and senior residents on the TSC checklist



1.5.2.2 Score progression by level of residency

The average scores by level of residency were compared to one another. Significant differences were found between the scores of PGY-2 and PGY-5, between PGY-3 and PGY-5 and between PGY-4 and PGY-5, on both GAS and TSC ($p < 0,05$) (Figure 13 p.60). Comparing scores of PGY-2, PGY-3 and PGY-4 did not show any statistically significant differences. However, the curve plotting average results by residency level does show a progression for every year of residency. It is possible that there is an improvement in scores through residency levels, but that the power in this study is not sufficient to allow a significant difference to be seen. It is also possible that because residents have access to a greater number of surgical cases with increasing level of residency, PGY-5 might see their learning curve increase faster than other levels because of increased exposition. They may also have more experience and might learn at a faster rate, making their scores increase significantly faster. The average scores on the GRS and TSC checklists were also plotted and compared by residency level. The resulting curves follow the same trend closely, and after a Student T-test, no significant difference could be found between the average scores on each checklist for the same level of residency ($p < 0,05$).

Figure 13: Average scores by year of residency for GRS and TSC checklists

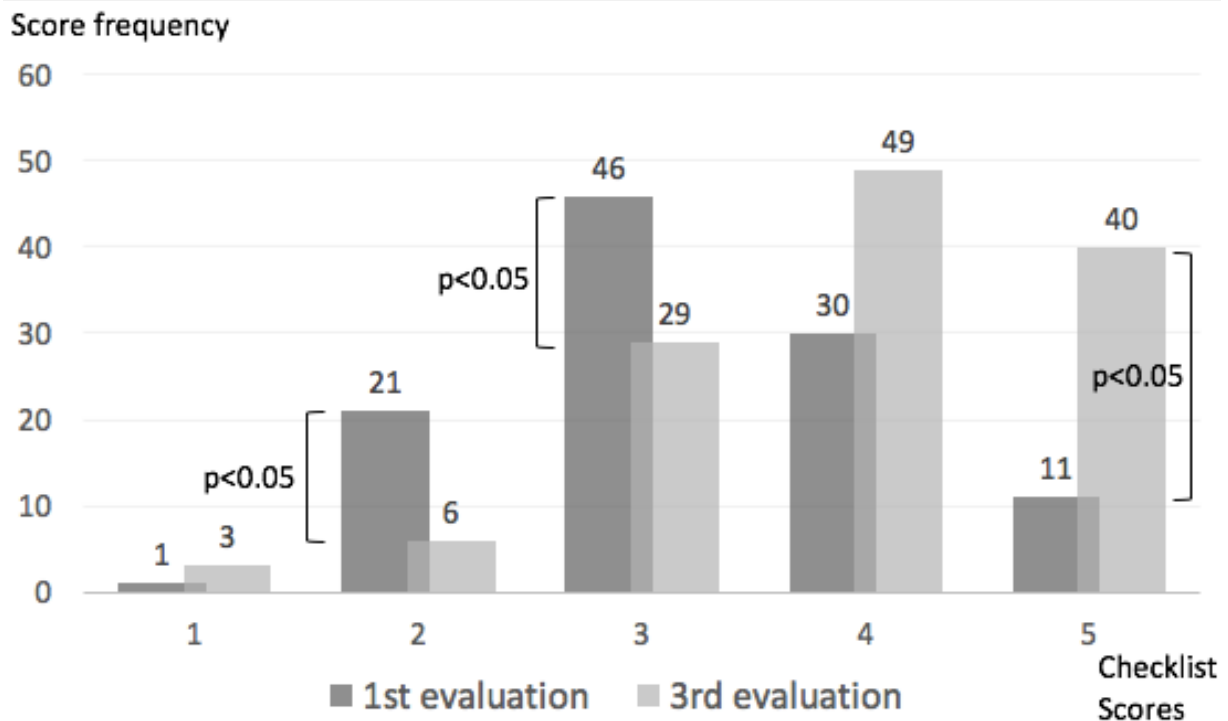


1.5.2.3 Score progression in time

Resident score progression was assessed by comparing scores obtained on their first evaluation, with their third evaluation, approximately one month later. It is not possible to say exactly how many neck dissections each resident had performed in between these recorded assessments, as not all neck dissections performed were included in this study. This was either because of time constraints, lack of habit within the program or in some cases, many residents were asked to work on the surgical case, an exclusion criterion for our study. Residents could thus have progressed unequally in between recorded assessments, but we tried to maintain a similar timeframe to approximate progression. For this study, junior and senior results were pooled and score frequencies were compared over time (Figure 14 p.61). A clear progression can be observed with residents scoring mostly 2, 3 and 4 on their first evaluation and quickly

progressing to 3, 4 and 5 ratings on the subsequent third assessment. Differences in score frequency were statistically significant for scores 2,3 and 5 ($p<0,05$).

Figure 14: Resident score progression



1.5.2.4 *Autonomy*

Residents were judged able to perform the procedure independently in 24,1% of cases, all of which were performed by PGY-5 residents. Alternatively, all PGY-5 residents did achieve autonomy.

1.5.3 **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 in their original French format, as well as an English translation (Annexe 2, Table I and

II, p.V-VII). No significant differences in responses were identified for residents when comparing pre and post study responses ($p < 0,05$). We also compared post-study resident response with staff responses. Residents and staff differed in opinion only as to the summative use of the tool, with residents less prone to use this tool for summative purposes ($p < 0,05$). There was no statistically significant difference for any other questions.

1.5.4 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% of cases (28 of the 29 the evaluations performed), in the way described previously. This demonstrates, that feedback was generally easy to include in the process, especially when done immediately after filling out the checklist.

1.6 Discussion

A prior review of literature by the authors has shown that the combined checklist tool was the most frequent method used to assess surgical competency for OTL-HNS procedures. These OSATS-based instruments being easy to implement, low in cost and overall simple to use⁵¹, a similar methodology was followed to develop the tool for assessment of surgical competency in neck dissection.

One of the main goals of this 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 allow an objective means for staff and program directors to keep track of resident progression to offer focused assistance if required. We believe we have succeeded in creating a tool that can measure resident progression in a reliable and standardized fashion. Our results show that our tool has the sensitivity to demonstrate significant progression in score results with increasing resident seniority, hence demonstrating the tool's construct validity. The tool also demonstrates rapid and significant resident score progression over a relatively short period of time. On their third assessment, residents had visibly 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 if they did not match their expectations, and received the necessary guidance to improve on any areas of weakness. Another point that must be mentioned, is the possibility of a learning bias⁷⁸, which implies that the residents' scores might have partially improved as a result of them getting habituated to using the tool or having learned or memorised parts of the questionnaire, allowing them to anticipate steps. This type of bias would be desirable in this study as it could help residents in achieving their goals and increasing scores. Furthermore, the results on the TCS checklist correlate with the results obtained on the previously validated GRS checklist,

following the same trend closely, supporting the accuracy of the newly developed TCS 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 competency.

Another objective of this study was to ensure that it displayed reasonable acceptability in our academic centers. We believe this to be the case, as anonymous surveys to residents and staff have shown that residents did not have an apprehensive attitude towards the evaluations and did not feel it 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. The only divergence in results involved the 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 by the average time to complete the tool, which was under 5 minutes and the percentage of feedback given to the resident by the attending, which was 96,6%. Its use in between operating room cases is hence 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 et al. in a similar study⁸⁰.

Feedback rates were excellent in this 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 in itself to be of invaluable help to residents, who benefit from this insight and time to reflect on the case. It could be argued that resident progression might be a result of regular feedback by staff instead of direct use of the tool, but since one of the major goals of the neck assessment dissection tool was to increase feedback, we believe that the end goal of resident skill progression is achieved.

There are some inherent limitations to this study. The first is the possibility of a Hawthorne effect, where the behavior of the subject is altered by the awareness of being observed. Also,

attendings were not blinded to the identity of the subjects, which might induce a bias related to preconceived ideas and impressions of the individual being evaluated (halo or horn effect). Also, different attendings might offer different assistance or guidance through the cases, depending on their personal approach, despite our attempt to standardise their implication through detailed instructions. These instructions explain that residents should be given the most autonomy possible within reasonable time and security constraints. They should be allowed to complete as many steps as possible and position their assistant themselves. The document detailing these instructions is available in the annex and was given to all participating staff for reference (Annex 3, Document 2, p. X). The same was done with participating residents (Annex 3, Document 3, p.XI). These limitations could have been prevented by having the procedures filmed and then scored anonymously afterwards, blinding the assessor. However, we felt that despite the imperfections of this method, it is the situation that best reproduces the actual conditions in which the tool will be used, and thus better reflects reality. Moreover, a recent study has shown that there was a high inter-rater reliability among blinded (use of video recordings for assessment) and nonblinded (direct observation in the operative room) which facilitates the evaluation of residents and avoids the complexity of arranging video recordings and having blinded raters⁸¹. Another potential bias in this study could be attributed to the fact that the same unblinded staff were given the task of assessing residents and giving out feedback, which could lead to them either consistently under or over score residents based on subjective perception or as a result of their own variable level of comfort with the assessment process. However, as explained previously, it was not possible and entirely desirable for us to use video footage, and although it is possible that staff evaluation scores may vary over time, the objectivity of the checklist criteria limit this to some extent. It would be interesting to include intra-rater reliability studies to assess this, but it was not possible due to the small sample size.

Delays between the surgery and the evaluation can induce a recall bias, where the surgeon might not exactly 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. Another important limitation is the small number of cases recorded which might have affected statistical significance for some analysis. Power calculations and analyses could not be done for this study as we were working with a convenience sample.

Additional study would be necessary to complete these studies with inter and intra-rater reliability studies and to gain long term insight on the use of the tool.

1.7Conclusions

We have successfully developed the first tool designed to assess surgical competency for neck dissection. This tool showed good construct validity, was demonstrated acceptable by residents and staff and finally feasible in an academic setting. To allow a broader use of the tool, a second similar study is ongoing to validate the tool in English. Implementing the regular use of these kinds of tools promotes resident progression by encouraging case discussion and stimulating consistent individualised support. This is in line with the objectives of the milestone curriculum, where programs are tailoring to individual needs and progression. We believe that opening the discussion and encouraging a dialogue between mentors and learners is the first step in helping tomorrow's surgeons achieve competency.

1.8 Acknowledgments

The authors would like to acknowledge the implication of Paule Bodson-Clermont and Mylène Baptista, for their work and guidance on statistical analysis and interpretation. A special thank you to Nathan Yang for his help and involvement in statistical revisions.

Chapitre 4: Discussion

1. Limites du projet

1.1 Biais d'inclusion

Malgré tous nos efforts pour respecter les plus hauts standards de méthodologie dans la conception et la réalisation de ce projet, certaines limites sont inhérentes à sa mise en œuvre. Nous avons identifié certains biais potentiels que nous voulons exposer. Tout d'abord, il existe un potentiel de biais d'inclusion lié à l'échantillonnage des participants à notre projet. Nous avons procédé à un « échantillonnage de convenance » en choisissant d'inclure seulement les résidents en ORL-CCF de l'Université de Montréal comme participants à l'étude. Ce type d'échantillonnage exclut la composante de probabilité et met plutôt à profit l'accessibilité d'une certaine population⁸³. Ces échantillonnages sont souvent utilisés dans un contexte d'études pilotes ou comme première approche à un sujet. Ceci nous a permis d'avoir accès à un nombre convenable de participants, en limitant les difficultés logistiques et d'uniformisation associées à la mise en œuvre d'un projet de recherche dans différents centres ou différentes villes. Cependant, cette méthode a pour conséquence que tous les étudiants proviennent du même centre et composent un échantillonnage plutôt uniforme et étroit sur le plan démographique. Nous obtenons ainsi une appréciation des tendances pour notre population locale, mais les résultats de l'étude pourraient être difficiles à extrapoler à des populations différentes ou élargies. Dans le contexte où notre étude se faisait dans le but d'un usage local de l'outil chez une population québécoise francophone, nous croyons qu'il s'agit d'un échantillon approprié à cette fin.

1.2 Étude sans double insu

D'autres limites de notre projet découlent du fait que les évaluations réalisées avec la nouvelle grille n'ont pas été faites à double insu. Les évaluateurs provenaient du même centre que les résidents évalués et les deux groupes étaient familiers avant le début de l'étude. Ceci peut donner lieu à un biais de type « halo or horn effect » ou « ange ou démon », qui est un type de biais cognitif où le jugement d'une personne peut être influencé par des idées préconçues ou des impressions non fondées en lien avec une personne. L'évaluateur pourrait ainsi être influencé par les performances précédentes du résident pour la même chirurgie ou pour d'autres procédures, par son niveau de résidence ou par des facteurs personnels.

De la même façon, la performance du résident évalué peut être influencée par le fait qu'il ait conscience de la présence son évaluateur. Ceci correspond au « Hawthorne effect » où un individu modifie sa conduite suite à la réalisation qu'il est observé et correspond à un type de biais d'observation. Par exemple, le résident pourrait se montrer plus anxieux ou faire des erreurs inattendues du fait de se savoir examiné. Nous ne croyons cependant pas que ce biais prenne une place importante dans notre étude, puisque les résidents sont toujours sous observation au bloc opératoire et qu'il s'agit des mêmes évaluateurs avec lesquels ils sont habitués de travailler. Il y a donc peu de chances que leur performance soit significativement différente qu'à l'habitude. De plus, nous jugeons que malgré que cette situation soit imparfaite, elle représente le mieux les conditions réelles d'application de l'outil dans le futur et reflète donc davantage la réalité.

Un troisième effet découlant du fait d'avoir réalisé une étude non à double insu est la possibilité d'induire un biais d'intervieweur. Ceci survient lorsque l'intervieweur ou dans ce cas-ci l'évaluateur induit un changement dans les réponses ou le comportement du répondant

ou du résident évalué par sa façon de mener l'évaluation⁸⁴. Ainsi, si un évaluateur intervient excessivement d'une manière pouvant influencer positivement ou négativement la performance du sujet évalué, il introduit alors un biais dans l'étude. Nous avons tenté de limiter la possibilité d'un tel biais et standardiser les évaluations en distribuant aux évaluateurs des instructions claires quant à la quantité d'encadrement et d'assistance à apporter aux résidents lors des évaluations. Ces instructions, détaillées précédemment, sont disponibles dans les annexes (Annexe 3, Documents 2 et 3, p.X-XI).

1.3 Biais de rappel

Dans l'éventualité où une longue période de temps s'écoulerait entre la réalisation d'une évaluation et la complétion de la grille, un biais de rappel pourrait s'installer. Ceci survient lorsqu'un participant à l'étude oublie des éléments d'information de façon à modifier ses réponses⁸⁵. Ici, un évaluateur pourrait oublier les détails de certaines étapes de la chirurgie ou se confondre avec d'autres procédures survenues entretemps. Afin d'éviter que ceci survienne, nous avons demandé aux évaluateurs de compléter les grilles le plus rapidement possible après la chirurgie, idéalement dans un délai de 48 heures.

1.4 Traduction de la grille OSATS

Une limite additionnelle de notre étude réside dans le fait que nous avons utilisé une version traduite en français de la grille générale de l'OSATS pour développer notre outil d'évaluation. La grille OSATS a été validée en anglais, mais ne l'a pas encore été en français. Comme nous voulions rendre notre outil aussi accessible que possible pour l'utilisation locale dans notre centre, nous avons pris l'initiative de la faire traduire par des services professionnels de traduction. La traduction française a été révisée par les membres du groupe d'oncologie afin d'assurer une clarté et un maintien de la signification globale. Cette nouvelle version n'a pas été

validée seule dans notre étude, mais ceci ne faisait pas partie de nos objectifs initiaux. L'objectif principal était de valider conjointement la grille générale avec notre liste spécifique originale en français, puisque c'est ce format qui sera utilisé en pratique.

1.5 Biais d'apprentissage

Étant donné que les résidents avaient connaissance de la grille à l'étude et des critères formant la base de leur évaluation, il est possible qu'un certain biais d'apprentissage puisse teinter les résultats. Les résidents pourraient voir une partie de leur score s'améliorer en raison d'un rappel ou d'une certaine mémorisation des critères et du processus d'évaluation, les rendant meilleurs au fil des évaluations par simple habitude. Cependant, ce biais a un impact positif sur leur progrès et serait même souhaitable. Nous encourageons d'ailleurs les résidents à se référer aux critères d'évaluation au cours de l'étude dans le but de les aider à progresser et nous croyons que cet effet participe positivement à l'apprentissage des résidents, malgré le fait qu'il puisse contribuer à surestimer l'effet de la grille seule.

2. Suites du projet

2.1. Version anglophone

Un projet subséquent est de procéder à la traduction anglophone de notre outil et d'en faire la validation d'une manière similaire à la validation francophone. Comme la majorité de la communauté médicale communique en anglais, l'existence d'une version anglophone de l'outil aurait le potentiel de rejoindre une vaste portion de la population du Canada et d'ailleurs dans le monde. Ceci aurait pour effet d'augmenter significativement la validité externe de notre étude.

2.2. Participation d'autres centres

Une prochaine étape à cette étude serait de valider notre outil d'évaluation avec la participation d'un plus grand nombre de centres académiques. Ceci augmenterait le nombre de participants, le nombre d'évaluations remplies et permettrait d'avoir une perspective plus globale avec un échantillonnage diversifié. Cette étape pourrait être réalisée en collaboration avec les autres programmes de médecine du Québec dans le cas de la grille francophone et avec d'autres centres canadiens ou l'Université McGill dans le cas de la grille anglophone.

2.3. Étendre à d'autres procédures

Une autre suite potentielle à ce projet est l'extrapolation du même processus de création et de validation à d'autres procédures chirurgicales de l'ORL-CCF. Plusieurs autres procédures, entre autres dans le champ de l'oncologie, demeurent encore inexplorées. Quelques exemples incluent la laryngectomie, la parotidectomie, les microchirurgies laryngées, les lambeaux locaux et régionaux et l'exérèse de lésions cutanées. Un processus similaire à celui employé pour l'évidement cervical pourrait aisément être appliqué pour créer des outils propres à ces chirurgies, dans le but éventuel d'avoir des outils spécialisés et validés pour toutes les procédures évaluées dans le cadre de la résidence en ORL-CCF.

Chapitre 5: Conclusion

La réalisation de ce travail de maîtrise nous a permis de faire le point sur les moyens à la disposition des programmes académiques d'ORL-CCF pour remplir les nouvelles recommandations découlant de la réforme en éducation médicale. Afin de faciliter ce processus, nous avons développé le premier outil évaluant les compétences chirurgicales des résidents en ORL-CCF pour l'évidement cervical. Nous avons démontré la validité et l'acceptabilité d'un tel outil en français, et croyons qu'il pourrait être intégré comme outil de référence en centre académique, encourageant une discussion éducationnelle ciblée entre résidents et patrons.

La prochaine étape de ce projet consistera à valider le même outil en anglais et ainsi en permettre un usage pan national et même international. La transition vers l'apprentissage par compétence est en plein essor et continuera sans doute de prendre racine à travers le monde. Comme l'un des premiers pays impliqués dans le mouvement, nous nous trouvons à l'avant-garde du processus d'implémentation et d'adaptation. Par conséquent, il en revient de notre responsabilité de concevoir et développer les ressources qui permettront et à nos institutions de faire la transition avec succès, ouvrant la voie à d'autres après nous. Par le travail décrit dans ces pages, nous nous rapprochons un peu plus de ce but et plaçons les fondations d'un programme qui encouragera la réussite et la compétence chez nos médecins de demain.

Bibliographie

1. Bârsu, C. (2017). History of Medicine between Tradition and Modernity. *Clujul Medical*. Iuliu Hatieganu University of Medicine and Pharmacy, Repéré à www.ncbi.nlm.nih.gov/pmc/articles/PMC5433581/
2. Moll, W. (1968). History of American Medical Education. *Medical Education*. Repéré à <https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2923.1968.tb01765.x/pdf>
3. Fulton, JF. (1953). History of Medical Education. *British Medical Journal*. Repéré à <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2029428/>
4. Taradejna, C. (2007). History of Medical Education. Repéré à <http://www.acgme.org/About-Us/Overview/History-of-Medical-Education>
5. Bailey, P. Douglas W. (2018). Medical Education. *The Canadian Encyclopedia*. Repéré à www.thecanadianencyclopedia.ca/en/article/medical-education/
6. Carnegie Foundation for the Advancement of Teaching. (2018). Foundation History. Repéré à www.carnegiefoundation.org/about-us/foundation-history/
7. Howell, JD. (2016). A History of Medical Residency. *John Hopkins University Press*, Repéré à muse.jhu.edu/article/612115/pdf
8. Centers for Medicare & Medicaid Services. (2018). History. Repéré à www.cms.gov/About-CMS/Agency-information/History/
9. New England Journal of Medicine. The Next GME Accreditation System - Rationale and Benefits. *Oxford University Press*, Repéré à www.nejm.org/doi/10.1056/NEJMSr1200117?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub=www.ncbi.nlm.nih.gov

10. ACGME. Common Program Requirements. Repéré à www.acgme.org/What-We-Do/Accreditation/Common-Program-Requirements
11. ACGME. (2017) ACGME Common Program Requirements. *Common Program Requirements*, Repéré à www.acgme.org/Portals/0/PFAssets/ProgramRequirements/CPRs_2017-07-01.pdf
12. Rosner, L. (1992). Thistle on the Delaware: Edinburgh Medical Education and Philadelphia Practice, *Social History of Medicine*, Volume 5, Issue 1, Pages 19–42, Repéré à <https://academic.oup.com/shm/article-abstract/5/1/19/1617087>
13. Wood, DF. (2003). Problem based learning. *British Medical Journal*, Repéré à www.ncbi.nlm.nih.gov/pmc/articles/PMC1125189/
14. NEJM Knowledge. (2016). Exploring the ACGME Core Competencies (Part 1 of 7). Repéré à knowledgeplus.nejm.org/blog/exploring-acgme-core-competencies/
15. NEJM. (2018). The Next GME Accreditation System - Rationale and Benefits. Repéré à <https://www.nejm.org/doi/full/10.1056/NEJMSr1200117>
16. Austin, M. (2013). The Next Accreditation System: A Resident Perspectiv. *The Next Accreditation System: A Resident Perspectiv*, Accreditation Council for Graduate Medical Education, Repéré à www.acgme.org/portals/0/pdfs/resident-services/9nasresidentsmay2014.pdf
17. Holmboe, Eric, et al. (2016). The Milestones Guidebook Version 2016. *ACGME*, Repéré à www.acgme.org/Portals/0/MilestonesGuidebook.pdf
18. Holmboe, ES, (2010). The Role of Assessment in Competency-Based Medical Education. *Med Teach*. Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/20662580>

19. Weggemans, M. et al. (2017). The Postgraduate Medical Education Pathway: An International Comparison. *GMS Journal for Medical Education*. Repéré à <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5704606/>
20. Rotgans JI. (2012). The themes, institutions, and people of medical education research 1988–2010: content analysis of abstracts from six journals people of medical education research. *Adv Heal Sci Educ*. Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/21971993>
21. Jaarsma D, Scherpbier A, Van der Vleuten C, Ten Cate O. (2013). Stimulating medical education research in the Netherlands. *Med Teach*. Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/23281880>
22. Ten Cate O. (2007). Medical education in the Netherlands. *Med Teach*. Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/18236272>
23. Australian Medical Council. AMC Examinations (Standard pathway), Repéré à www.amc.org.au/assessment/pathways/standard/exams
24. Bundesärztekammer. (2018). About the German Medical Association. Repéré à www.bundesaerztekammer.de/weitere-sprachen/english/german-medical-association/
25. Deutschland.de. (2018). How to Become a Medical Doctor in Germany. Repéré à www.deutschland.de/en/topic/knowledge/how-to-become-a-medical-doctor-in-germany
26. Conseil National De L'Ordre Des Médecins. Etudes De Médecine En France. Repéré à www.conseil-national.medecin.fr/article/les-etudes-de-medecine-en-france-400
27. Hôpitaux universitaires Genève. (2016). Devenir Médecin. Repéré à www.hug-ge.ch/medecins

28. Medical Residency Database. (2016) Switzerland. Repéré à www.residencydatabase.com/switzerland/
29. ISFM Institut Suisse Pour La Formation Médicale Postgraduée Et Continue. (2013). Spécialiste En Oto-Rhino-Laryngologie Programme De Formation Postgraduée. *Spécialiste En Oto-Rhino-Laryngologie*. Repéré à www.fmh.ch/files/pdf21/orl_version_internet_fl.pdf
30. Royal College of Physicians and Surgeons of Canada. Mise en oeuvre de la CPC. Repéré à www.royalcollege.ca/rcsite/cbd/competence-by-design-cbd-f
31. McMaster University. (2018). Competency by design. *McMaster Ranked Top Canadian University for Health*, Repéré à fhs.mcmaster.ca/main/news/news_2018/Competence_by_Design_The_transition_to_competency_based_medical_education.html
32. Royal College of Physicians and Surgeons of Canada. À propos de CanMEDS. Repéré à www.royalcollege.ca/rcsite/canmeds/about-canmeds-f
33. Ahmmad, Z. (2011). Assessment Methods for Surgical Skill. *World Academy of Science, Engineering and Technology*, International Journal of Biomedical and Biological Engineering, Repéré à waset.org/publications/2872/assessment-methods-for-surgical-skill
34. Royal College of Physicians and Surgeons of Canada. (2017). Applying competency-based practices to residency education. Repéré à <http://www.carms.ca/en/rcpsc/>
35. Resident Doctors of Canada. (2017). Competency-Based Medical Education. Repéré à <http://residentdoctors.ca/training/cbme/>
36. Mount, CA, et al. (2014). An End-of-Year Oral Examination for Internal Medicine Residents: An Assessment Tool for the Clinical Competency Committee. *Current*

- Neurology and Neuroscience Reports*. U.S. National Library of Medicine, Repéré à www.ncbi.nlm.nih.gov/pubmed/25210583
37. ACGME. (2017). Milestones. Repéré à <http://www.acgme.org/What-We-Do/Accreditation/Milestones/Overview>.
 38. Royal College of Physicians and Surgeons of Canada. (2017). Rationale why CBD. Repéré à <http://www.royalcollege.ca/rcsite/cbd/rationale-why-cbd-e>
 39. ISCP. (2017). Procedure Based Assessment (PBA). https://www.iscp.ac.uk/curriculum/surgical/assessment_pba.aspx
 40. Lin SY, Laeeq K, Ishii M, et al. (2009). Development and pilot-testing of a feasible, reliable, and valid operative competency assessment tool for endoscopic sinus surgery. *Am J Rhinology Allergy*. Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/19490815>
 41. Stack BC Jr., Siegel E, Bodonner D, et al. (2010). A study of resident proficiency with thyroid surgery: creation of a thyroid-specific tool. *Otolaryngology Head Neck Surgery*. Repéré à <https://journals.sagepub.com/doi/abs/10.1016/j.otohns.2010.02.028>
 42. Ahmed A, Ishman SL, Laeeq K, et al. (2013). Assessment of improvement of trainee surgical skills in the operating room for tonsillectomy. *Laryngoscope*. Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/23483535>
 43. Brown DJ, Thompson RE, Bhatti NI. (2008). Assessment of operative competency in otolaryngology residency: Survey of US Program Directors. *Laryngoscope*. Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/18641530>
 44. Sullivan GM. (2011). A Primer on the Validity of Assessment Instruments. *J Grad Med Educ*. Repéré à <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3184912/>
 45. Mount, C A, et al. (2014). An End-of-Year Oral Examination for Internal Medicine Residents: An Assessment Tool for the Clinical Competency Committee. *Current*

- Neurology and Neuroscience Reports*. U.S. National Library of Medicine, Repéré à www.ncbi.nlm.nih.gov/pubmed/25210583
46. Martin JA, Regehr G, Reznick R, et al. (1997). Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg*. Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/9052454>
47. Jordan A, Antomarchi J, Bongain A, et al. (2016). Development and validation of an objective structured assessment of technical skill tool for the practice of breech presentation delivery. *Arch Gynecol Obstet*. Repéré à <https://link.springer.com/article/10.1007%2Fs00404-016-4063-4>
48. Noureldin YA, Elkoushy MA, Aloosh M, et al. (2016). Objective Structured Assessment of Technical Skills for the Photoselective Vaporization of the Prostate Procedure: A Pilot Study. *J Endourol*. Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/27198163>
49. Noordin S, Allana S. (2015). OSATS for total knee replacement: Assessment of surgical competence in the operating room. *J Pak Med Assoc*. Repéré à <https://pdfs.semanticscholar.org/de2a/f15854592f7f39febc756b621993e6d68441.pdf>
50. MacEwan MJ, Dudek NL, Wood TJ, et al. (2016). Continued Validation of the O-SCORE (Ottawa Surgical Competency Operating Room Evaluation): Use in the Simulated Environment. *Teach Learn Med*. Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/26787087>
51. Niitsu H, Hirabayashi N, Yoshimitsu M, et al. (2013). 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*. Repéré à <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3574562/>

52. Obeid AA, Al-Qahtani KH, Ashraf M, et al. (2014). Development and testing for an operative competency assessment tool for nasal septoplasty surgery. *Am J Rhinol Allergy*. Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/25197910>
53. Ahmidi N, Poddar P, Jones JD, et al. (2015). Automated objective surgical skill assessment in the operating room from unstructured tool motion in septoplasty. *Int J Comput Assist Radiol Surg*. Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/25895080>
54. Awad Z, Taghi A, Sethukumar P, et al. (2015). Construct validity of the ovine model in endoscopic sinus surgery training. *Laryngoscope*. Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/25200556>
55. Kerwin T, Wiet G, Stredney D, et al. (2012). Automatic scoring of virtual mastoidectomies using expert examples. *Int J Comput Assist Radiol Surg*. Repéré à <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3403744/>
56. Jabbour N, Dobratz EJ, Dresner HS, et al. (2011). Assessing fundamental 2-dimensional understanding of basic soft tissue techniques. *J Surg Educ*. Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/21481800>
57. Awad Z, Hayden L, Robson AK, et al. (2015). Reliability and validity of procedure-based assessments in otolaryngology training. *Laryngoscope*. Repéré à <https://www.ncbi.nlm.nih.gov/m/pubmed/25346528/>
58. Ahmed A, Ishman S L, Laeeq K, et al. (2013). Assessment of improvement of trainee surgical skills in the operating room for tonsillectomy. *Laryngoscope*. Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/23483535>
59. Ishman SL, Benke JR, Johnson KE, et al. (2012). Blinded evaluation of interrater reliability of an operative competency assessment tool for direct laryngoscopy and rigid

- bronchoscopy. *Arch Otolaryngol Head Neck Surg.* Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/22986639>
60. Diaz Voss Varela DA, Malik MU, Thompson CB, et al. (2012). Comprehensive assessment of thyroidectomy skills development: a pilot project. *Laryngoscope.* Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/22183631>
61. Ahmidi N, Ishii M, Fichtinger G, et al. (2012). An objective and automated method for assessing surgical skill in endoscopic sinus surgery using eye-tracking and tool-motion data. *Int Forum Allergy Rhinol.* Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/22696449>
62. Oussi N, Loukas C, Kjellin A, et al. (2017). Video analysis in basic skills training: a way to expand the value and use of BlackBox training? *Surg Endosc.* Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/28664435>
63. Simkin, DJ, et al. (2017). The Death of Animals in Medical School. *N Engl J Med.* Repéré à <https://www.nejm.org/doi/full/10.1056/NEJMp1612992>
64. Laeeq K, Bhatti NI, Carey JP, et al. (2009). Pilot testing of an assessment tool for competency in mastoidectomy. *Laryngoscope.* Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/19885831>
65. Laeeq K, Infusino S, Lin SY, et al. (2010). Video-based assessment of operative competency in endoscopic sinus surgery. *Am J Rhinol Allergy.* Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/20021742>
66. Swing SR. (2007). The ACGME Outcome Project: retrospective and prospective. *Med Teach.* Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/18236251>
67. Bhatti, N. I., Cummings C. W. (2007). Competency in surgical residency training: defining and raising the bar. *Acad Med.* Repéré à

<https://www.ncbi.nlm.nih.gov/pubmed/17525542>

68. Brown, D. J., et al. (2008). Assessment of operative competency in otolaryngology residency: Survey of US Program Directors. *Laryngoscope*. Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/18641530>
69. Mercier E, et al. (2017). Objective Assessment of Technical Skills in Otorhinolaryngology–Head and Neck Surgery Residents: A Systematic Review, *Otolaryngology–Head and Neck Surgery*, Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/28925316>
70. Faulkner, H, et al. (1996). Validation of an objective structured assessment of technical skill for surgical residents. Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/9114900>
71. Martin, J. A, et al. (1997). Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg*. Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/9052454>
72. Niitsu, H, et al. (2013). 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*. Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/22941345>
73. Hopmans, C. J, et al. (2014). 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*. Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/25231747>
74. Lin, S. Y, et al. (2009). Development and pilot-testing of a feasible, reliable, and valid operative competency assessment tool for endoscopic sinus surgery. *Am J Rhinol Allergy*. Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/19490815>

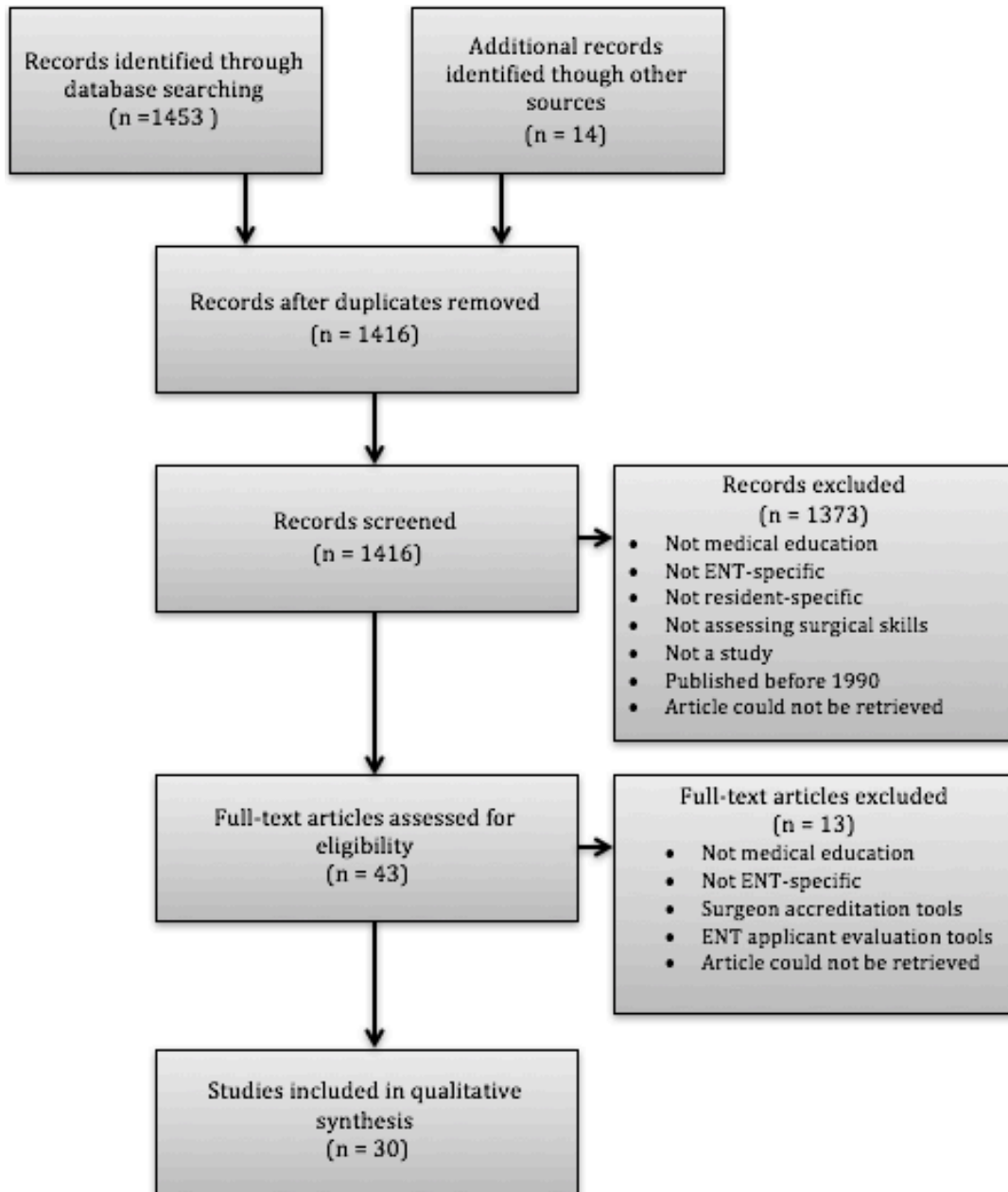
75. Ishman, S. L, et al. (2010). Development and pilot testing of an operative competency assessment tool for pediatric direct laryngoscopy and rigid bronchoscopy. *Laryngoscope*. Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/20939072>
76. Marglani, O, et al. (2012). Development of a tool for Global Rating of Endoscopic Surgical Skills (GRESS) for assessment of otolaryngology residents. *B-ENT*. Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/23113382>
77. Obeid, A. A, et al. (2014). Development and testing for an operative competency assessment tool for nasal septoplasty surgery. *Am J Rhinol Allergy*. Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/25197910>
78. Choi, B CK, Pak WP A. (2004) A Catalog of Biases in Questionnaires. *US National Library of Medicine National Institutes of Health*. Prev Chronic Dis. Repéré à www.ncbi.nlm.nih.gov/pmc/articles/PMC1323316/
79. McAlister, E. D, Goldstein, D. P, Rotstein, L. E. (2014). Redefining classification of central neck dissection in differentiated thyroid cancer. *Head Neck*, Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/23508631>
80. Ahmed, A, et al. (2013). Assessment of improvement of trainee surgical skills in the operating room for tonsillectomy. *Laryngoscope*. Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/23483535>
81. Schwartz J, et al. (2016). Objective assessment of Myringotomy and tympanostomy tube insertion: A prospective single-blinded validation study. *Laryngoscope*. Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/26498973>

82. Brown, D. J, et al. (2008). Assessment of operative competency in otolaryngology residency: Survey of US Program Directors. *Laryngoscope*. Repéré à <https://www.ncbi.nlm.nih.gov/pubmed/18641530>
83. Research-Methodology. Convenience Sampling. Repéré à research-methodology.net/sampling-in-primary-data-collection/convenience-sampling/
84. Encyclopedia.com. (2018). Interview bias. *A Dictionary of Sociology*. Repéré à <https://www.encyclopedia.com/social-sciences/dictionaries-thesauruses-pictures-and-press-releases/interview-bias>
85. Catalog of Bias. (2018). Recall Bias. Repéré à catalogofbias.org/biases/recall-bias/

Annexe 1: Figures

Chapitre 2

Figure 1: Search Flow Chart



Chapitre 3

Figure 8: Objective Structured Assessment of Technical Skills (OSATS) tool⁷

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	

Figure 9: French global assessment tool

Numéro du résident : _____ Niveau du résident : _____
 Évaluateur(s) : _____
 Date de la chirurgie : _____
 Date de complétion du questionnaire : _____
 Zones de l'évidement : _____
 Cou post-radique? Oui Non Difficulté du cas : Standard Difficile

Grille d'évaluation globale (GÉG)

Respect des tissus				
1	2	3	4	5
Utilisation de force non nécessaire ou dommages causés par l'utilisation inappropriée des instruments		Manipulation prudente des tissus, mais dommages occasionnels par inadvertance		Manipulation appropriée des tissus de manière constante, avec dommages minimaux
Gestion du temps et des mouvements				
1	2	3	4	5
Mouvements inutiles fréquents		Utilisation efficace du temps et des mouvements, mais quelques mouvements inutiles		Économie claire des mouvements et efficacité maximale
Utilisation des instruments				
1	2	3	4	5
Faits des gestes hésitants ou maladroits avec ses instruments de manière répétée		Utilisation compétente des instruments, mais occasionnellement rigide ou maladroit		Mouvements fluides avec les instruments, aucune maladresse
Utilisation de l'assistance				
1	2	3	4	5
Mauvais positionnement de l'assistant ou n'utilise pas l'assistant		Bon usage de l'assistant la majorité du temps		Utilise stratégiquement l'assistant à son avantage en tout temps
Connaissance the instruments				
1	2	3	4	5
Demande fréquemment le mauvais instrument ou utilise un instrument inapproprié		Connait le nom de la majorité des instruments et utilise le bon instrument pour la tâche		Très familier avec les instruments requis et leur nom
Connaissance de la procédure				
1	2	3	4	5
Connaissances déficientes et besoin d'instructions durant la plupart des étapes		Connait les étapes importantes de l'opération		Très familier avec tous les aspects de l'opération
Fluidité de la chirurgie et préparation				
1	2	3	4	5
Arrêts fréquents de la procédure et incertitude face à la prochaine étape		Démontre un certain niveau de planification quant aux prochaines étapes et progresse de manière consistante dans la chirurgie		Progression de la chirurgie planifiée de manière évidente et fluidité naturelle en toute temps
Avez-vous donné de la rétroaction au résident?				Oui
Croyez-vous que le résident est compétent à faire la chirurgie de manière indépendante?				Oui

TOTAL GÉG : _____/35
 Temps de complétion du questionnaire : _____

Figure 10: French task specific checklist tool for neck dissection

Grille d'évaluation spécifique à l'évidement cervical (GÉS)

	Incapable d'effectuer l'étape	Étape effectuée avec beaucoup d'aide	Étape effectuée avec de l'aide	Étape effectuée avec un minimum d'aide	Effectue l'étape aisément et avec fluidité	
Positionnement du patient	1	2	3	4	5	N/D
Placement de l'incision cutanée	1	2	3	4	5	N/D
Incision de la peau et du platysma	1	2	3	4	5	N/D
Élévation du lambeau sous-platysmal	1	2	3	4	5	N/D
Délimitation de la zone IA	1	2	3	4	5	N/D
Identification et préservation de la branche marginale	1	2	3	4	5	N/D
Identification +/- préservation des vaisseaux faciaux	1	2	3	4	5	N/D
Inclusion des tissus adipo-ganglionnaires jusqu'au rebord de la mandibule	1	2	3	4	5	N/D
Libération du muscle mylohyoïde	1	2	3	4	5	N/D
Identification et libération du nerf lingual	1	2	3	4	5	N/D
Libération du plan profond de la glande	1	2	3	4	5	N/D
Défascialisation du muscle SCM	1	2	3	4	5	N/D
Identification et dissection du nerf XI ad digastrique	1	2	3	4	5	N/D
Dissection et délimitation du digastrique	1	2	3	4	5	N/D
Nerf XI mis en corde à linge	1	2	3	4	5	N/D
Dissection de la zone rétrospinale	1	2	3	4	5	N/D
Libération de l'angle de la veine jugulaire interne - XI	1	2	3	4	5	N/D
Identification des branches du plexus cervical profond	1	2	3	4	5	N/D
Décollement du spécimen du plan profond	1	2	3	4	5	N/D
Délimitation du bord inférieur de la dissection	1	2	3	4	5	N/D
Défascialisation de la veine jugulaire interne	1	2	3	4	5	N/D
Identification et préservation du NC X	1	2	3	4	5	N/D
Dissection du tronc thyro-linguo-facial	1	2	3	4	5	N/D
Limite médiane de l'évidement	1	2	3	4	5	N/D

TOTAL GÉS: _____

GRAND TOTAL : _____

Annexe 2 : Tableaux

Chapitre 3

Tableau 1: Questionnaire d'acceptabilité

Questions	Résidents			Patrons	
	Pré- étude	Post- étude	p*	Post-étude	p**
Q1 Le projet va accroître votre anxiété lorsque vous effectuerez la procédure	3,40	2,88	(0,53)	4,00	(0,37)
Q2 Le projet va nuire à votre performance comparativement à votre niveau habituel	2,65	2,38	(0,82)	1,75	(0,28)
Q3 Le projet aura un impact favorable sur vos habiletés techniques en général	6,80	6,25	(0,47)	7,50	(0,15)
Q4 Le projet aura un impact favorable sur vos habiletés techniques spécifiques à l'évidement cervical	7,60	7,25	(0,86)	7,00	(1,00)
Q5 Vous souhaitez que ce genre d'évaluation demeure formatif	7,15	7,75	(0,60)	4,75	(0,15)
Q6 Vous souhaitez que ce genre d'évaluation serve éventuellement à vos évaluations de stage	4,70	4,00	(0,43)	7,25	(0,042)***
Q7 Vous pensez que ce genre d'évaluation est un bon moyen pour	7,25	7,50	(0,71)	9,00	(0,15)

évaluer vos compétences à réaliser les procédures chirurgicales les plus fréquentes en ORL					
Q8 Vous croyez que ce type d'évaluation sera représentatif de vos compétences pour votre niveau	6,45	7,38	(0,17)	9,00	(0,073)
Q9 Vous croyez que ce type d'évaluation augmentera la quantité de rétroaction que vous recevrez	8,55	8,63	(1,00)	9,00	(0,68)
Q10 Vous croyez que ce type d'évaluation vous permettra de mieux cibler vos forces et faiblesses au point de vue chirurgical	8,70	7,75	(0,24)	8,25	(0,68)
Q11 Vous croyez que ce type d'évaluation permettra une meilleure reproductibilité entre vos évaluations d'un patron à l'autre	7,65	6,50	(0,22)	7,75	(0,37)
Q12 Vous croyez que ce type d'évaluation est applicable dans le cadre de la résidence	7,90	7,50	(0,78)	9,25	(0,15)

*valeurs p comparant les scores pré et post étude chez les résidents

**valeurs p comparant les scores des résidents et des patrons post-étude

***Indique une valeur statistiquement significative ($p < 0.05$)

Tableau II: Traduction anglophone du questionnaire d'acceptabilité

Survey Questions	Residents			Attendings	
	Pre-study	Post-study	p*	Post-study	p**
Q1 Will being evaluated increase your <u>anxiety</u> ?	3,40	2,88	(0,53)	4,00	(0,37)
Q2 Will being evaluated <u>negatively</u> <u>impact your performance</u> ?	2,65	2,38	(0,82)	1,75	(0,28)
Q3 Will there be a <u>favorable impact</u> on your <u>global surgical skills</u> ?	6,80	6,25	(0,47)	7,50	(0,15)
Q4 Will there be a <u>favorable impact</u> on your <u>skills for neck dissection</u> ?	7,60	7,25	(0,86)	7,00	(1,00)
Q5 You wish for these kind of evaluations to remain <u>formative</u> ?	7,15	7,75	(0,60)	4,75	(0,15)
Q6 You wish for these kind of evaluations to be used in <u>summative assessments</u> ?	4,70	4,00	(0,43)	7,25	(0,042)***
Q7 You believe these kind of evaluations are a good way to <u>assess surgical skills in OTL-HNS</u> ?	7,25	7,50	(0,71)	9,00	(0,15)
Q8 You believe these evaluations will represent <u>your level of competence</u> ?	6,45	7,38	(0,17)	9,00	(0,073)

Q9 You believe these evaluations will increase the amount of <u>feedback</u> you receive?	8,55	8,63	(1,00)	9,00	(0,68)
Q10 You believe these evaluations will help you identify your <u>strengths and weaknesses</u> ?	8,70	7,75	(0,24)	8,25	(0,68)
Q11 You believe these evaluations will allow better <u>reproducibility</u> in your evaluations?	7,65	6,50	(0,22)	7,75	(0,37)
Q12 You believe these evaluations are <u>applicable in residency</u> ?	7,90	7,50	(0,78)	9,25	(0,15)

*p-values for comparison of resident pre and post-study scores

**p-values for comparison of resident and attending scores post-study

***Indicates a statistically significant value (p<0.05)

Annexe 3: Documents supplémentaires

Document 1: Validité selon les critères de Sullivan

EDITORIAL

A Primer on the Validity of Assessment Instruments

GAIL M. SULLIVAN, MD, MPH

1. What is reliability?¹

Reliability refers to whether an assessment instrument gives the same results each time it is used in the same setting with the same type of subjects. Reliability essentially means *consistent* or *dependable* results. Reliability is a part of the assessment of validity.

2. What is validity?²

Validity in research refers to how accurately a study answers the study question or the strength of the study conclusions. For outcome measures such as surveys or tests, validity refers to the *accuracy* of measurement. Here validity refers to how well the assessment tool actually measures the underlying outcome of interest. Validity is not a property of the tool itself, but rather of the interpretation or specific purpose of the assessment tool with particular settings and learners.

Assessment instruments must be both reliable and valid for study results to be credible. Thus, reliability and validity must be examined and reported, or references cited, for each assessment instrument used to measure study outcomes. Examples of assessments include resident feedback survey, course evaluation, written test, clinical simulation observer ratings, needs assessment survey, and teacher evaluation. Using an instrument with high reliability is not sufficient; other measures of validity are needed to establish the credibility of your study.

3. How is reliability measured?²⁻⁴

Reliability can be estimated in several ways; the method will depend upon the type of assessment instrument. Sometimes reliability is referred to as internal validity or internal structure of the assessment tool.

For *internal consistency* 2 to 3 questions or items are created that measure the same concept, and the difference among the answers is calculated. That is, the correlation among the answers is measured.

Cronbach alpha is a test of internal consistency and frequently used to calculate the correlation values among the answers on your assessment tool.⁵ Cronbach alpha calculates correlation among all the variables, in every combination; a high reliability estimate should be as close to 1 as possible.

Gail M. Sullivan, MD, MPH, is Editor-in-Chief, *Journal of Graduate Medical Education*.

Corresponding author: Gail M. Sullivan, MD, MPH, Editor-in-Chief, *Journal of Graduate Medical Education*, 515 N State St, Suite 2000, gsullivan@nsor.uchc.edu

DOI: 10.4300/JGME-D-11-00075.1

For *test/retest* the test should give the same results each time, assuming there are no interval changes in what you are measuring, and they are often measured as correlation, with Pearson *r*.

Test/retest is a more conservative estimate of reliability than Cronbach alpha, but it takes at least 2 administrations of the tool, whereas Cronbach alpha can be calculated after a single administration. To perform a *test/retest*, you must be able to minimize or eliminate any change (ie, learning) in the condition you are measuring, between the 2 measurement times. Administer the assessment instrument at 2 separate times for each subject and calculate the correlation between the 2 different measurements.

Interrater reliability is used to study the effect of different raters or observers using the same tool and is generally estimated by percent agreement, kappa (for binary outcomes), or Kendall tau.

Another method uses analysis of variance (ANOVA) to generate a *generalizability coefficient*, to quantify how much measurement error can be attributed to each potential factor, such as different test items, subjects, raters, dates of administration, and so forth. This model looks at the overall reliability of the results.⁶

5. How is the validity of an assessment instrument determined?⁶⁻⁸

Validity of assessment instruments requires several sources of evidence to build the case that the instrument measures what it is supposed to measure.^{9,10} Determining validity can be viewed as constructing an evidence-based argument regarding how well a tool measures what it is supposed to do. Evidence can be assembled to support, or not support, a specific use of the assessment tool. Evidence can be found in *content*, *response process*, *relationships to other variables*, and *consequences*.

Content includes a description of the steps used to develop the instrument. Provide information such as who created the instrument (national experts would confer greater validity than local experts, who in turn would have more validity than nonexperts) and other steps that support the instrument has the appropriate content.

Response process includes information about whether the actions or thoughts of the subjects actually match the test and also information regarding training for the raters/observers, instructions for the test-takers, instructions for scoring, and clarity of these materials.

Relationship to other variables includes correlation of the new assessment instrument results with other performance outcomes that would likely be the same. If

there is a previously accepted “gold standard” of measurement, correlate the instrument results to the subject’s performance on the “gold standard.” In many cases, no “gold standard” exists and comparison is made to other assessments that appear reasonable (eg, in-training examinations, objective structured clinical examinations, rotation “grades,” similar surveys).

Consequences means that if there are pass/fail or cut-off performance scores, those grouped in each category tend to perform the same in other settings. Also, if lower performers receive additional training and their scores improve, this would add to the validity of the instrument.

Different types of instruments need an emphasis on different sources of validity evidence.⁷ For example, for observer ratings of resident performance, interrater agreement may be key, whereas for a survey measuring resident stress, relationship to other variables may be more important. For a multiple choice examination, content and consequences may be essential sources of validity evidence. For high-stakes assessments (eg, board examinations), substantial evidence to support the case for validity will be required.⁹

There are also other types of validity evidence, which are not discussed here.

6. How can researchers enhance the validity of their assessment instruments?

First, do a literature search and use previously developed outcome measures. If the instrument must be modified for use with your subjects or setting, modify and describe how, in a transparent way. Include sufficient detail to allow readers to understand the potential limitations of this approach.

If no assessment instruments are available, use content experts to create your own and pilot the instrument prior to using it in your study. Test reliability and include as many sources of validity evidence as are possible in your paper. Discuss the limitations of this approach openly.

7. What are the expectations of JGME editors regarding assessment instruments used in graduate medical education research?

JGME editors expect that discussions of the validity of your assessment tools will be explicitly mentioned in your

manuscript, in the methods section. If you are using a previously studied tool in the same setting, with the same subjects, and for the same purpose, citing the reference(s) is sufficient. Additional discussion about your adaptation is needed if you (1) have modified previously studied instruments; (2) are using the instrument for different settings, subjects, or purposes; or (3) are using different interpretation or cut-off points. Discuss whether the changes are likely to affect the reliability or validity of the instrument.

Researchers who create novel assessment instruments need to state the development process, reliability measures, pilot results, and any other information that may lend credibility to the use of homegrown instruments. Transparency enhances credibility.

In general, little information can be gleaned from single-site studies using untested assessment instruments; these studies are unlikely to be accepted for publication.

8. What are useful resources for reliability and validity of assessment instruments?

The references for this editorial are a good starting point.

References

- 1 American Educational Research Association, American Psychological Association, National Council on Measurement in Education. *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association; 1999.
- 2 Downing SM. Reliability: on the reproducibility of assessment data. *Med Educ*. 2004;38(9):1006–1012.
- 3 Beckman TJ, Ghosh AK, Cook DA, Erwin PJ, Manderkar JN. How reliable are assessments of clinical teaching?: a review of the published instruments. *J Gen Intern Med*. 2004;19(9):971–977.
- 4 Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments. *Am J Med*. 2006;119(2):166e7–166e16.
- 5 Bland JM, Altman DG. Statistics notes: Cronbach’s alpha. *BMJ*. 1997;314:572.
- 6 Brennan RL. *Generalizability Theory*. New York, NY: Springer-Verlag; 2001.
- 7 Downing SM. Validity: on the meaningful interpretation of assessment data. *Med Educ*. 2003;37(9):830–837.
- 8 Downing SM, Haldyna TM. Validity threats: overcoming interference with proposed interpretations of assessment data. *Med Educ*. 2004;38(3):327–333.
- 9 Kane M. Validating high-stakes testing programs. *Educ Meas Issues Pract*. 2002;131–41.
- 10 Kane M. The assessment of professional competence. *Eval Health Prof*. 1992;15(2):163–182.

Document 2: Consignes pour les patrons d'ORL

Consignes pour les patrons d'ORL faisant l'évaluation des compétences chirurgicales des résidents pour l'évidement cervical

L'étude à laquelle vous participez vise à déterminer la validité, la fiabilité, la reproductibilité et l'applicabilité d'une grille développée pour évaluer les compétences chirurgicales des résidents en ORL qui apprennent à faire un évidement cervical.

Afin de standardiser l'évaluation des résidents, voici quelques consignes à respecter pendant et après la période d'évaluation.

- Le résident doit être averti qu'il s'agit d'une procédure durant laquelle il sera évalué.
- Le résident doit être mis au courant des critères sur lesquels sera basée son évaluation.
- Le patron a la liberté de décider à quelles parties de la chirurgie il juge que le résident peut participer.
- Lorsque le résident est invité à participer à la chirurgie, le patron doit le moins possible offrir son aide et laisser le résident demander lui-même de l'assistance pour le positionnement, la rétraction et les instruments.
- Le patron doit donner au résident le maximum d'autonomie jugé sécuritaire durant sa partie de l'opération. Il peut évidemment reprendre la barre s'il juge que c'est dans le meilleur intérêt du patient.
- Le patron peut donner de la rétroaction orale durant le cas.
- Le patron doit remplir l'évaluation écrite du résident en dedans de 48h et le résident doit être mis au courant de ses résultats.
- Le patron ne doit pas inscrire le nom du résident sur l'évaluation, mais plutôt utiliser le numéro attribué au résident.
- Le patron doit s'identifier sur la feuille d'évaluation.
- Le patron doit baser son évaluation sur l'ensemble de la chirurgie, incluant les portions non réalisées par le résident. Le résident recevra 0 pour les parties auxquelles il n'a pas participé.

Si vous avez des questions, n'hésitez pas à contacter le Dr Érika Mercier, soit au (418)255-2256 ou à erika.mercier@umontreal.ca.

Merci de votre participation.

Érika Mercier, MD
R2 oto-rhino-laryngologie et chirurgie cervico-faciale
Université de Montréal

Document 3: Consignes pour les résidents d'ORL

Consignes pour les résidents d'ORL lors de leur évaluation

L'étude à laquelle vous participez vise à déterminer la validité, la fiabilité, la reproductibilité et l'applicabilité d'une grille développée pour évaluer les compétences chirurgicales des résidents en ORL qui apprennent à faire un évidement cervical.

Afin de standardiser l'évaluation, voici quelques consignes à respecter pendant et après la période d'évaluation.

- Vous serez averti préalablement qu'il s'agit d'une procédure durant laquelle vous serez évalué.
- Vous serez mis au courant des critères sur lesquels sera basée votre évaluation.
- Le patron a la liberté de décider à quelles parties de la chirurgie il juge que vous pouvez participer.
- Lorsque vous êtes invité à participer à la chirurgie, le patron doit le moins possible offrir son aide et il doit vous laisser demander vous-même l'assistance pour le positionnement, la rétraction et les instruments.
- Le patron doit vous donner le maximum d'autonomie jugé sécuritaire durant votre partie de l'opération. Il peut évidemment reprendre la barre s'il juge que c'est dans le meilleur intérêt du patient.
- Le patron peut donner de la rétroaction orale durant le cas.
- Le patron doit remplir votre évaluation écrite en dedans de 48h et vous serez mis au courant de vos résultats.
- Le patron ne doit pas inscrire votre nom sur l'évaluation, mais plutôt utiliser le numéro qui vous est attribué.
- Le patron doit s'identifier sur la feuille d'évaluation.
- Le patron doit baser son évaluation sur l'ensemble de la chirurgie, incluant les portions non réalisées par vous. Vous recevrez donc 0 pour les parties auxquelles vous n'avez pas participé.

Si vous avez des questions, n'hésitez pas à contacter le Dr Érika Mercier, soit au (418)255-2256 ou à erika.mercier@umontreal.ca.

Merci de votre participation.

Érika Mercier, MD
R2 oto-rhino-laryngologie et chirurgie cervico-faciale
Université de Montréal