

Université de Montréal

Développement d'un outil d'évaluation des techniques
chirurgicales en plastie

Par Saoussen SALHI

Programme de Sciences Biomédicales, Faculté de Médecine

Mémoire présenté à la Faculté de Médecine en vue de l'obtention du grade de M.Sc. en
Sciences Biomédicales option Musculosquelettique.

Dépôt Octobre 2016.

Copyright, Saoussen SALHI, 2016.

Développement d'un outil d'évaluation des techniques chirurgicales en plastie

Résumé

Introduction

Pendant la formation d'un résident de chirurgie, l'évaluation des habiletés techniques se fait au moyen de méthodes subjectives et peu reproductibles.

Le but de cette étude est de développer un nouvel outil d'évaluation valide et fiable, pour examiner les techniques de microchirurgie des résidents de plastie lors de la performance d'une réparation de nerf digital.

Matériaux et méthodes

Des résidents de différents niveaux de formation ainsi que des patrons en chirurgie plastique ont effectué une réparation de nerf digital sur un modèle humain cadavérique du membre supérieur. Ces performances ont été filmées et évaluées par deux évaluateurs au moyen d'un nouvel outil d'évaluation développé. Cet instrument compte 12 étapes, chacune évaluée au moyen d'une échelle de Likert à 5 points. La fiabilité inter-évaluateur a été déterminée en comparant pour chaque candidat, les évaluations par les deux évaluateurs et la validité de l'outil en comparant les résultats de participants de différents niveaux d'expertise.

Résultats

Dix candidats ont pris part dans l'étude. La fiabilité inter-évaluateur de l'outil était acceptable, avec un coefficient de corrélation intraclasse de 0.690. La cohérence interne du test était élevée avec un coefficient de Cronbach alpha qui s'élevait à 0.960. Enfin, la validité de l'outil était démontrée par la différence statistiquement significative ($p=0.017$) entre les résultats de candidats de différents niveau d'expertise.

Conclusion

Nous avons développé un nouvel instrument d'évaluation, dont la fiabilité et la validité élevées en font un outil utile dans l'enseignement et l'évaluation des résidents en chirurgie plastique.

Mots clés

Outil d'évaluation ; instrument d'évaluation ; pédagogie ; enseignement chirurgical ; techniques chirurgicales.

Summary

Introduction

During surgical residencies, technical skills continue to be assessed using subjective evaluation methods.

We aim to develop an objective, valid and reliable evaluation tool to assess plastic surgery residents' microsurgical skills in performing a digital nerve repair.

Materials and methods

Plastic surgery residents of different levels of training and staff surgeons performed a digital nerve repair on a human cadaveric hand model. These performances were recorded and assessed by blinded raters using a newly developed evaluation tool. This consisted of 12 steps, each rated using a 5-point Likert scale. The tool's reliability was determined by comparing each candidate's evaluation by the two separate raters and its validity determined by comparing performances of candidates of different levels of expertise.

Results

Ten candidates took part in this study. The inter-rater reliability was considered acceptable, with an intraclass correlation coefficient of 0.690. The test's internal consistency was excellent with a Cronbach's alpha coefficient of 0.960. The test's validity was demonstrated by the fact that higher levels of expertise were associated with higher performance scores ($p=0.017$).

Conclusion

We developed a new evaluation tool for a microsurgical procedure, that performed well in terms of reliability and validity. This makes it a pedagogically useful tool that can be incorporated in plastic surgery residents' training to improve surgical skills teaching and assessment.

Key words

Evaluation tool ; evaluation instrument ; pedagogy ; surgical teaching ; surgical skills.

Table des matières

Page titre.....	i
Résumé.....	ii
Mots clés.....	iii
Summary.....	iv
Key words.....	v
Liste des tableaux.....	x
Liste des figures.....	xi
Liste des abréviations.....	xii
Remerciements.....	xiii
Introduction.....	14
1. La compétence en chirurgie.....	14
a. Définition de la compétence.....	14
b. Composantes de la compétence clinique.....	14
c. Importances des habiletés techniques.....	15
2. Évaluation de la compétence en chirurgie.....	16
a. Définition de l'évaluation.....	16
b. Le rôle de l'évaluation dans l'apprentissage.....	16
c. L'évaluation des différentes composantes de la compétence clinique.....	17
d. L'évaluation des techniques chirurgicales.....	17
3. Pressions pour développer des méthodes d'évaluation des techniques chirurgicales objectives.....	20

4. Les caractéristiques d'un outil d'évaluation objectif.....	21
a. Faisabilité.....	21
b. Validité.....	21
c. Fiabilité.....	23
5. Types d'outil d'évaluation.....	24
a. Les échelles d'évaluation globale.....	26
b. Les listes de vérification.....	27
c. Échelle procédurale spécifique.....	28
6. Objectifs de l'étude.....	29
a. Pourquoi la microchirurgie.....	29
Matériaux et méthodes.....	31
1. Revue de la littérature.....	31
2. Conception de l'outil d'évaluation.....	32
a. Décomposition de la chirurgie de micro-neurorrhaphie de nerf digital en une série d'étapes.....	32
b. Étapes de la chirurgie retenues dans l'outil d'évaluation.....	32
3. Évaluation de l'outil développé.....	35
a. Création d'une installation pour tester l'outil d'évaluation.....	35
b. Étude pilote.....	36
c. Création du matériel d'évaluation.....	37
d. Recrutement de participants.....	38
e. Analyses statistiques.....	39
Résultats.....	40

1. Version finale de l’outil d’évaluation conçu.....	40
2. Évaluation de l’outil.....	41
a. Fiabilité inter-évaluateur.....	42
b. Cohérence interne.....	43
c. Validité conceptuelle.....	44
Discussion.....	44
1. Évaluation en chirurgie.....	44
2. Méthodes objectives d’évaluation des techniques chirurgicales.....	46
a. Échelles d’évaluation globale.....	46
b. Listes de vérification.....	47
c. Notre outil d’évaluation.....	48
d. Combinaison d’une échelle d’évaluation globale et d’un outil spécifique à une procédure.....	51
3. Situation propice à l’évaluation du résident.....	51
a. Salle d’opération.....	51
b. Enregistrement vidéo.....	52
c. Enregistrement vidéo en salle d’opération.....	54
4. Impact d’un tel outil d’évaluation.....	54
a. Impact éducatif.....	54
b. Utilisation pour l’accréditation.....	55
5. Autres outils d’évaluation de la microchirurgie.....	56
6. Nouvelles tendances dans l’évaluation des techniques chirurgicales.....	58
7. Limitations de notre étude.....	59

a. Conception de l'outil d'évaluation.....	59
b. Utilisation des modèles cadavériques.....	59
c. Matériel audio-visuel.....	60
d. Validation de l'instrument.....	60
Conclusion.....	61
Bibliographie.....	95

Listes des tableaux

Tableau I : Participants.....92

Tableau II : Variance de la cohérence interne de l’outil d’évaluation selon
la déletion de chacune de ses étapes.....93

Tableau III : Analyse post-hoc des différences de scores entre les groupes
de participants.....94

Liste des figures

Figure 1 : Exemple d'échelle d'évaluation globale.....	62
Figure 2 : Exemple de liste de vérification.....	63
Figure 3 : Exemple d'échelle procédurale spécifique.....	64
Figure 4 : Étapes d'une micro-neurorrhaphie digitale.....	66
Figure 5 : Outil d'évaluation.....	69
Figure 6 : Installation pour tester l'outil d'évaluation.....	74
Figure 7 : Microscope chirurgicale Leica M525 F50.....	75
Figure 8 : Image tirée d'une vidéo finale démontrant l'intégration des images des deux caméras.....	76
Figure 9 : Formulaire de consentement- Version résident.....	77
Figure 10 : Outil d'évaluation- Version finale.....	84
Figure 11 : Variation de la performance des participants selon le niveau d'expertise.....	87
Figure 12 : University of Western Ontario Microsurgical Skills Acquisition instrument : UWOMSA.....	88
Figure 13 : Structured Assessment of Microsurgical Skills (SAMS).....	90
Figure 14 : Liste d'erreurs.....	91

Liste des abréviations

CanMEDS : Canadian Medical Education Directives for Specialists

ABSITE : American Board of Surgery In-Training Exam

ACGME : Accreditation Council for Graduate Medical Education

MeSH : Medical Subject Headings

ECOS : Examen Clinique Objectif Structuré

ANOVA : Analyse de variance

PGY : Post-Graduate Year

GRS : Global Rating Scale

OSATS : Objective Structured Assessment of Technical Skill

GOALS : Global Operative Assessment of Laparoscopic Skills

UWOMSA : University of Western Ontario Microsurgical Skills Acquisition instrument

SAMS : Structured Assessment of Microsurgical Skills

Remerciements

Je tiens à remercier Dr Jenny Lin pour avoir supervisé mon projet de maîtrise et m'avoir guidé non seulement dans la réalisation de ce projet mais également dans la rédaction de ce mémoire.

Je remercie, par ordre alphabétique, Dr Patricia Bortoluzzi, Dr André Chollet, Dr Louise Laberge, Dr Stephen Nicolaidis, et Dr Dominique Tremblay, pour avoir accepté de participer à la conception de notre outil d'évaluation.

Enfin je remercie tous les résidents et tous les chirurgiens qui ont pris part dans la performance de la procédure de micro-neurorrhaphie digitale sur les modèles cadavériques.

Introduction

1. La compétence en chirurgie

a. Définition de la compétence

La compétence clinique est définie comme l'utilisation judicieuse de capacités de communication, connaissances, habiletés techniques, émotions, et valeurs ainsi que du raisonnement clinique et de la réflexion au profit d'un individu ou d'une communauté que l'on sert sur une base quotidienne (1). Cette compétence est le résultat d'expériences acquises et ponctuées de rétroaction positive, c'est à dire d'évaluation objective ayant pour but l'amélioration des habiletés cliniques (2).

b. Composantes de la compétence clinique

Le développement de la compétence clinique fait appel à plusieurs aptitudes, aussi bien cognitives que techniques (3), parmi lesquelles on retrouve l'application de connaissances théoriques (2, 4, 5), la prise de décision, l'acquisition d'habiletés techniques (2, 4, 6), l'art de la communication (2, 4-6), le travail en équipe, le jugement clinique et la compassion (5).

Les différentes organisations régionales réglementant les soins de la santé, telles que le Collège Royal des médecins et chirurgiens du Canada, Le Collège Royal des chirurgiens du Royaume Uni, le « American Board of Medical Specialists » des États Unis, et les différents corps de licence médicale dans les pays Européens, ont fixé des exigences pour décider de la compétence clinique des médecins finissants (1). Comme exemple, le Collège Royal des médecins et chirurgiens du Canada a rassemblé ces exigences sous l'appellation CanMEDS, abréviation pour « Canadian Medical Education Directives for Specialists ». Celles-ci incluent :

- **La promotion de la santé**, c'est à dire l'usage d'une expertise médicale dans le but d'améliorer la santé et de garantir le bien-être d'une population de patients (7).
- **Le professionnalisme** qui fait référence au devoir du médecin de promouvoir la santé de ses patients selon les normes médicales actuelles et tout en respectant un code d'éthique qui inclut des valeurs telles que l'intégrité et l'honnêteté (7).
- **La collaboration** avec différents professionnels de la santé au sein d'équipes multidisciplinaires en vue de fournir les meilleurs soins de santé aux patients (7).
- **L'érudition** qui fait référence à l'engagement du médecin à toujours faire preuve d'excellence dans la pratique médicale en participant au processus de formation continue et à l'avancement des sciences médicales par la recherche (7).
- **Le leadership** du médecin à titre d'administrateur, de clinicien ou d'enseignant dans l'optique de faire évoluer le système de santé vers l'excellence (7).
- **La communication** avec le patient et ses proches pour assurer un échange facile et efficace d'informations essentielles à la prestation de soins supérieurs (1, 7).
- **La gestion** de pratiques durables et de ressources pour promouvoir un système de santé efficace (7).
- L'application de tous ces attributs ainsi que le raisonnement diagnostique, le jugement clinique, la prise de décision, l'administration de traitements adéquats et la maîtrise d'habiletés techniques contribue à la création d'un **expert médical** (7).

c. Importances des habiletés techniques

Parmi toutes ces aptitudes nécessaires à la compétence clinique, les habiletés techniques jouent un rôle primordial dans la formation d'un chirurgien (8, 9) et dans le succès chirurgical

(2, 10-13). On estime que 25% du résultat chirurgical dépend des habiletés techniques du chirurgien (14, 15).

De plus en plus, les compétences chirurgicales prennent de l'ampleur au sein des disciplines médicales. En effet, le développement de procédures interventionnelles dans des spécialités traditionnellement considérées comme médicales (4, 16), telles que la radiologie, la cardiologie, la gastroentérologie ou la pneumologie, fait en sorte que le développement et la maîtrise de techniques chirurgicales sont essentiels dans la formation de tous les médecins, pas seulement dans celle des chirurgiens (4).

2. Évaluation de la compétence en chirurgie

a. Définition de l'évaluation

Le mot évaluation prend son étymologie du latin « a + valere » qui signifie attribuer une valeur à l'objet étudié (17). Dans son sens large, une évaluation est définie comme le jugement porté sur la performance, le progrès ou l'accomplissement d'un individu lorsque ceux-ci sont comparés à une référence et à des critères prédéfinis (5, 12, 17-19). Dans un contexte éducatif, une évaluation a pour objectif d'estimer le degré de maîtrise des connaissances et des habiletés techniques, de mesurer le progrès effectué sur une période de temps, de définir les forces et les lacunes de l'individu évalué, et de classer différentes personnes lorsqu'il s'agit d'un processus de sélection ou d'exclusion (19).

b. Le rôle de l'évaluation dans l'apprentissage

Le processus d'évaluation est essentiel dans la formation médicale (12). Il a deux grands objectifs. D'une part, les évaluations à faible enjeu possèdent un rôle de formation (5, 19) et

poussent à l'érudition (1, 18-21) en permettant à un apprenti d'identifier ses lacunes et de les adresser au moyen d'un apprentissage ciblé (1, 11). D'autre part, les évaluations sommatives, ou à enjeu élevé, ont pour but d'examiner et de certifier (1, 5, 18, 19, 21). En d'autres termes, les évaluations formatives sont des évaluations **pour** l'apprentissage, tandis que les évaluations sommatives sont des évaluations **de** l'apprentissage (22).

c. L'évaluation des différentes composantes de la compétence clinique

Pour établir la compétence clinique d'un médecin, chacun des attributs qui la constituent doit être évalué. Si l'on revient sur l'exemple des exigences CanMEDS du Collège Royal des médecins et chirurgiens du Canada, **la promotion de la santé**, dont le but principal est d'assurer la sécurité des patients, peut être évaluée de manière indirecte à travers des données de morbidité et de mortalité (1). **Le professionnalisme** d'un médecin peut être jaugé en observant ses interactions avec les patients et ses pairs ou au biais de sondages de satisfaction conduits auprès de ses patients (1). De même, **le leadership**, les capacités de **communication** et de **gestion** d'un clinicien sont facilement évalués au moyen de l'observation. Des exemples d'outil d'évaluation des capacités de **collaboration** d'un médecin incluent le guide d'observation Calgary-Cambridge, ou encore le « Kalamazoo essential elements » (1). La méthode d'évaluation la plus répandue est celle de **l'érudition** et du jugement clinique qui se fait au moyen d'examens écrits ou oraux (1, 23-25) ou encore au moyen de discussions basées sur des cas cliniques ou de modules de formation continue (1).

d. L'évaluation des techniques chirurgicales

Actuellement, la réussite aux examens d'accréditation dépend de la performance dans ces examens écrits et oraux (23, 26-28). Bien qu'ils constituent une bonne méthode d'évaluation du bagage de connaissances du résident, ils ne reflètent pas ses habiletés techniques (5, 8, 26, 27). Il a été démontré par exemple qu'il n'existait aucune corrélation entre la performance au « American Board of Surgery In-Training Exam (ABSITE) » et l'exécution chirurgicale (4, 19, 23).

Une méthode objective d'évaluation des techniques chirurgicales demeure insaisissable (27) et actuellement, le Collège Royal des médecins et chirurgiens du Canada ne possède pas de moyen objectif et standardisé pour décider de la compétence chirurgicale des résidents finissants (8, 29). De même, le « American Board of Plastic Surgery » mesure la compétence chirurgicale selon des examens écrits et oraux, et selon le volume d'exposition du résident finissant évalué par un échantillonnage de son journal de bord (28).

L'enseignement traditionnel des procédures chirurgicales, introduit, il y a de cela, plus d'un siècle (19), continue à suivre le modèle d'apprenti de Halstead (19, 25, 30) dans lequel le résident commence par apprendre à maîtriser l'anatomie pertinente par des lectures et des cours didactiques (31). Par la suite, le résident acquiert l'opportunité d'observer un chirurgien expert opérer en salle d'opération (31-33). Progressivement, le résident est promu du stade d'observateur à celui d'assistant au chirurgien (31). Selon ce même modèle, le résident est évalué quant à sa capacité à exécuter une procédure chirurgicale en étant à son tour observé par un chirurgien (31, 32). Il se voit alors attribuer une autonomie grandissante en salle d'opération s'il satisfait son précepteur (34, 35).

Ce modèle génère une évaluation globale et subjective des habiletés techniques du résident (9, 10, 13, 15, 23, 27, 29, 31, 35-40). En effet, l'évaluation est souvent complétée à la fin de

l'année académique ou à la fin d'une rotation clinique (8, 15, 39), donc à distance de la performance du résident et dépend ainsi de la mémoire du superviseur (8, 12, 15, 27, 39, 41-43). L'évaluation peut donc être faussement influencée par les événements les plus récents (15). Elle peut également être teintée par les interactions passées entre le chirurgien superviseur et le résident évalué (9, 44, 45). Par exemple, il a été démontré que les résidents séniors sont souvent perçus comme étant plus compétent que les résidents juniors malgré des performances chirurgicales comparables (23). Plusieurs études suggèrent également que les superviseurs ont tendance à attribuer des évaluations plus favorables aux résidents avec qui ils sont familiers (23).

D'autre part, différents chirurgiens conviennent rarement du niveau de performance du résident évalué (24).

En outre, l'évaluation produite est souvent vague et générale (41) et ne s'appuie pas sur des critères spécifiques, donc elle ne permet pas une rétroaction ciblée sur les faiblesses du résident et sur des avenues pour améliorer ces lacunes (32, 37).

Par ailleurs, ce modèle d'apprentissage n'a de rendement pédagogique que dans un contexte de répétition, impliquant un grand volume de cas cliniques et chirurgicaux qui offrent aux résidents de multiples opportunités de pratique (2, 9, 27, 46, 47). Cette condition d'exposition volumineuse est de plus en plus difficile à garantir dans le cadre de la réduction du temps de travail (2, 5, 9, 27, 40, 46, 48) imposée aux résidents par les différentes fédérations de formation des médecins (5). Ainsi, le «Accreditation Council for Graduate Medical Education» ou ACGME aux États-Unis a exigé, en 2003, une restriction du temps de travail des résidents de toutes les spécialités médicales et chirurgicales dans le but de promouvoir un apprentissage et une formation de haute qualité, mais aussi afin d'assurer la sécurité des

patients traités dans les centres hospitaliers universitaires (46). De même, au Royaume-Uni (32) ainsi que dans le reste de l'Europe (48), les programmes de résidences chirurgicales ont souffert de la réduction des heures de formation dictée par la Directive Européenne sur le Temps de Travail (32, 45). Ceci soulève une inquiétude quant à la qualité des formations chirurgicales (32) et crée un besoin de développer des méthodes d'évaluation objectives pour garantir la compétence des résidents finissants (4, 14, 27, 33, 46, 49, 50). En effet, en imposant des semaines de travail de 80 heures, l'expérience clinique d'un résident dans un programme de formation de cinq ans se voit réduite de 6 à 12 mois (46). En d'autres termes, l'expérience hospitalière du résident, incluant la prise en charge de cas urgents et semi-urgents ainsi que l'assistance chirurgicale, a subi une réduction estimée entre 33% et 50% (46). Les résidents graduant de programmes chirurgicaux aujourd'hui ont certainement moins d'exposition chirurgicale et des habiletés techniques moins développées que les résidents ayant gradué auparavant (2, 47). Ceci contribue sans doute au fait qu'environ 80% des résidents chirurgicaux finissants appliquent pour une sur-spécialité, en quête de formation additionnelle (33).

3. Pressions pour développer des méthodes d'évaluation des techniques chirurgicales objectives

À cette réduction du temps de travail vient s'ajouter le problème grandissant de la diffusion des erreurs médicales et des morbidités, en particulier chirurgicales, par les médias (4, 10, 19, 41, 51). Ceci soulève au sein du grand public une inquiétude généralisée quant à la compétence clinique des chirurgiens (4, 41, 51, 52) et remet en question les méthodes actuelles qui mènent à leur accréditation (10). Par conséquent, les comités d'accréditation sont

appelés à démontrer de manière objective la compétence d'un chirurgien finissant (12, 19, 31, 33).

De plus, la complexité croissante des nouvelles techniques chirurgicales (4, 40, 50, 52, 53), l'emphase portée sur l'efficacité en salle d'opération (19) ainsi que les problèmes éthiques soulevés par l'apprentissage et la pratique sur des patients (54) représentent d'autres facteurs qui poussent l'évaluation chirurgicale dans la même direction.

4. Les caractéristiques d'un outil d'évaluation objectif

a. Faisabilité

Une bonne méthode d'évaluation doit être dans un premier lieu faisable (5, 10, 18, 27, 44) et pratique (3). Pour ce faire, il faut qu'elle évalue un nombre de critères limité (44), qu'elle soit peu coûteuse, acceptable aux yeux des participants et facile à instaurer (27).

b. Validité

Un outil d'évaluation n'a d'utilité pédagogique que s'il est prouvé valide et fiable (3, 5, 6, 10, 18, 19, 44, 55).

Selon le psychométricien Downing, la validité correspond à l'ensemble des preuves avancées pour appuyer l'interprétation attribuée aux résultats émis par l'outil d'évaluation (16, 30). En termes plus simples, un outil d'évaluation est considéré valide s'il mesure réellement ce qu'il prévoit mesurer (5, 8, 49, 55, 56). Ainsi, la bonne performance d'un résident selon un outil d'évaluation valide devrait indiquer que le résident en question est un chirurgien compétent (8). La validité est le sine qua non de l'outil d'évaluation car, sans validité, cet outil n'aurait aucune valeur éducative (16, 30).

Cinq types de validité ont été décrits:

- **La validité manifeste (face validity)** correspond à la ressemblance entre la tâche évaluée par l'outil et la procédure chirurgicale performée dans la situation réelle (19, 34). La validité manifeste est souvent assurée en incorporant, dans la conception de l'outil, une méthodologie de Delphi (47, 49).
- **La validité de contenu (content validity)** est prouvée si le contenu de l'outil d'évaluation est en rapport direct avec la technique chirurgicale que l'outil doit évaluer (5, 19, 34, 46). La validité de contenu (4, 47) est souvent établie en incorporant, lors de la conception de l'outil, des revues de littérature, de livres de référence, et l'opinion de groupes d'experts (méthode de Delphi) (4).
- **La validité critérielle ou concourante (concurrent/criterion validity)** repose sur la corrélation des résultats obtenus par l'outil dont la validité est à prouver avec les résultats obtenus par un autre instrument déjà validé et considéré comme étalon d'or (5, 19, 34).
- **La validité prédictive (predictive/outcome validity)** correspond à la capacité de l'outil d'évaluation à prédire une performance dans un contexte différent de celui dans lequel il a été testé (19, 34), notamment pour évaluer les techniques chirurgicales en salle d'opération lorsque l'outil a été testé dans un environnement de laboratoire qui simule la salle d'opération (19, 34).
- **La validité conceptuelle (construct validity)** fait référence à la correspondance entre les résultats de l'outil d'évaluation et les résultats attendus (19, 34, 57). Il est attendu que des experts, avec une expérience chirurgicale plus vaste, performant mieux que des novices (49). La validité conceptuelle est donc souvent démontrée par la capacité de

l'instrument à discriminer entre la performance d'experts et celle de novices (5, 9, 19, 24, 49, 52, 57). La validité conceptuelle est probablement la caractéristique la plus importante d'un outil d'évaluation (25) et devrait être une propriété de tout instrument utilisé pour mesurer une performance (57).

c. Fiabilité

D'autre part, un test est fiable s'il produit des résultats constants et stables (8, 56) lorsqu'il est répété (8). Il existe trois types de fiabilité.

- **La fiabilité test-retest** est démontrée par un outil d'évaluation qui, lorsqu'il est utilisé pour évaluer un même candidat, à plusieurs reprises donne des résultats comparables (5, 9, 19, 58). La prémisse est que le niveau du candidat n'ait pas évolué entre les différentes évaluations (58).
- **La fiabilité inter-évaluateur** est démontrée lorsque l'outil d'évaluation, utilisé par deux évaluateurs indépendants pour évaluer un même candidat, donne des résultats similaires (5, 9, 19, 58). Autrement dit, la fiabilité inter-évaluateur mesure si la différence de résultats obtenus par l'instrument d'évaluation reflète une réelle différence de performance plutôt qu'une différence d'évaluateur (52). Elle représente une caractéristique importante de l'instrument d'évaluation puisqu'un outil avec une fiabilité inter-évaluateur élevée peut être utilisé par différents évaluateurs sans que les résultats qu'il donne ne soient affectés par la variation d'évaluateur. Toute variation de résultats serait strictement le reflet d'une variation de performance (59). La fiabilité inter-évaluateur est celle qui a le plus grand impact sur la fiabilité d'un outil

d'évaluation, administré par des examinateurs, comme c'est le cas dans le domaine médical (60).

- **La cohérence interne** d'un instrument d'évaluation est elle aussi une mesure de la fiabilité d'un outil d'évaluation (8). Elle correspond à la corrélation entre les notes obtenues à chaque question ou étape de l'outil d'évaluation (9). Cette propriété reflète la contribution de chacun des domaines de l'outil au score final obtenu (9).

En somme, les différentes mesures de fiabilité d'un test donnent une estimation des erreurs aléatoires, générées par le test, dans l'évaluation d'une performance (60).

Les scores de fiabilité varient entre 0 (aucune fiabilité) et 1 (parfaite fiabilité) (5, 16, 19). Une fiabilité de l'ordre de 0.8 est exigée des évaluations à enjeu élevé, telle que celles destinées à la certification d'un médecin (5, 19, 23).

5. Types d'outil d'évaluation

On retrouve neuf différents types de méthode d'évaluation employés dans l'éducation chirurgicale (43) :

- **Le carnet de bord** (1, 43), souvent imposé par les programmes de résidence chirurgicale, est un registre de toutes les procédures auxquelles le résident a participé et de l'étendue de sa participation (1, 2, 23, 26). Bien que ces journaux de bord reflètent l'expérience acquise par le résident, ils ne garantissent ni sa compétence (2, 5, 18, 19, 23, 44) ni la qualité de ses gestes chirurgicaux (1, 10, 24, 61). En effet, Charles Mayo a affirmé que « l'expérience peut vouloir dire répéter constamment la même erreur » (2). Comme il est mentionné plus haut, la compétence naît de l'expérience, c'est à dire de la pratique, mais seulement si celle-ci est ponctuée de rétroaction

positive (2). Ainsi, ces journaux de bord ne font pas preuve de validité de contenu puisqu'ils n'indiquent que la participation du résident à la procédure et non pas sa capacité à mener la chirurgie (19, 26, 28). Ils sont également peu fiables (57, 62).

- Les données des **morbidités et mortalités** (43) sont souvent utilisées par les hôpitaux pour évaluer leur performance et leur capacité à garantir la sécurité de leurs patients (1). En d'autres termes, elles permettent d'évaluer l'aspect de la promotion de la santé des exigences CanMEDS (1). Cependant, évaluer les techniques chirurgicales au moyen de ces données requièrent un grand nombre de morbidités et mortalités et courent le risque d'accumuler des résultats défavorables pendant le processus d'évaluation (4, 19). De plus, les morbidités et mortalités sont souvent le résultat de plusieurs variables et ainsi ne reflètent pas seulement l'expertise chirurgicale de l'individu (10, 19, 24).
- **L'observation directe de la performance chirurgicale sans critères d'évaluation** (43) correspond au modèle d'apprenti préalablement décrit. Ce type d'observation engendre des évaluations à faible fiabilité, qu'il s'agisse de la fiabilité test-retest ou de la fiabilité inter-évaluateur, puisqu'il a été démontré que des chirurgiens experts sont souvent en désaccord quant à leur évaluation de la compétence d'un résident (19).
- Les nouvelles méthodes d'évaluation incluent les évaluations faisant usage de **modèle animalier ou de laboratoire, l'enregistrement vidéo de cas chirurgicaux, des listes de vérification spécifiques à une procédure, des échelles d'évaluation globale, des systèmes d'analyse de dextérité**, et les évaluations faisant usage de **simulateurs virtuels de réalité** (43). Parmi tous ces instruments d'évaluation, l'observation directe

se basant sur des critères spécifiques au moyen de listes de vérification ou d'échelle d'évaluation globale fait preuve de la plus haute validité et fiabilité (19, 43).

a. Les échelles d'évaluation globale

Les échelles d'évaluation globale ont pour but d'évaluer des habiletés techniques génériques et communes à plusieurs procédures (4, 10) telles que la manipulation d'instruments chirurgicaux et le déroulement global de la procédure (48). Ces habiletés sont typiquement notées sur une échelle de Likert avec des points d'ancrage qui servent de directives à l'évaluateur (19). Elles ne sont pas spécifiques à une procédure particulière (48) et ont donc l'avantage de pouvoir s'appliquer à différentes techniques chirurgicales (4). Cependant comme elles font usage d'une échelle de Likert, elle requiert que l'évaluateur fasse un jugement et sont prônes à refléter le biais de l'évaluateur. Ceci peut en affecter la fiabilité. Cependant, il a été démontré qu'en familiarisant les évaluateurs avec l'outil d'évaluation, les échelles d'évaluation globale font preuve d'une plus grande fiabilité que les listes de vérification (19).

Un des inconvénients majeur de l'utilisation des échelles d'évaluation globale dans un but éducatif est que, pour qu'elles soient applicables à différentes opérations, elles sont souvent rédigées avec des termes vagues et peu précis. Pour cette raison, elles perdent partiellement leur capacité à engendrer une rétroaction exhaustive, précise et ciblée (4, 52).

La figure 1 montre l'échelle d'évaluation globale développée par Martin et al. Cette échelle est utilisée pour évaluer des candidats sur des modèles de laboratoire ou des modèles animaliers (63).

b. Les listes de vérification

Les listes de vérification sont spécifiques à une procédure chirurgicale (4, 10). Elles divisent celle-ci en une séquence d'étapes et elles permettent d'évaluer si ces étapes ont été accomplies ou non (4, 48). Les listes de vérification ont très longtemps été utilisées dans le domaine de l'aviation et dans la formation des pilotes dont les tâches peuvent être organisées en une séquence logique et dont la performance peut être évaluée comme satisfaisante ou non à chaque étape de cette séquence (11). Par la suite, un programme d'entraînement personnalisé est instauré pour adresser les faiblesses identifiées par la liste de vérification (11).

Elles sont également communément utilisées dans le domaine médical, notamment sous formes des ECOS, ou Examens Cliniques Objectifs Structurés, dans lequel l'étudiant en médecine ou résident est noté selon une grille de tâches à accomplir (19).

L'avantage principal des listes de vérification est qu'elles transforment l'évaluateur en observateur, éliminant le besoin d'interprétation et donc le risque de subjectivité (19). En outre, en divisant une procédure en étapes, elles permettent d'identifier exactement où se situent les lacunes du résident et permettent donc de fournir une rétroaction précise et ciblée (4, 52).

Cependant, elles sont souvent critiquées sur deux plans. Elles n'évaluent que si une étape a été complétée, et non pas si elle a été complétée adéquatement (10, 19). De plus, elles suggèrent qu'il n'existe qu'une seule manière rigide de compléter une procédure (10) et ne récompense le résident que s'il performe la chirurgie selon la séquence d'étapes définie par la liste de vérification (10). Il peut en résulter qu'un novice, qui apprend à conduire une procédure chirurgicale en la décomposant en étapes successives, performe mieux selon la liste de vérification qu'un expert qui aurait plus tendance à voir une procédure comme un but à

atteindre et qui pourrait avoir une approche différente de celle utilisée par la liste de vérification. Elles ont donc tendance à démontrer une faible validité conceptuelle (19, 52). Elles ont également tendance à être moins fiables que les échelles d'évaluation globale (52).

La figure 2 montre un exemple de liste de vérification développée par de Montbrun et al pour évaluer les capacités des résidents à réaliser un examen sous anesthésie (Exam Under Anesthesia) (23). Pour chaque étape sur la liste, le candidat est noté selon s'il a complété l'étape, ou s'il ne l'a pas complétée ou mal complétée (23).

c. Échelle procédurale spécifique

Un hybride de ces deux types d'évaluation, l'échelle procédurale spécifique, a été développé afin de combiner les qualités de validité et de fiabilité des échelles d'évaluation globale avec les capacités de rétroaction des listes de vérification (52). Tout comme une liste de vérification, elle divise une chirurgie particulière en étapes considérées comme importantes. Cependant, au lieu de simplement noter si l'étape a été complétée ou non, l'échelle procédurale spécifique permet d'évaluer chacune de ces étapes au moyen d'une échelle de Likert (52). Une telle échelle a le potentiel d'évaluer de manière objective les techniques chirurgicales d'un résident et de lui fournir une rétroaction pertinente pour mener à son progrès (52).

Un exemple d'échelle procédurale spécifique apparaît sur la figure 3. Ici, les auteurs ont divisé la procédure évaluée, à savoir la colectomie du sigmoïde par laparoscopie, en étapes. Pour chacune de ces étapes, le candidat était évalué selon une échelle de Likert de 1 à 5, avec des descripteurs aux points 1, 3 et 5 (52)

6. Objectifs de l'étude

Nous nous sommes donc posés deux objectifs : le premier, celui de développer une telle échelle procédurale spécifique qui permettrait d'évaluer les techniques des résidents de chirurgie plastique lors de l'exécution d'une procédure de microchirurgie, et qui servirait un but formatif dans leur cursus ; et le deuxième, d'en étudier la validité et la fiabilité.

a. Pourquoi la microchirurgie ?

La microchirurgie a été introduite au monde médical il y a de cela plus d'un siècle, par Alexis Carrel, chirurgien et biologiste français, qui a emporté le prix Nobel en physiologie ou médecine en 1912 pour sa recherche sur les anastomoses vasculaires (19). Par la suite, dans les années 1960, la microchirurgie a connu de grandes avancées, en particulier grâce à Nakayama, un chirurgien cardio-thoracique japonais, qui a été le premier à décrire le transfère de segments intestinaux vascularisés dans la reconstruction de l'œsophage cervical suite à sa résection pour un traitement de néoplasie (19).

La microchirurgie a ensuite été popularisée en chirurgie plastique, dans un premier temps par son usage dans les anastomoses vasculaires pour la réimplantation digitale. Dans les années 70s, son usage s'est étendu aux cas cliniques électifs, en particulier avec le développement de lambeaux libres et de greffes nerveuses. Ainsi, le premier lambeau libre pour une reconstruction mammaire, un lambeau myocutané basé sur l'artère glutéale supérieure, a été performée en 1975 (34). Aujourd'hui la microchirurgie est devenu un outil indispensable dans l'artillerie d'un chirurgien plasticien (12).

La place qu'occupe la microchirurgie dans le monde de la chirurgie plastique ne peut que croître avec la popularisation récente des allo-transplantations tissulaires composites (34, 64).

Ces nouvelles applications chirurgicales de la microchirurgie augmentent son champ d'impact en plastie mais aussi mettent le patient à risque de morbidité postopératoire dans l'échéance où les techniques de microchirurgie ne sont pas maîtrisées (34).

Malgré des taux de succès actuels qui approchent les 95%, ces enjeux élevés associés avec la réussite des techniques de microchirurgie crée un besoin de développer des outils pour évaluer la compétence en microchirurgie (34). Ce besoin ne concerne pas seulement la chirurgie plastique, mais toutes les spécialités chirurgicales, puisque la microchirurgie fait partie intégrante de plusieurs domaines chirurgicaux (49, 64). Cette nécessité d'évaluer les techniques de microchirurgie grâce à des méthodes objectives est d'autant plus pressante que l'apprentissage de ces techniques est semé de difficultés (64).

Malheureusement, la formation en microchirurgie ne fait pas exception à la formation dans les autres domaines chirurgicaux, puisqu'elle suit elle aussi le modèle Halstedian (19, 65).

Nous avons choisi d'évaluer la technique de réparation de lacération d'un nerf digital (micro-neurorrhaphie), puisqu'elle requiert la maîtrise de techniques microchirurgicales de base, commune à toutes les procédures microchirurgicales, quelque soit la spécialité chirurgicale, telles que la manipulation de tissus délicats, en l'occurrence ici un nerf, et le contrôle d'instruments de microchirurgie (64). De plus, elle représente une tâche qui est facilement déconstruite en une séquence d'étapes critiques qui doivent être complétées (66), ce qui facilite son évaluation à l'aide d'une échelle procédurale spécifique.

En outre, il s'agit d'une procédure microchirurgicale commune dans le domaine de la chirurgie plastique, à laquelle le résident est exposé tôt dans sa formation, et qu'il doit maîtriser avant la fin de sa résidence, et ce même s'il décide de ne pas poursuivre une carrière microchirurgicale.

Enfin, il existe très peu de variantes anatomiques dans un doigt qui peuvent affecter la performance d'une micro-neurographie d'un nerf digital. Ainsi, la chirurgie est reproductible et sa performance reflète quasi exclusivement les habiletés techniques du résident (65).

Matériaux et méthodes

1. Revue de la littérature

Nous avons effectué une revue de la littérature dans le but de créer un répertoire des outils d'évaluation déjà développé par les différentes spécialités chirurgicales, et aussi dans celui de délimiter une méthodologie pour le développement de notre outil d'évaluation. Nous avons recherché la base de données Medline en utilisant les termes MeSH (Medical Subject Headings) suivant : « evaluation ; assessment ; instrument ; tool ; surgery ; skill ; teaching ». Les articles retenus sont ceux rédigés en français ou en anglais qui traitaient d'un outil d'évaluation des techniques chirurgicales de résidents, pendant que ceux-ci réalisaient une procédure. Les critères d'exclusion incluaient les articles dont le but était la validation d'un modèle de simulation et ceux qui utilisaient des analyses de mouvements de main ou de réalité virtuelle comme méthode d'évaluation. Nous avons fait une première sélection des articles basée sur leurs titres. Ainsi, nous avons rassemblé un total de 540 articles. Par la suite, nous avons lu les résumés de chacun de ces articles et nous en avons exclu 246 qui ne correspondaient pas à nos critères d'inclusion. Nous avons lu le texte des 294 articles restants, et vérifié leurs références pour des articles additionnels qu'on aurait omis lors de la revue de littérature. Nous avons retenu 84 articles qui respectaient les critères détaillés ci-haut.

Les données soutirées des articles retenues sont les suivantes : le titre de l'article et ses auteurs, le journal et l'année de publication, le pays dans lequel l'étude était conduite, le but

de l'étude, le développement de l'outil d'évaluation, le type et le niveau d'expertise des sujets sur lesquels l'outil était testé, le processus et la méthode de validation de l'outil, les analyses statistiques utilisées dans ce processus, la qualité de l'outil d'évaluation développé et son impact potentiel dans l'éducation des résidents.

2. Conception de l'outil d'évaluation

a. Décomposition de la chirurgie de micro-neurorrhaphie de nerf digital en une série d'étapes

Dans un premier temps, il fallait décomposer la procédure de micro-neurorrhaphie digitale en une série d'étapes successives, qui serait faciles à évaluer. Pour déterminer ces étapes, nous avons formé un groupe de discussion composé de trois chirurgiens expérimentés en chirurgie de la main et en microchirurgie (Drs AC, JL et DT). Ce groupe de discussion a produit une liste où apparaissait une série d'étapes successives qui constituent la chirurgie de micro-neurorrhaphie digitale (figure 4).

b. Étapes de la chirurgie retenues dans l'outil d'évaluation

Une méthode modifiée de Delphi a ensuite été employée pour arriver à un consensus d'experts sur les étapes critiques d'une micro-neurorrhaphie digitale.

La méthode de Delphi a été introduite par le RAND Corporation en 1948 (53). Il s'agit d'un processus anonyme dans lequel les participants sont sollicités pour leur opinion d'experts (53, 67). Leurs réponses sont rassemblées et analysées, et le processus se poursuit jusqu'à ce qu'un consensus de groupe soit atteint (53, 67). La méthode de Delphi a pour avantage d'éliminer la nécessité d'organiser une rencontre physique entre les différents experts. Elle améliore ainsi la

faisabilité de l'étude et en réduit les coûts. Elle permet également de faire appel à des experts de différents endroits géographiques. Enfin, son caractère anonyme élimine le risque d'avoir un expert dominant qui influence le consensus atteint (53). Lorsqu'utilisé dans le développement d'un outil d'évaluation, la méthode de Delphi permet d'en assurer la validité manifeste (18, 47, 49, 68, 69) et de contenu (4, 18, 47, 68, 69). Il n'existe pas de consensus dans la littérature quant au nombre minimal d'experts nécessaire pour conduire une méthode de Delphi.

Ainsi, nous avons envoyé un courriel à tous les chirurgiens du programme de chirurgie plastique et reconstructive de l'Université de Montréal, avec une surspécialisation en chirurgie de la main ou en microchirurgie. Cinq chirurgiens (Drs PB, AC, LL, SN et DT), ont accepté de faire partie du groupe d'experts. La liste de vérification préalablement construite (figure 5) leur a été envoyée par courriel. Nous avons demandé à ce que chacun d'entre eux révise et considère la pertinence de chacune des étapes de cette liste de vérification, la clarté de sa formulation et globalement celle de l'outil de l'évaluation. Chacun des experts a par la suite présenté son opinion avec des arguments la supportant dans un courriel adressé aux responsables de l'étude (Dr JL et SS). Ces opinions ont ensuite été compilées par ceux-ci et présentées à nouveau au groupe d'experts. Ceux-ci ont alors révisé leurs opinions respectives à la lumière de celles des autres chirurgiens. Ce processus a été répété jusqu'à ce qu'un consensus soit atteint par les experts. En somme, les experts étaient d'accord que les points importants de l'approche chirurgicale incluaient la planification de l'incision cutanée tout en respectant les principes généraux de la chirurgie de la main (incision de Bruner, incision midaxiale), la dissection de lambeaux robustes, et l'identification rapide des pédicules neurovasculaires. Cependant, le consensus d'experts voulait l'ajout d'un critère additionnel, à

savoir que l'incision cutanée soit complétée avec confiance, et non pas par de multiples passages du bistouri. Les experts étaient d'accord avec les deux étapes sous la catégorie « general microsurgery » et n'ont suggéré aucune modification à cette catégorie. La catégorie de « micro-neurorrhaphie » a été reformulée par les experts en 7 étapes qu'ils ont jugé pertinent de décrire avec plus de détails, à savoir, la microdissection des deux extrémités du nerf sectionné, la résection des extrémités nerveuses endommagées, la vérification de l'absence de tension entre les deux extrémités du nerf, la manipulation du nerf par son épinèvre, le passage de la suture à travers cet épinèvre seulement sans attraper des fascicules, l'exécution de nœuds carrés, et enfin la mise en place d'un nombre de sutures suffisant pour apposer les fascicules nerveux sans en causer l'extrusion. Les experts étaient d'accord d'inclure une étape sur l'adéquation de la fermeture cutanée. Ils ont également jugé que la manipulation d'instruments de microchirurgie ainsi que la manipulation des tissus pouvaient affecter le résultat microchirurgical. Les experts ont également jugé qu'il était adéquat d'inclure une question ouverte à la fin de l'outil d'évaluation dans laquelle l'évaluateur pouvait décider d'échouer un résident s'il jugeait que celui-ci avait commis une erreur qui pouvait affecter le résultat chirurgical.

Pour faire de notre outil une échelle procédurale spécifique, chacune de ces étapes devait être évaluée à l'aide d'une échelle de Likert allant de 1 à 5. Cette échelle de Likert permet d'attribuer une valeur quantitative à une donnée qualitative afin de pouvoir utiliser cette valeur dans des analyses statistiques (47). Les échelles de Likert vont souvent de 1 à 5 ou de 1 à 7, typiquement construite autour d'une option neutre (47). Les échelles de 0 à 3 sont trop étroites. De plus, un score de 0 est un symbole d'échec et lorsqu'attribué, peut être préjudiciable à l'apprenti et à la portée éducative de l'outil d'évaluation (20). Les échelles de

Likert à intervalle plus élargi laissent plus de place à l'amélioration de l'apprenti lorsque l'évaluation de la même tâche est répétée (20). Cependant, augmenter l'intervalle d'une échelle de Likert crée des difficultés quant au développement de critères définis pour chacun de ses scores (20). De tels critères aident les évaluateurs à attribuer aux apprentis des scores plus adéquats, ainsi permettant d'améliorer la fiabilité inter-évaluateur. Ils aident également à prodiguer aux apprentis une rétroaction précise et ciblée (20). Sur l'outil d'évaluation, nous avons conçus les échelles de Likert pour qu'elles mesurent exactement 8 cm. L'évaluateur plaçait un « X » sur l'échelle pour indiquer le score qu'il désirait assigner au candidat. Le score était par la suite déterminé numériquement en mesurant où sur l'échelle de 8 cm le X était placé (70).

À l'issue de ce processus, l'outil d'évaluation (figure 5) était composée de 18 étapes, chacune d'entre elles évaluées au moyen d'une échelle continue visuelle de Likert à 5 points. Chacun des points sur cette échelle était associé à un descripteur. Ces descripteurs étaient les mêmes pour chaque étape de l'outil d'évaluation.

3. Évaluation de l'outil développé

a. Création d'une installation pour tester l'outil d'évaluation

Pour tester l'outil d'évaluation, nous avons décidé de reproduire un environnement similaire à celui d'une salle d'opération dans lequel les candidats pouvaient effectuer une réparation de lacération de nerf digital du début jusqu'à la fin. Un tel environnement permet de garantir une certaine validité prédictive à notre outil. Pour ce faire, nous avons choisi d'utiliser, dans un environnement de laboratoire, des modèles humains cadavériques frais congelés de membre supérieur, s'étendant de l'épaule jusqu'aux doigts. En effet, un membre supérieur cadavérique

représente ce qu'il y a de plus proche d'un patient vivant (47). Angelo et al. ont démontré qu'un modèle cadavérique d'épaule avait une validité conceptuelle dans la comparaison de la performance d'une procédure arthroscopique de Bankart (47).

Chaque doigt d'un membre supérieur était utilisé pour évaluer une procédure et donc un participant. Nous avons préparé ces modèles cadavériques de manière identique, à savoir en effectuant une incision transverse par rapport à l'axe du doigt, sur la face palmaire de la phalange proximale du doigt. Cette lacération était portée jusqu'à l'os pour assurer que le pédicule neurovasculaire était lésé.

b. Étude pilote

Avant d'avoir des candidats effectuer la procédure de micro-neurorrhaphie digitale sur les modèles cadavériques ainsi préparés, nous avons voulu nous assurer de la faisabilité de la procédure en milieu de laboratoire, sur un modèle cadavérique. Ainsi, le chirurgien superviseur (Dr JL) a effectué, sous l'observation du résident assistant (SS), la procédure sur un modèle cadavérique. Cette étude pilote nous a permis de décider de l'emplacement des caméras pour capturer la procédure telle que pratiquée par un candidat, de décider des instruments nécessaires à la performance de la procédure, à savoir : un bistouri #15, des ciseaux de dissection Stevens, une pince Adson, un porte-aiguille, des ciseaux Iris ; et des instruments de microchirurgie, à savoir : une pince bijoutier, des ciseaux micro courbes et un porte-aiguille micro. Un fil de Nylon 9-0 pour la micro-neurorrhaphie et un fil de prolène 4-0 pour la fermeture cutanée étaient fournis. Nous avons également décidé de mettre à la disponibilité du candidat du ruban adhésif et des punaises pour qu'il puisse installer le membre supérieur dans une position propice au travail.

Cette étude pilote a également permis de réviser l'outil d'évaluation pour en assurer la faisabilité. Aucun changement ne lui a été apporté à ce stade-ci.

c. Création du matériel d'évaluation

Chaque candidat a eu comme instruction d'exposer la lacération nerveuse, puis de procéder à la micro-neurorrhaphie, tout en faisant abstraction des lacérations des structures adjacentes, à savoir l'artère digitale ou le tendon fléchisseur, s'il y avait lieu. Ils avaient à leur disposition les instruments chirurgicaux et les sutures nécessaires à la réalisation de la procédure (figure 6), ainsi qu'un microscope chirurgical (figure 7).

Nous avons préféré filmer les participants plutôt que de les avoir évalués par un chirurgien présent alors qu'ils réalisaient la procédure pour deux raisons. Dans un premier temps, filmer les participants permettait aux évaluateurs de noter les performances à tête reposée, sans qu'ils ne soient contraints par le temps. De plus, ce système de vidéo a permis d'établir un anonymat, ainsi éliminant les biais qui peuvent provenir de la reconnaissance d'un candidat et de son niveau d'expertise par l'évaluateur et qui peuvent donc affecter l'évaluation.

Toutes les vidéos de performance étaient produites de la même manière. Deux caméras étaient installées pour capturer la performance de la procédure. Une première caméra était installée sur un trépied face au participant et permettait de filmer la position du candidat évalué, sa posture et ses gestes pendant qu'il ou elle opérait sur le modèle cadavérique, sans toutefois dévoiler son visage et son identité. La position du candidat et sa posture alors qu'il performe la procédure microchirurgicale peuvent affecter la fluidité de ses mouvements et la précision de ses gestes chirurgicaux. En effet, il a été démontré dans d'autres disciplines nécessitant une coordination entre les yeux et les mains, comme les disciplines musicales ou

sportives, qu'un bon positionnement et une bonne posture sont associées avec une meilleure performance technique et un meilleur résultat (49). Cette caméra était démarrée lorsque la première incision était effectuée et n'était arrêtée que lorsque la fermeture cutanée était complétée. Une seconde caméra permettait de capturer le champ opératoire sous le microscope chirurgical. Elle filmait ainsi exactement ce que le candidat voyait à travers le microscope et permettait d'apprécier la précision des mouvements, essentiels à maîtriser, pour réaliser des procédures microchirurgicales. Ainsi, le produit vidéo final débutait avec le film de la première caméra pendant que le candidat planifiait l'incision, exposait la lacération nerveuse et la disséquait. Une fois que le candidat approchait le microscope, l'image principale devenait celle capturée par le microscope, alors que l'image de la caméra sur trépied était incorporée en plus petit format dans l'écran de manière à ne pas obstruer l'image enregistrée par le microscope. L'image de la première caméra permettait à ce moment-ci de la chirurgie d'évaluer la manipulation du microscope par le candidat, qui était un élément d'évaluation de la procédure (figure 8). Une fois la partie de la procédure, nécessitant le microscope, terminée, l'image principale redevenait celle filmée par la caméra sur trépied.

Tous les participants étaient habillés de blouses jaunes et portaient des gants chirurgicaux pour assurer leur anonymat aux yeux des évaluateurs. Dans la même optique, les vidéos ont été compilées de sorte que le produit final ne contenait pas de son qui pouvait divulguer l'identité des participants.

d. Recrutement de participants

Un courriel de recrutement, expliquant les grandes lignes du projet, a été envoyé à tous les résidents et patrons du programme de chirurgie plastique et reconstructive de l'Université de

Montréal. Le résident conduisant le projet de recherche (SS) a été exclu des participants potentiels. Cependant, le chirurgien superviseur (Dr JL) ainsi que certains chirurgiens experts ayant participé à la conception de l'outil d'évaluation par la méthode de Delphi, ont pris part dans la validation de l'outil.

La participation au projet de recherche se faisait sur une base strictement volontaire, et il était clairement expliqué que la performance selon l'outil d'évaluation n'aurait aucun impact sur la formation des candidats résidents. Cette information, entre autres, figurait sur un formulaire de consentement (figure 9) qui a été lu et signé par tous les participants volontaires.

Chacun des participants a donc effectué la micro-neurorraphie digitale du début à la fin sur un modèle cadavérique préparé de la manière décrite ci-haut, et a été enregistré par les deux caméras mentionnées ci-haut. Les participants n'ont bénéficié d'aucune orientation avant de performer la micro-neurorraphie et aucune assistance ne leur a été prodiguée pendant la procédure. Les vidéos ont par la suite été compilées sans édition, assignées un numéro au hasard puis évaluées par deux évaluateurs experts (Drs PB et JL), séparément, et notées selon l'outil d'évaluation développé. Les numéros associant une vidéo particulière au nom d'un participant figuraient sur une liste Excel à laquelle seul le résident impliqué dans le projet (SS) avait accès. Ce dernier n'avait aucun impact sur l'évaluation des vidéos. Quant aux évaluateurs, ils n'avaient connaissance ni de l'identité ni du niveau d'expertise des candidats dont ils évaluaient la performance.

Pour chaque candidat, la note totale obtenue selon l'instrument d'évaluation correspondait à la somme des scores obtenus à chaque étape selon l'échelle de Likert.

e. Analyses statistiques

Nous avons étudié la validité conceptuelle de l'outil d'évaluation en comparant les notes de participants de différents niveaux d'expertise au moyen du test d'analyse de variance à un facteur ANOVA avec un niveau significatif à 0.05.

La fiabilité inter-évaluateur de l'instrument a été déterminée en calculant le coefficient de corrélation intraclasse.

Enfin, la cohérence interne de l'outil d'évaluation a été déterminée au moyen du calcul du coefficient de Cronbach's alpha.

Les analyses statistiques ont été effectuées au moyen du logiciel statistique « IBM SPSS Statistics version 23 ».

Le projet de recherche a été approuvé par le comité d'éthique et de recherche de l'Hôpital Maisonneuve-Rosemont où le projet a pris place.

Résultats

1. Version finale de l'outil d'évaluation conçu

L'outil d'évaluation obtenu à l'issue du consensus du groupe d'expert a été testé avant d'être utilisé par les évaluateurs pour noter les vidéos. Le chirurgien superviseur du projet a ainsi utilisé l'instrument pour évaluer une vidéo choisie au hasard, dans le but d'en tester la compréhensibilité et la facilité d'utilisation. À l'issue de ce test, l'opinion du chirurgien superviseur était que la formulation des étapes d'évaluation était trop générale, peu spécifique et donc laissait trop de place à l'interprétation et à l'opinion subjective de l'évaluateur. En particulier, l'outil d'évaluation utilisé pour évaluer la même vidéo, par le même chirurgien responsable, à deux occasions différentes, donnait des résultats de performance variables. Ceci

pouvait donc mener au développement d'un instrument d'évaluation avec une fiabilité test-retest relativement faible.

Pour remédier à ce problème, l'outil d'évaluation a été reformulé afin de mieux orienter les évaluateurs potentiels. Une liste d'instructions a été ajoutée à l'entête de chaque formulaire d'évaluation afin de guider les évaluateurs dans leur emploi de l'outil. Les évaluateurs n'ont pas bénéficié de formation additionnelle pour l'utilisation de l'outil d'évaluation hormis ces instructions qui figuraient sur l'outil. De plus, pour chaque étape évaluée, nous avons attribué une liste de 3 éléments qui devaient être respectés. Sur l'échelle de Likert, une note de 1 à 2 devait être attribuée au candidat si aucun de ces trois éléments n'était respecté. Une note entre 2 et 3 devait être donnée si un des 3 critères était respecté. Une note entre 3 et 4 était donnée si le candidat respectait 2 critères sur 3 et enfin une note entre 4 et 5 pouvait être accordée si les 3 critères étaient respectés. Les évaluateurs étaient incités à faire usage de leur jugement d'expert pour décider où la performance d'un candidat se situait dans chacune des tranches prédéterminées. De plus, vis à vis de chaque étape à évaluer, une section de commentaires était disponible pour les évaluateurs (figure 10). Les descripteurs sur l'échelle de Likert et les 3 critères associés à chaque étape de l'outil d'évaluation avaient pour but non seulement de guider l'évaluateur dans son utilisation de l'outil, mais aussi de fournir au candidat une rétroaction spécifique et ciblée (71).

2. Évaluation de l'outil

Dix participants ont pris part dans le projet de recherche, parmi lesquels trois patrons experts, un résident PGY-1, un résident PGY-2, un résident PGY-3, deux résidents PGY-4 et deux résidents PGY-5 (table 1).

Chaque participant a été filmé alors qu'il ou elle effectuait la procédure de micro-neurographie digitale et chacune de ces vidéos a été évaluée par deux évaluateurs séparément. Pour chaque candidat, la note obtenue à l'évaluation était le résultat de l'addition des notes obtenues à chacune étape. Puisque l'outil d'évaluation final comptait 12 étapes, chacune notée selon une échelle de Likert qui variait entre 1 et 5, les notes pouvaient aller d'un minimum de 12 à un maximum de 60. Pour déterminer la note obtenue à chaque étape, nous avons mesuré où sur l'échelle de Likert la notation de l'évaluateur avait été faite. Puisque chaque échelle de Likert allait d'une note de 1 à 5, et que chaque échelle mesurait 8 cm, nous pouvions déterminer à quelle note entre 1 et 5 la notation de l'évaluateur correspondait. Les notes attribuées par l'évaluateur # 1 variaient entre 21.75 et 52.25. Celles attribuées par l'évaluateur # 2 variaient entre 14.7 et 60 (tableau I).

a. Fiabilité inter-évaluateur

Pour le calcul de la fiabilité inter-évaluateur, nous avons comparé pour chacun des candidats les notes obtenues selon l'évaluateur # 1 avec celles obtenues selon l'évaluateur # 2. Ceci a été fait au moyen du coefficient de corrélation intraclasse. Ce coefficient mesure la fiabilité en comparant la variabilité des évaluations d'un même candidat à la variabilité totale des résultats. Mais comment calculer ce coefficient ? Considérons une étude impliquant un nombre r d'évaluateurs qui doivent prendre m nombre de mesures sur n nombre de candidats ; y_{ijk} est la représentation abstraite du résultat quantitatif obtenu par le candidat i , lorsqu'évalué par l'évaluateur j à l'occasion k . Ce modèle se traduit mathématiquement par la formule suivante :

$$y_{ijk} = \mu + s_i + e_{ijk}$$

où μ représente la valeur moyenne, s_i est l'effet du sujet i , et e_{ijk} est l'effet de l'erreur. Ces deux effets sont considérés aléatoires, indépendants et suivent une distribution normale avec une moyenne de 0 et des variances respectives σ_s^2 et σ_e^2 .

Le coefficient de corrélation intraclasse, comme mesure de la fiabilité inter-évaluateur, est une mesure de la corrélation entre deux résultats y_{ijk} et $y_{ij'k}$ pour un même candidat i , une même occasion k , mais par deux évaluateurs différents j et j' . Ce coefficient, nommé ρ , est calculé par la formule mathématique suivante

$$\rho = \sigma_s^2 / (\sigma_s^2 + \sigma_e^2)$$

et représente donc la fraction de la variance du sujet par la variance totale qui est la somme de la variance du sujet et de la variance de l'erreur (72).

À noter, que pour le calcul de ce coefficient nous avons exclus les résultats du candidat # 8. En effet, son évaluation par l'évaluateur # 2 était incomplète avec deux étapes non évaluées puisque cet évaluateur considérait la vidéo de mauvaise qualité, rendant l'évaluation de ces deux étapes difficiles. Le coefficient de corrélation intraclasse ainsi calculé était de 0.690.

b. Cohérence interne

La cohérence interne de l'outil d'évaluation a été calculée au moyen du coefficient de Cronbach alpha. Si k est le nombre de questions dans l'outil d'évaluation, s_i^2 la variance associée avec la question i et s_T^2 la variance associée avec la somme des notes à toutes les questions k , le coefficient de Cronbach alpha est défini par la formule mathématique suivante :

$$\alpha = \frac{k}{k-1} \left(1 - \frac{1}{s_T^2} \sum_{i=1}^k s_i^2 \right)$$

(73). Celui-ci s'élevait à une valeur de 0.960. Nos analyses ont également démontré qu'il n'y aurait aucune amélioration dans le coefficient de Cronbach alpha, si une des étapes du test était omises (tableau II).

c. Validité conceptuelle

Enfin, nous avons évalué si les résultats des candidats variaient en fonction de leur niveau d'expertise. Pour ce faire, les candidats ont été divisés en trois groupes de niveau d'expertise différent : un groupe de résidents juniors rassemblant les résidents PGY-1 et PGY-2, un groupe de résidents séniors (PGY-3, PGY-4 et PGY-5) et un groupe de patrons. Les résultats de ces candidats ont été comparés selon le test d'analyse de variance à un facteur (ANOVA). Celui-ci a révélé une différence statistiquement significative entre les résultats des candidats par niveau d'expertise ($p=0.017$) (figure 11). Une analyse post-hoc a révélé qu'il y avait une différence dans les résultats des candidats qui était statistiquement significative entre le groupe de résidents juniors et le groupe de résidents séniors d'une part ($p=0.032$), et entre le groupe de résidents juniors et le groupe de patrons d'autre part ($p=0.017$). Cependant, bien que le groupe de patrons ait mieux performé que le groupe de résidents séniors, cette différence n'était pas statistiquement significative ($p=0.626$) (tableau III).

Discussion

1. Évaluation en chirurgie

La compétence chirurgicale dépend autant des connaissances cliniques que de la maîtrise des habiletés techniques (1, 4, 20). Les connaissances cliniques sont traditionnellement évaluées au moyen d'examens écrits et oraux (1). Ceux-ci ne permettent toutefois pas d'apprécier la maîtrise des techniques chirurgicales puisqu'il a été démontré qu'il n'y avait aucune corrélation entre la note obtenue à l'examen ABSITE et les techniques chirurgicales (4, 19). Bien que les habiletés techniques d'un chirurgien aient un impact important sur le succès chirurgical (11), leur évaluation se fait au moyen de méthodes relativement subjectives (11, 37), qui ne permettent pas de donner au résident une rétroaction structurée et suffisante pour formuler un plan de remédiation (32). Ces méthodes incluent entre autres des journaux de bord qui documentent le nombre de procédures auxquelles un résident a été exposé, mais qui ne reflète pas sa capacité à effectuer, de manière indépendante, ces procédures (2). Cependant, pour éviter les erreurs chirurgicales, il est important de développer des méthodes objectives pour quantifier la compétence chirurgicale afin de générer une rétroaction utile (4, 11). De plus, la complexité croissante des nouvelles techniques chirurgicales (4) combinée à la réduction de l'exposition clinique dans le cadre de la diminution du temps de travail des résidents (4, 32, 47, 48), appuient ce besoin de développer de nouvelles méthodes objectives pour évaluer la compétence et la performance des résidents en spécialité chirurgicale (4).

L'acquisition des techniques de microchirurgie ne fait pas exception à ce phénomène, et suit, elle aussi, un modèle d'apprenti subjectif (65). L'évaluation objective des techniques de microchirurgie est d'autant plus importante que la microchirurgie fait partie intégrante de plusieurs disciplines chirurgicales, telles que la neurochirurgie, la chirurgie vasculaire, ou la chirurgie de la base du crâne (49). En particulier, la microchirurgie est un atout indispensable dans l'artillerie d'un chirurgien plasticien puisqu'elle joue un rôle essentiel non seulement

dans les reconstructions tissulaires complexes (12) mais aussi dans les réimplantations de doigts amputés (34). De plus, la microchirurgie connaît actuellement une expansion de son impact avec la popularisation récente des allotransplantations tissulaires composites (34).

2. Méthodes objectives d'évaluation des techniques chirurgicales

Une revue de la littérature des évaluations des techniques chirurgicales effectuées par Ahmed et al a identifié trois catégories principales d'outil d'évaluation (4).

a. Échelles d'évaluation globale

Ces échelles permettent d'évaluer des techniques chirurgicales génériques, qui ne sont pas propres à une procédure chirurgicale spécifique (4). Parmi ces échelles, la plus populaire est probablement celle qui a été développée par Martin et al, le Global Rating Scale (GRS) comme composante de l'outil OSATS (Objective Structured Assessment of Technical Skill) qui incluait également une liste de vérification (4). Cette échelle a été utilisée dans sa version originale, ou légèrement altérée, non seulement dans des procédures de chirurgie générale, autant dans les chirurgies ouvertes que laparoscopiques que robotiques, mais aussi dans des procédures de gynécologie, de chirurgie cardiaque, orthopédique ou ophtalmologique (4). Cette échelle globale a démontré de manière constante un bon niveau de validité conceptuelle et de fiabilité (4). Cependant, il a été observé que la validité de l'outil GRS atteignait un plafond parmi les chirurgiens experts (4). En effet, les habiletés techniques, telles que mesurées par le GRS, ne permettent pas de distinguer les différents niveaux d'expérience parmi les chirurgiens experts (4).

Un autre outil d'évaluation, le Global Operative Assessment of Laparoscopic Skills (GOALS) a été étudié de manière extensive pour les procédures laparoscopiques (4). Il a été démontré à plusieurs reprises comme étant un outil faisable, et avec une grande fiabilité et validité (4, 35). Il a également été démontré comme supérieur aux listes de vérification pour l'évaluation des techniques chirurgicales (4). Cependant, tout comme le GRS, GOALS atteint un plafond dans l'évaluation des experts (35).

Les échelles d'évaluation globale font usage d'une échelle de Likert avec des points d'ancrage qui servent de repères aux évaluateurs (19). Cependant, les échelles de Likert requièrent une certaine mesure de subjectivité des évaluateurs (19, 49). Ceci peut affecter la fiabilité de l'outil d'évaluation (19). Toutefois, il a été démontré que les chirurgiens évaluateurs ont une conception commune de la réussite et de l'échec chirurgical lorsqu'il s'agit d'évaluer des techniques chirurgicales (26, 41).

Par ailleurs, les outils d'évaluation faisant usage d'une échelle de Likert sont susceptibles aux erreurs de tendance centrale, c'est à dire que les évaluateurs sont souvent réticents à attribuer des notes basses aux candidats (27, 46). Pour remédier à ces deux problèmes, augmenter la constance des évaluations par les deux évaluateurs et donc assurer une meilleure fiabilité inter-évaluateur, il a été suggéré d'entraîner les évaluateurs à utiliser l'outil d'évaluation (19, 46).

Les échelles d'évaluation globale sont également critiquées du fait de leur forme générique qui empêche d'identifier exactement où se situent les lacunes de l'apprenti, et donc de prodiguer une rétroaction ciblée qui guide l'amélioration des techniques chirurgicales (35).

b. Listes de vérification

Avec les listes de vérification, il ne s'agit non pas de déterminer si les techniques chirurgicales sont effectuées de manière adéquate, mais si elles sont effectuées ou pas (19). Elles font donc de l'évaluateur un observateur et non pas un interprète d'une performance chirurgicale, ce qui élimine tout risque de subjectivité (19). Ces listes de vérification sont spécifiques à une procédure chirurgicale particulière. Des listes de vérification ont été décrites pour plusieurs procédures chirurgicales, incluant l'endoscopie gastro-intestinale, la fundoplication de Nissen laparoscopique, la colectomie du sigmoïde, et des procédures vasculaires et de thoracotomie (4) et ont été démontrées faisables, valides et fiables (4). Cependant, tout comme les échelles d'évaluation globale, les listes de vérification atteignent un plafond lorsqu'il s'agit de distinguer la performance de chirurgiens experts (4). De plus, il est souvent noté qu'elles ont une moins bonne fiabilité que les échelles d'évaluation globale (57). Elles sont également critiquées car elles font preuve d'une validité conceptuelle inférieure à celles des échelles d'évaluation globale (2, 24). En effet, puisqu'elles divisent une procédure particulière en une séquence rigide d'étapes, les novices, qui ont tendance à approcher une procédure en la décomposant en une série d'étapes, ont tendances à mieux performer que des experts qui approchent une procédure comme un but à atteindre (19). Cependant, par le même fait qu'elles divisent une procédure en une série d'étapes, elles permettent d'identifier quelles étapes représentent une difficulté pour le résident évalué et amènent donc un avantage éducatif (2, 13, 24).

c. Notre outil d'évaluation

Nous avons développé un outil d'évaluation que l'on décrit comme une échelle d'évaluation procédurale. Cet outil est à mi-chemin entre une échelle d'évaluation globale et une liste de

vérification. Il décompose la chirurgie de micro-neurorrhaphie digitale en une série d'étapes, tout comme une liste de vérification, dont la performance est évaluée selon une échelle de Likert, à l'image d'une échelle d'évaluation globale. Le but derrière la conception d'un tel outil est de combiner la force de validité et de fiabilité d'une échelle d'évaluation globale avec la capacité d'une liste de vérification d'identifier les lacunes spécifiques d'un résident.

La cohérence interne de notre outil d'évaluation, telle que calculée par un coefficient de Cronbach alpha de 0.960, était élevée, ce qui signifie qu'il existe une excellente corrélation entre les différentes étapes du test et que celles-ci contribuent à évaluer le même objet, ici la micro-neurorrhaphie digitale (44, 46). De plus, aucune amélioration ne pouvait être apportée à la cohérence interne du test en retirant une des étapes de l'instrument développé.

Le niveau nécessaire de fiabilité inter-évaluateur d'un outil d'évaluation dépend de l'objectif de l'outil (74). La fiabilité inter-évaluateur de notre outil d'évaluation, estimée au moyen du coefficient de corrélation intra-classe, était de 0.690. Ceci correspond à une fiabilité acceptable pour un outil dont l'objectif est d'optimiser la formation des résidents en chirurgie plastique (60), de suivre leur progrès et de leur donner des rétroactions structurées (74). En effet, Downing estime que pour de telles évaluations à but formatif et administré par des membres locaux de la faculté, une fiabilité de l'ordre de 0.70 à 0.79 est attendue (60).

Cependant, pour des outils d'évaluation à enjeu modéré, comme des examens de fin d'année, on s'attend à une fiabilité dans les 0.80 à 0.89 (60). Pour les outils d'évaluation qui aboutissent à la certification ou l'accréditation d'un médecin, une fiabilité de 0.90 ou plus est exigé (60). Plusieurs méthodes ont été décrites pour améliorer la fiabilité d'un outil d'évaluation. Il a été mentionné que entraîner les évaluateurs et leur donner des instructions simples et précises peut améliorer la fiabilité de l'outil (75, 76). Aussi, augmenter le nombre d'observations pour

chaque résident permet d'améliorer la fiabilité (11, 76). De plus, évaluer plus qu'une procédure par résident peut également augmenter la fiabilité (60). Les ancrages au niveau de l'échelle de Likert font de l'outil un instrument plus objectif et contribuent également à en augmenter la fiabilité (35). En effet, dans l'étude de Gofton et al, les évaluateurs ont indiqué que ces ancrages ont facilité l'évaluation des apprentis parce qu'ils leur évitaient de devoir décider si la performance de ceux-ci était en-dessous ou au-dessus des attentes, ou si elle rencontrait les attentes (21, 26). En somme, plus les évaluateurs sont familiers avec l'outil d'évaluation, plus la fiabilité de l'outil d'évaluation est haute (16).

Nos analyses statistiques ont démontré que l'outil d'évaluation développé, démontre une validité conceptuelle puisqu'il existe une différence statistiquement significative entre les résultats de candidats de différents niveaux d'expertise ($p=0.017$). En effet, les participants ont été séparés en trois groupes selon leur niveau de formation : un groupe de résidents juniors (PGY-1 et 2), un groupe de résidents séniors (PGY-3, 4 et 5) et un groupe de chirurgiens experts. Les résidents séniors ont eu des résultats de performance meilleurs que ceux des résidents juniors ($p=0.032$) et les patrons ont mieux performés que les résidents juniors ($p=0.017$). Cependant, la différence de performance entre les résidents séniors et les patrons n'a pas atteint un niveau statistiquement significatif ($p=0.626$). Ceci peut s'expliquer par le fait que le groupe de résidents séniors rassemblait trois différents niveaux d'expertise avec des résultats très variables. De plus, le petit nombre de candidats participants à l'étude a fait en sorte que le calcul des moyennes était facilement affecté par les valeurs extrêmes. De ce fait, des résidents séniors avec des habiletés techniques au-dessus de la moyenne (eg. Candidat # 7) peuvent affecter la moyenne globale du groupe des résidents séniors ce qui rend la différence avec la moyenne du groupe des chirurgiens experts statistiquement non significative.

Nous aurions également pu étudier la validité conceptuelle de notre outil en évaluant une corrélation potentielle entre la performance des candidats et leur expérience passée en microchirurgie (65). D'autres ont démontré la validité conceptuelle de leur outil en corrélant le score obtenu par le candidat au temps qu'il a pris pour compléter la procédure (77). Nous avons intentionnellement omis d'évaluer une telle corrélation puisque nous sommes d'avis que le temps pris pour compléter une procédure peut être affecté par plusieurs facteurs, pas seulement par l'expertise du candidat, et ne reflète pas la qualité de la performance de la procédure (19).

La validité manifeste et de contenu de l'outil ont été établies lorsque l'opinion d'experts en microchirurgie a été suscitée pendant la conception de l'instrument de l'évaluation (21, 69).

d. Combinaison d'une échelle d'évaluation globale et d'un outil spécifique à une procédure

Ceci correspond à la troisième catégorie d'outil identifiée par la revue de littérature menée par Ahmed et al. Une combinaison d'outils représente peut-être une approche plus compréhensive de l'évaluation des résidents (4).

3. Situation propice à l'évaluation du résident

a. Salle d'opération

L'évaluation du résident basée sur l'observation de ses habiletés techniques en salle d'opération représente l'étalon d'or en terme de validité de contenu et manifeste (5, 18, 23).

Un outil valide et fiable qui permet d'évaluer les techniques chirurgicales en salle d'opération a plusieurs avantages. Parmi ceux-ci, le plus important est que toutes les questions sur le

réalisme d'une situation créée en laboratoire ne se posent plus (27). Cependant, une performance chirurgicale en salle d'opération peut être affectée par plusieurs variables. Souvent en salle d'opération, de l'aide verbale et/ou physique est apportée au résident par le chirurgien superviseur ce qui peut teinter la performance du résident et dissimuler son réel niveau d'expertise (11, 23, 78). De plus, lorsque l'évaluation est réalisée en salle d'opération, l'anonymat du résident ne peut être assuré (11, 27). Ainsi, son évaluation peut être biaisée par une opinion préconçue que l'évaluateur a sur lui/elle (11, 27). En outre, si le chirurgien complète l'évaluation à distance de la performance du résident, l'évaluation peut être sujette à un biais de mémoire (11, 62, 78). En effet, Laeeq et al ont démontré que les évaluations complétées en dedans de six jours de la performance ont une validité conceptuelle plus grande que celles des évaluations complétées après cette limite de six jours (11). Enfin, le simple fait d'être observé peut affecter la performance du résident évalué (6). Avoir une audience peut d'une part, pousser certains résidents à vouloir plaire et donc à mieux performer et d'autre part, susciter de l'anxiété chez d'autres résidents qui verraient alors leur performance se détériorer (2). Cependant, on peut argumenter que l'observation fait partie intégrante de la réalité d'un chirurgien (2), puisqu'alors qu'il réalise une opération, il est souvent observé par des infirmières ainsi que l'équipe anesthésique, bien que ce ne soit pas dans un but d'évaluation.

b. Enregistrement vidéo

L'usage de vidéos peut remédier à plusieurs de ces problèmes. Il s'agit d'une méthode souvent utilisée par les athlètes professionnels et leur coach pour évaluer leur performance, améliorer leur technique et augmenter leur confiance (28, 79). Dans le domaine chirurgical,

l'enregistrement vidéo permet d'évaluer la compétence chirurgicale des résidents de manière impartiale (28, 54, 60, 80) puisqu'il permet d'assurer l'anonymat de ceux-ci (44, 45, 80).

L'enregistrement vidéo d'une procédure peut contribuer à augmenter la faisabilité d'un outil d'évaluation (80) et ce de plusieurs manières. En effet, d'une part, il permet à l'évaluateur d'effectuer l'évaluation à tête reposée (12, 45, 57). D'autre part, il bénéficie aussi de l'option d'avancer la vidéo réduisant ainsi le temps d'évaluation de 50%, sans risque d'omettre des parties importantes de la procédure (5, 45).

Une vidéo permet aussi d'augmenter la fiabilité d'un outil d'évaluation en permettant à l'évaluateur de suspendre ou rembobiner la vidéo, ainsi augmentant la précision de l'évaluation (80). La fiabilité peut d'autant plus être améliorée que les vidéos peuvent être évaluées par plusieurs évaluateurs de manière indépendante (54, 81).

La vidéo peut être visualisée par l'apprenti lui-même, dans un but de réviser sa performance à la lumière de la rétroaction qu'il/elle reçoit (3, 12, 57).

Une excellente corrélation a été démontrée par Beard et al entre l'observation directe et l'évaluation aveuglée d'une vidéo de dissection saphéno-fémorale (2, 5, 45). Ceci supporte le fait qu'une procédure peut être enregistrée et évaluée, omettant ainsi le besoin d'observation directe (5). Scott et al a trouvé qu'il n'y avait pas de corrélation entre l'observation directe et les vidéos éditées dans l'évaluation globale de cholécystectomie laparoscopique (5, 45, 80).

Nous avons donc choisi de ne pas éditer les vidéos des participants à notre étude afin d'éviter d'introduire un biais de sélection (82).

Les enregistrements vidéo ont cependant été critiqués car ils ne permettent de capturer que le champ opératoire et ne permettent pas d'apprécier ce qui se déroule dans l'environnement de la salle d'opération, bien que ceci puisse affecter la performance d'une procédure chirurgicale

(6, 35, 44, 45). Cependant, nous avons adressé cette limitation en installant une seconde caméra qui capturerait la posture du candidat, sa manipulation des instruments et du microscope. Les enregistrements vidéos rajoutent également au coût de ces évaluations (62).

c. Enregistrement vidéo en salle d'opération

Un moyen d'évaluer les résidents en salle d'opération tout en maintenant les avantages qu'offre l'enregistrement vidéo est de filmer les résidents pendant qu'ils performant en salle d'opération. Cependant, le champ opératoire peut être dissimulé par les personnes autour de la table opératoire. De plus, il peut être difficile de positionner la caméra à proximité de la table opératoire lorsque celle-ci est entourée de tables d'instruments et de matériel chirurgical (45). Berger et al ont conçu un moyen de contourner ces difficultés en construisant des caméras que les résidents peuvent porter autour de leur tête alors qu'ils opèrent (28). Ceci permet de capturer le déroulement chirurgical tel que perçu par la personne qui opère (28). Ceci permet ainsi de bénéficier de tous les avantages que présentent les enregistrements vidéo.

4. Impact d'un tel outil d'évaluation

a. Impact éducatif

Un tel outil d'évaluation peut avoir plusieurs impacts bénéfiques sur la formation des résidents. Dans un premier temps, il aide à diminuer la subjectivité du processus d'évaluation en définissant clairement pour l'évaluateur les éléments à observer lors de la performance d'une procédure (21, 68). L'outil profite également à l'apprenti, puisqu'il définit clairement ce qui est attendu d'une performance adéquate (21, 68).

D'autre part, il permet d'identifier les faiblesses de chaque résident et par conséquent concevoir un plan de remédiation spécifique au résident (11, 33, 41, 47, 57). Sans évaluation objective et rétroaction utile, l'apprenti est à risque de répéter les mêmes erreurs, sans jamais avoir l'opportunité de devenir compétent (8). De plus, il aide à suivre la progression d'un résident durant chacune de ses rotations, et durant sa résidence (9, 11, 13, 40, 65, 74, 83). En fin de résidence, il permettrait de pousser les résidents seniors à travailler sur des points spécifiques pour acquérir un niveau d'expert (11). Il peut également servir à étudier l'efficacité de formations académiques en évaluant les résidents avant et après ces formations (65, 74, 84).

b. Utilisation pour l'accréditation

Il a été suggéré par plusieurs qu'un tel outil d'évaluation peut éventuellement être incorporé dans le processus d'accréditation des résidents finissants et dans le processus de ré-accréditation des chirurgiens déjà en pratique (33, 54, 74). La ré-accréditation a pour but d'améliorer les services de santé en évaluant de manière objective la qualité des soins apportés par les cliniciens en pratique (6). Cependant, l'usage de ces outils dans cet optique serait difficile vu le plafond dont font preuve ces outils dans la distinction du niveau de performance d'experts (11).

De plus, tout comme notre étude, la majorité des groupes ayant développé des instruments d'évaluation, s'attardent sur la démonstration de la fiabilité et de la validité conceptuelle de l'outil (46). Très peu d'études ont cherché à démontrer d'autres niveaux de validité (46). Pour cette raison, il a été suggéré de limiter l'usage de ces outils à la formation des résidents, et non pas à leur accréditation (46). Avant d'introduire ces outils dans les domaines d'accréditation

ou de ré-accréditation, il est suggéré de les soumettre à des processus de validation plus extensifs (46, 70) et de déterminer leur impact sur l'éducation des résidents (46).

5. Autres outils d'évaluation de la microchirurgie

L'instrument que nous avons développé est le premier instrument conçu spécifiquement pour évaluer les procédures de micro-neurorrhaphie digitale.

Par ailleurs, nous avons identifié deux autres outils d'évaluation développés en microchirurgie, au travers de notre revue de littérature. Notamment, Temple et al ont fait usage de cuisses de poulets commerciales sur lesquelles les candidats avaient à effectuer des microanastomoses de vaisseaux. Ce modèle a été utilisé pour évaluer la validité et la fiabilité d'un outil d'évaluation de microchirurgie, le « University of Western Ontario Microsurgical Acquisition/Assessment (UWOMSA) instrument » (figure 12) développé par le groupe. Celui-ci a fait preuve de validité conceptuelle, en démontrant une association entre des scores élevés d'une part, et un niveau d'expertise plus avancé ainsi qu'un temps chirurgical plus rapide d'autre part. Sa fiabilité inter-évaluateur était de 0.75 avec deux évaluateurs examinant les performances (77). Notre outil a pour avantage sur celui-ci d'avoir été testé sur un modèle de cadavre humain qui se rapproche plus de la réalité d'un cas en salle d'opération qu'un modèle de vaisseau sur une cuisse de poulet. Il démontre donc une validité manifeste et prédictive plus élevées que celle de l'instrument UWOMSA. De plus, notre outil se rapporte à une procédure spécifique et non pas à des gestes de microchirurgie généraux et a donc plus d'impact comme outil de formation.

Chan et al ont développé un outil d'évaluation de type échelle d'évaluation globale, le « Structured Assessment of Microsurgery Skills (SAMS) » (figure 13), qui évalue les

techniques générales, nécessaires à la performance d'une microanastomose. Une liste d'erreur a été ajoutée à l'outil d'évaluation afin de pouvoir prodiguer des rétroactions ciblées. L'outil a été testé sur des enregistrements vidéo, anonymes, de microanastomoses veineuses effectuées sur des patients en salle d'opération. Cet outil a fait preuve de validité conceptuelle avec des résultats démontrant une différence statistiquement significative entre le groupe expert et le groupe de résidents, et parmi les résidents évalués au début et à la fin de leur année de formation. L'outil a également démontré une très bonne fiabilité inter-évaluateur à 0.78. L'obtention d'un taux de fiabilité aussi élevé s'explique probablement par la sollicitation d'un grand nombre d'évaluateurs (cinq) (12). L'avantage de la méthodologie utilisée par Chan et al est que l'outil a été directement testé dans un contexte de salle d'opération. Ses validités manifeste et prédictive sont donc irréprochables. Cependant, il se rapporte, tout comme l'outil de Temple et al, à des gestes de microchirurgie généraux, et bien qu'une liste d'erreur (figure 14) ait été ajoutée à l'outil dans le but d'en faire un meilleur outil formatif, nous jugeons que notre outil permet de mieux discerner les faiblesses d'un résident et donc a plus de poids comme outil pédagogique.

Selber et al ont utilisé ce même outil, SAMS, pour suivre la progression des techniques chirurgicales de sept fellows en microchirurgie pendant leur année de formation dans le centre « MD Anderson ». Les fellows participant ont été évalués par des chirurgiens experts pour tout les cas de microchirurgie dans lesquels ils prenaient part pendant une période d'un mois. Ce processus a été répété trois fois pendant leur année de formation pour pouvoir suivre leur progression. Tous les fellows ont démontré une amélioration dans leurs techniques de microchirurgie statistiquement significative pendant les six premiers mois de leur formation. Après cette période, la progression s'est maintenue mais n'atteignait plus la signification

statistique. Une deuxième partie de l'étude avait pour but de valider l'outil dans un modèle de laboratoire. Les sept fellows ont effectué au début et à la fin de leur année de formation, la microanastomose d'une artère fémorale sur un rat. Leurs performances ont été filmées en dissimulant leur identité, et ces vidéos ont été évaluées au moyen du SAMS par cinq évaluateurs. La fiabilité inter-évaluateur de cette deuxième portion de l'étude s'élevait à 0.67, ce qui était considéré comme acceptable. De plus, l'outil a démontré qu'il y avait une amélioration, statistiquement significative, de la performance des candidats entre la performance accomplie au début de l'année et celle effectuée à la fin de leur année de formation (66).

6. Nouvelles tendances dans l'évaluation des techniques chirurgicales

La tendance récente dans les instruments d'évaluation des techniques chirurgicales est l'usage d'analyse de dextérité au moyen de système d'analyse de mouvements. Ces systèmes font typiquement usage d'un ordinateur qui mesure le temps, le nombre total de mouvements, et la distance parcourue par les mains du candidat (19). Il a été démontré que ce genre de système reflète la qualité de la performance technique en microchirurgie et qu'il corrèle avec les taux de fuites anastomotiques (19). L'avantage de ces systèmes est qu'ils offrent une méthode très précise et entièrement objective de mesurer les habiletés techniques (48). Le jugement d'un expert, et le risque de subjectivité qu'il apporte, est complètement éliminé (19). Cependant, de tels systèmes sont coûteux et difficilement applicable en salle d'opération (48). En effet, ils nécessitent l'installation de traqueurs électromagnétiques sur le dos des mains du chirurgien ce qui peut facilement contaminer le champ opératoire et diminuer la dextérité du

chirurgien, en particulier dans une procédure microchirurgicale où la précision des gestes chirurgicaux a un impact important sur le succès de l'opération (12).

7. Limitations de notre étude

a. Conception de l'outil d'évaluation

Plusieurs chirurgiens experts qui ont aidé dans la conception de l'outil ont aussi pris part dans l'étude comme participants. Ceci aurait pu faussement augmenter leurs performances. En effet, ce phénomène, appelé l'effet Hawthorne, correspond à l'amélioration de la performance d'une tâche lorsqu'on sait d'avance ce qui est évalué (49).

b. Utilisation de modèles cadavériques

Bien qu'un modèle humain frais congelé de membre supérieur est ce qui se rapproche le plus de la réalité de la salle d'opération, le spécimen cadavérique n'a pas la même qualité de tissu que le modèle vivant (47). De plus, il ne saigne pas à la manière des tissus humains vivants, il est associé à un coût élevé, et est souvent difficile à procurer (58). En outre, certains modèles cadavériques étaient certainement moins bien préservés que d'autres, ce qui aurait pu affecter la performance de certains candidats (47).

De plus, bien que nous ayons tenté d'effectuer la lésion nerveuse de la même manière sur tous les cadavres en effectuant l'incision à l'aide d'un bistouri #15, sur la face palmaire de la phalange proximale des doigts, il nous était impossible d'effectuer la lésion nerveuse de manière uniforme (47). Dans certains cas, une fois que le candidat avait exposé le pédicule neurovasculaire, nous nous sommes aperçus que le nerf n'était pas divisé, auquel cas nous

l'avons lésé sous observation directe. On peut reprocher que ceci aurait pu affecter la performance du candidat.

c. Matériel audio-visuel

Des dysfonctionnements des appareils audio-visuels ont engendré des mauvaises qualités d'enregistrement pour certaines performances, notamment celle du candidat #8. Pour ce dernier, l'évaluation complétée par l'évaluateur #2 était incomplète pour deux critères sur douze. Les évaluations de ce candidat ont donc été exclues dans le calcul de la fiabilité inter-évaluateur.

De plus, la caméra intégrée au microscope capturait une image rectangulaire alors que le candidat qui travaille sous le microscope perçoit une image circulaire. Ceci a fait en sorte que si le candidat travaillait dans la périphérie de son champ circulaire, ses mouvements n'étaient que partiellement capturés par le champ rectangulaire de la caméra du microscope.

d. Validation de l'instrument

Notre méthodologie de validation de notre outil d'évaluation est incomplète. En effet, nous n'avons démontré que la validité conceptuelle, manifeste et de contenu de l'outil, ainsi que sa fiabilité inter-évaluateur et sa cohérence interne. Cependant, pour compléter la validation de l'outil, nous devons également en étudier la fiabilité test-retest et les validités concurrente et prédictive.

Un moyen d'étudier la fiabilité test-retest est de sélectionner, de manière aléatoire, un nombre limité de vidéos à faire réévaluer par un des deux évaluateurs. Ceci doit être fait après une certaine période de temps pour éviter que l'évaluateur ne se souvienne des vidéos. Les

résultats de cette deuxième évaluation de ces vidéos pourraient alors être comparés aux résultats obtenus à leur évaluation initiale par le même évaluateur. La fiabilité test-retest serait prouvée si une corrélation significative était démontrée entre ces deux séries de résultats (79). La validité concurrente pourrait être étudiée en évaluant la corrélation entre les résultats obtenus par notre outil d'évaluation avec les résultats obtenus selon un outil d'évaluation préalablement évalué.

Conclusion

En somme, nous avons développé un outil d'évaluation des techniques de microchirurgie des résidents de plastie lors de l'exécution d'une micro-neurorrhaphie digitale. Sa fiabilité et sa validité en font un outil qui pourrait être intégré dans la formation des résidents, comme instrument d'évaluation mais aussi comme outil d'enseignement et d'évaluation de l'utilité d'interventions pédagogiques. Nous sommes d'avis que la diffusion de la méthodologie qui a abouti au développement et à la validation de cet outil d'évaluation alimentera la recherche pour créer de nouvelles avenues qui mèneront éventuellement à un curriculum chirurgical scientifique et objectif, pour ultimement assurer et améliorer la sécurité de nos patients.

Figures

Figure 1 : Exemple d'échelle d'évaluation globale (63)

Please rate the candidate's performance on the following scale:

	1	2	3	4	5
Respect for tissue	Frequently used unnecessary force on tissue or caused damage by inappropriate use of instruments.		Careful handling of tissue but occasionally caused inadvertent damage.		Consistently handled tissues appropriately with minimal damage.
Time and motion	Many unnecessary moves.		Efficient time/motion but some unnecessary moves.		Economy of movement and maximum efficiency.
Instrument handling	Repeatedly makes tentative or awkward moves with instruments.		Competent use of instruments although occasionally appeared stiff or awkward.		Fluid moves with instruments and no awkwardness.
Knowledge of instruments	Frequently asked for the wrong instrument or used an inappropriate instrument.		Knew the names of most instruments and used appropriate instrument for the task.		Obviously familiar with the instruments required and their names.
Use of assistants	Consistently placed assistants poorly or failed to use assistants.		Good use of assistants most of the time.		Strategically used assistant to the best advantage at all times.
Flow of operation and forward planning	Frequently stopped operating or needed to discuss next move.		Demonstrated ability for forward planning with steady progression of operative procedure.		Obviously planned course of operation with effortless flow from one move to the next.
Knowledge of specific procedure	Deficient knowledge. Needed specific instruction at most operative steps.		Knew all important aspects of the operation.		Demonstrated familiarity with all aspects of the operation.

Figure 2 : Exemple de liste de vérification (23)

ITEM	Not Done or Incorrect	Done Correctly
1. Performs a DRE	0	1
2. Uses local anesthetic	0	1
3. Uses anal retractor appropriately	0	1
4. Identifies fibrosis in posterior midline	0	1
5. Identifies external openings	0	1
6. Appropriate identification of internal opening	0	1
7. Tracks delineated with a fistula probe taking care to avoid creating a false passage	0	1
8. Identifies two tracts (leading to one internal opening)	0	1
9. Places seton drain through both tracts to internal opening appropriately	0	1
10. DOES not perform a fistulotomy	0	1
11. Ties seton in place appropriately	0	1
MAXIMUM TOTAL SCORE		(11)
TOTAL SCORE		<input type="text"/>

Figure 3 : Exemple d'échelle procédurale spécifique (52)

Table 1 Evaluation tool to assess technical skill in laparoscopic sigmoid colectomy

Access and port insertion					
Specific skill	1	2	3	4	5
Creating access and placing three or four additional ports under direct vision	Created clumsily and with difficulty		Created adequately		Created quickly and skillfully
Exposure					
Specific skill	1	2	3	4	5
Placing the patient in steep Trendelenburg	Not done		Done after delay and insufficiently		Done promptly allowing adequate exposure
Retracting the small bowel in the upper abdomen	Insufficient retraction of the small bowel		Bowl retracted adequately with some difficulty		Expert retraction of the small bowel allowing good exposure of the pelvis
Retracting the sigmoid colon from the pelvis in a cephalad and anterior direction	Difficulty achieving adequate retraction		Bowl retracted adequately with some difficulty		Expert retraction of the bowel
Identifying the left ureter	Inadequate identification of ureter		Identified with some difficulty		Ureter clearly identified
Identifying the origin of the inferior mesenteric artery (IMA)	Inadequate identification of IMA		Identified		Origin of IMA clearly identified with excellent exposure
Initial dissection					
Specific skill	1	2	3	4	5
Creating an incision in the rectorectal fascia, behind the IMA	Incision made in incorrect plane or extensive guidance required for incision. Tissue trauma and bleeding		Incision made tentatively, although in correct plane, some tissue trauma and bleeding		Incision made smoothly in correct plane
Identifying the presacral nerves	Inadequate identification		Identified		Clear identification
Developing a plane posterior to the IMA	Development of incorrect plane, tissue trauma, and bleeding		Plane developed adequately with some difficulty		Plane developed smoothly with minimal bleeding
Continuing the dissection in a plane anterior to the retroperitoneal fascia	Development of incorrect plane, tissue trauma, and bleeding		Plane developed adequately with some difficulty		Plane developed smoothly with minimal bleeding
Taking down the lateral attachments of the sigmoid and descending colon to the abdominal wall	Poor dissection technique. Bleeding		Lateral attachments taken down with minimal bleeding		Lateral attachments taken down in an expert manner with excellent hemostasis
Continuing the dissection inferiorly until the rectum is adequately mobilized	Poor technique, tissue trauma, and bleeding		Satisfactory mobilization with some difficulty		Excellent complete mobilization of rectum
Mobilizing the splenic flexure					
Specific skill	1	2	3	4	5
Dividing the splenicocolic ligament, renocolic ligament, and omental attachments of the splenic flexure	Poor technique, tissue trauma, and bleeding		Adequate mobilization of splenic flexure with some bleeding		Flexure dissected expertly with excellent mobilization
Checking the length of the mobilized specimen to ensure it will be of adequate length without tension	Length of specimen not checked		Specimen length checked briefly		Methodical checking of specimen length

Table 1 continued

Bowel resection					
Specific skill	1	2	3	4	5
Dividing the mesorectum at a right angle to the bowel	Divided with incorrect angle		Divided with an approximate right angle		Mesorectum divided at a right angle
Creating a minilaparotomy in the left lower quadrant or Pfannenstiel incision	Laparotomy created clumsily with inexperienced technique		Laparotomy created satisfactorily		Laparotomy created quickly in correct position
Utilizing a wound protector when the end of the proximal colon is delivered through the incision	Wound protector not utilized				Wound protector utilized
Anastomosis					
Specific skill	1	2	3	4	5
Checking the orientation of the bowel	Orientation not checked		Orientation checked briefly		Orientation checked in an expert fashion

Figure 4 : Étapes d'une microneurorrhaphie digitale

General hand surgery

Loop magnification

Incision planification (following Bruner principles)

Flap planification (base wide enough not more than 2:1)

Incision (number of tries to get to fat)

Tissue handling (respect vital structures)

Plane of dissection (away from zone of injury and above tendon sheath)

Exposure (flap traction aid)

Pedicle identification (proximal and distal to injury)

Pedicle isolation

Zone of injury examination

Microsurgery setup

Interocular distance

Comfortable chair height

Maximal zoom then adjust focus

Focus

Head straight

Elbows supported

Back straight

Shoulders relaxed

Dissecting zone of injury

Microinstruments resting on hands

Re-examine zone of injury

Separate artery from nerve from out of zone of injury towards zone of injury

Obliterate artery
Resect zone of injury
Check for healthy fascicles
Check amount of tension
Choice of suture (9-0 nylon)
Needle handling with needle holder under the microscope
First suture at 0 or 180°
Nerve handling by epineurium
Needle through epineurium
Pick up other side by epineurium
Align the fascicles
Needle through epineurium on other side
Surgical knot
Flat knots 3 times (no tension when finger in full extension)
Second stitch at 180 or 0°
Minimum of 3-4 sutures

Anastomosis technique

Sutures only through epineurium
No sutures through fascicles
No bunching of fascicles
Fascicles inside the nerve

Skin closure

Choice of suture (3-0, 4-0, 5-0)
Corner stitch first
No inversion of skin edges

No skin strangulation

Figure 5 : Outil d'évaluation

Resident's ID : _____ Evaluator's ID : _____

Digital Nerve Repair Practical Exam Evaluation Sheet

- Please evaluate the video-recorded performance by placing a mark on the line for each item using the visual 5-point Likert scales provided (0=step not performed; 1=performed poorly; 5=performed expertly). The criteria are presented such that each one should be scored sequentially as the procedure is carried out. Some general criteria will be presented at the end. Please do not stop the video if possible. Do not rewind the video.

- A score of 3 is the **minimum passing score** for each item (3=satisfactory). If for any item you assign a score of less than 3, please specify your reason(s) for doing so in the comment section next to the item, so that it can be communicated to the resident for educational purposes.

- Particularly important steps are noted in **BOLD** and with an asterisk *.

- To pass this exam, ALL of the **BOLD*** criteria must achieve a minimum score of 3. All of the remaining 8 criteria must achieve a minimum score of 2, and no more than 4 of these 8 can be less than 3.

Section I : Surgical approach

1)* **Incision planned respecting principles of hand surgery ***

0 1 2 3 4 5
Not done Poor Needs Improvement Satisfactory (PASS) Good Expert

Comments :

2) **Incision carried out appropriately and confidently**

0 1 2 3 4 5
Not done Poor Needs Improvement Satisfactory (PASS) Good Expert

Comments :

Resident's ID : _____ Evaluator's ID : _____

3)* **Flaps lifted and retracted adequately ***

0 1 2 3 4 5
Not done Poor Needs Improvement Satisfactory Good Expert

(PASS)

Comments :

4) **Pedicles identified appropriately and with ease**

0 1 2 3 4 5
Not done Poor Needs Improvement Satisfactory (PASS) Good Expert

Comments :

Section II : General Microsurgery

5) **Transitions to microscope dissection at appropriate time**

0 1 2 3 4 5
Not done Poor Needs Improvement Satisfactory (PASS) Good Expert

Comments :

6) **Properly adjusts microscope placement and settings to carry out microsurgical procedure without need for constant fiddling**

0 1 2 3 4 5
Not done Poor Needs Improvement Satisfactory (PASS) Good Expert

Comments :

Resident's ID : _____ Evaluator's ID : _____

Section III : Microneurraphy

7)* **Able to appropriately microdissect both cut ends of digital nerve in preparation for neurorraphy ***

0 1 2 3 4 5
Not done Poor Needs Improvement Satisfactory Good Expert
(PASS)

Comments :

8)* **Carries out appropriate "neuroma" resection (refreshes nerve ends) and checks for healthy fascicles on both nerve ends before attempting neurorrhaphy***

0 1 2 3 4 5
Not done Poor Needs Improvement Satisfactory Good Expert
(PASS)

Comments :

9)* **Verifies that there is not excessive tension before attempting neurorraphy ***

0 1 2 3 4 5
Not done Poor Needs Improvement Satisfactory Good Expert
(PASS)

Comments :

10)* **Handles nerve ends appropriately by the epineurium and without crushing fascicles ***

0 1 2 3 4 5
Not done Poor Needs Improvement Satisfactory Good Expert
(PASS)

Comments :

Resident's ID : _____ Evaluator's ID : _____

11)* Passes the suture needle through the epineurium only *

0 1 2 3 4 5
Not done Poor Needs Improvement Satisfactory Good Expert
(PASS)

Comments :

12)* Ties knot adequately and securely *

0 1 2 3 4 5
Not done Poor Needs Improvement Satisfactory Good Expert
(PASS)

Comments :

13)* Places adequate number of sutures to gently appose the fascicles without bunching or extrusion of fascicles *

0 1 2 3 4 5
Not done Poor Needs Improvement Satisfactory Good Expert
(PASS)

Comments :

Section IV : General

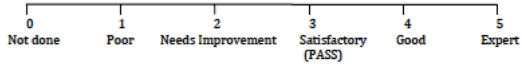
14) Carries out appropriate and adequate skin closure

0 1 2 3 4 5
Not done Poor Needs Improvement Satisfactory Good Expert
(PASS)

Comments :

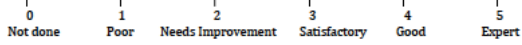
Resident's ID : _____ Evaluator's ID : _____

15) Handles surgical instruments adequately (including microsurgical instruments)



Comments :

16) Handles tissues gently and appropriately



Comments : (PASS)

17) **Automatic fail** : If you feel that the surgeon has carried out any step of the procedure in such a way that patient outcome would be **significantly adversely affected**, please describe the incorrect action here. This results in **FAILURE** of the practical examination, regardless of the rest of the evaluation.

[Empty box for automatic fail description]

Figure 6 : Installation pour tester l'outil d'évaluation

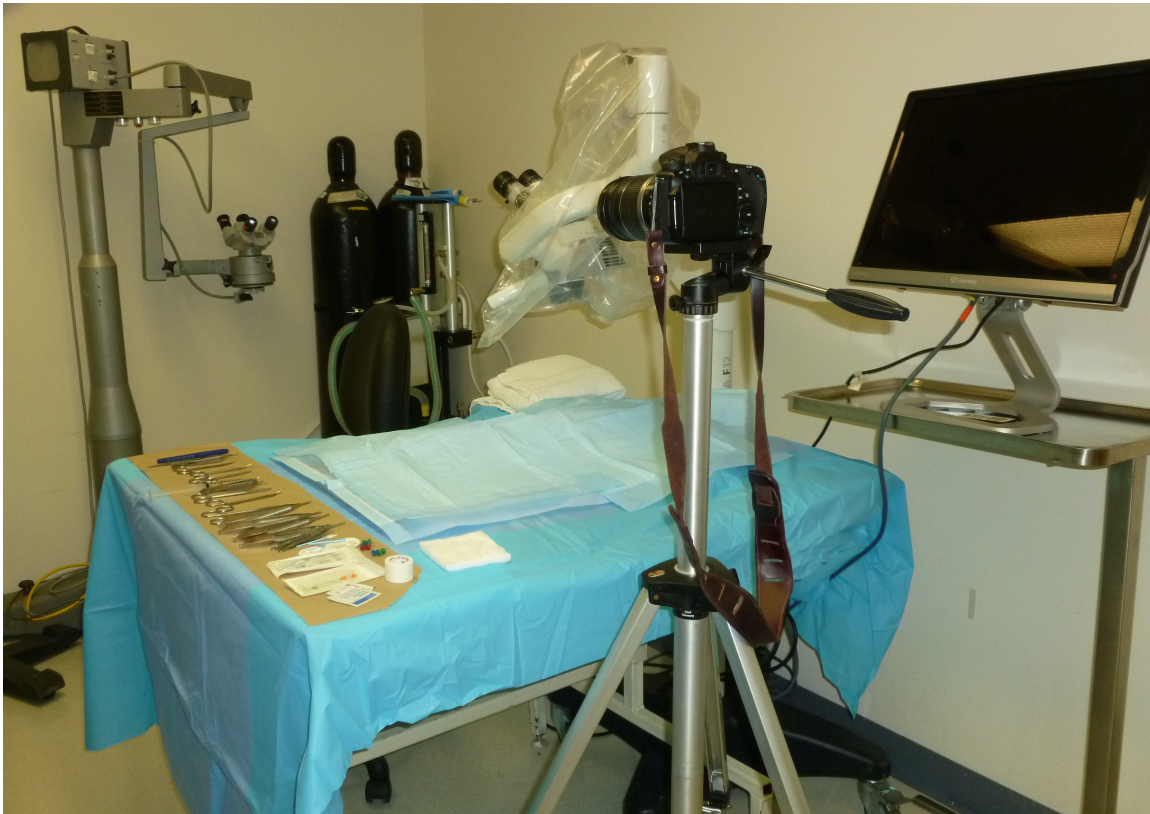


Figure 7 : Microscope chirurgical Leica M525 F50



Figure 8 : Image tirée d'une vidéo finale démontrant l'intégration des images des deux caméras



Figure 9 : Formulaire de consentement- Version résident

Formulaire d'information et de consentement- Résident

Titre du projet de recherche :

Nouvelle méthode d'évaluation des techniques de réparation de nerf digital chez le résident de
plastie

Chercheur responsable du projet de recherche :

Jenny C. Lin MD, PhD, FRCSC

Chirurgienne plastique

CHU Ste-Justine et Hôpital Maisonneuve-Rosemont

Co-chercheur:

Saoussen Salhi MD

Résidente en chirurgie plastique

Préambule:

Nous sollicitons votre participation à un projet de recherche en pédagogie médicale.

Cependant, avant d'accepter de participer à ce projet et de signer ce formulaire d'information

et de consentement, veuillez prendre le temps de lire, de comprendre et de considérer attentivement les renseignements qui suivent.

Nous vous invitons à poser toutes les questions que vous jugerez utiles au chercheur responsable du projet ou autres membres du personnel affecté au projet de recherche et à leur demander de vous expliquer tout mot ou renseignement qui n'est pas clair.

Nature et objectifs du projet de recherche:

Ce projet de recherche a pour but de développer des critères pour évaluer le niveau des résidents de chirurgie plastique dans le domaine de la réparation des nerfs digitaux. Une fois développés, ces critères seront évalués pour leur validité et leur fiabilité.

Déroulement du projet de recherche:

Cette étude s'adresse à tous les résidents de plastie de l'Université de Montréal, de tous niveaux qui acceptent d'y participer, ainsi qu'à un groupe de patrons consentants surspécialisés en microchirurgie.

Chaque résident qui accepte de prendre part dans cette étude aura à effectuer une réparation de nerf digital sur un modèle cadavérique. La procédure devrait durer entre 30 et 45 minutes en fonction du niveau d'expertise du résident. *Les seuls renseignements personnels utilisés seront votre niveau de résidence et le nombre de réparation de nerf digital que vous avez effectué dans le passé.* Votre performance sera chronométrée et filmée. *Seules vos mains apparaîtront*

sur l'enregistrement vidéo. De plus, toute bande sonore sera retirée afin de préserver votre anonymat. Cette vidéo sera par la suite visualisée par 7 examinateurs qui, séparément, évalueront le résident selon les critères développés. Une fois que vous aurez complété la procédure, un court sondage vous sera présenté afin de recueillir vos commentaires sur l'exercice. Vos réponses demeureront également confidentielles.

Collaboration du participant au projet de recherche:

Les conditions d'évaluation se feront le plus proche de celles d'une salle opératoire que possible. Une fois l'évaluation débutée, le résident devra à son tour faire comme si il se trouvait dans une salle d'opération en train d'opérer un patient. Votre participation à ce projet de recherche ne sera en aucun cas utilisée pour l'évaluation d'un de vos stages de résidence.

Inconvénients associés au projet de recherche :

Le déplacement et le temps nécessaires pour compléter la réparation nerveuse sont les inconvénients associés à ce projet de recherche.

Avantages :

Vous ne retirerez aucun bénéfice personnel de votre participation à ce projet de recherche.

Par ailleurs, les résultats obtenus de ce projet de recherche contribueront à améliorer l'enseignement ainsi que l'évaluation des résidents de chirurgie plastique dans le domaine des réparations nerveuses dans le cadre des réimplantations digitales.

Participation volontaire et possibilité de retrait :

Votre participation à ce projet de recherche est volontaire. Vous êtes donc libre de refuser d'y participer. Vous pouvez aussi vous retirer de ce projet à tout moment, sans avoir à donner de raisons, en faisant connaître votre décision au chercheur responsable du projet ou à l'un des membres du personnel affecté au projet.

Votre décision de ne pas participer à ce projet de recherche ou de vous en retirer n'aura aucune conséquence sur votre relation avec le chercheur responsable du projet.

Droits légaux :

En acceptant de participer à ce projet, vous ne renoncez à aucun de vos droits ni ne libérez les chercheurs ou l'établissement de leur responsabilité civile et professionnelle.

Confidentialité :

Durant votre participation à ce projet, le chercheur consignera dans un dossier de recherche seules les données pour répondre aux objectifs scientifiques de ce projet, décrits dans le formulaire d'information et de consentement. Tous les renseignements recueillis seront codés et demeureront strictement confidentiels dans les limites prévues par la loi. Ces données seront conservées pendant 7 ans par le chercheur responsable. Les données pourront être publiées dans des revues spécialisées ou faire l'objet de discussions scientifiques, mais il ne sera pas possible de vous identifier.

En cas de plainte, vous pouvez vous adresser au chercheur principal, Dr Lin, ou au co-chercheur, Dr Saoussen Salhi ou encore au bureau de l'Ombudsman de l'Université de Montréal dont voici les coordonnées :

Téléphone : 514-343-2100

Adresse courriel : ombudsman@umontreal.ca

Adresse civique : 3744, rue Jean Brillant, bureau 430

Adresse postale : Pascale Descary

Ombudsman

Université de Montréal

C.P. 6128, succursale Centre-ville

Montréal (QC), H3C 3J7

Financement du projet de recherche :

Le financement du projet de recherche se fera par des fonds internes au programme de Chirurgie Plastique *qui proviennent du CHU Ste-Justine.*

Identification des personnes-ressources :

Si vous avez des questions concernant le projet de recherche, vous pouvez communiquer avec le co-chercheur, Dr Saoussen Salhi, au numéro suivant : 514-582-8613 ou à l'adresse courriel ci-dessous : saoussen.salhi@mail.mcgill.ca

Surveillance des aspects éthiques du projet de recherche :

Le comité d'éthique de la recherche de l'HMR a approuvé ce projet de recherche et en assure le suivi. De plus, il approuvera au préalable toute révision et toute modification apportée au protocole et au formulaire d'information et de consentement. Vous pouvez le joindre au 514-252-3400 poste 5708.

Consentement :

J'ai pris connaissance du formulaire d'information et de consentement. Je reconnais qu'on m'a expliqué le projet, qu'on a répondu à mes questions, et qu'on m'a laissé le temps voulu pour prendre une décision.

Je consens à participer à ce projet de recherche aux conditions qui y sont énoncées. Une copie signée et datée de ce présent formulaire d'information et de consentement m'a été remise.

Nom et signature du *participant* de recherche

Date

Signature et engagement du chercheur responsable du projet :

Je certifie qu'on a expliqué au *participant* de recherche les termes du présent formulaire d'information et de consentement, que l'on a répondu aux questions que le *participant* de recherche avait à cet égard et qu'on lui a clairement indiqué qu'il demeure libre de mettre un terme à sa participation, et ce, sans préjudice.

Je m'engage, avec l'équipe de recherche, de respecter ce qui a été convenu au formulaire d'information et de consentement et à en remettre une copie signée au *participant* de recherche.

Nom et signature du chercheur responsable du projet de recherche

Date

Figure 10: Outil d'évaluation- Version finale

Resident's ID:

Evaluator's ID:

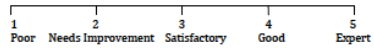
Digital nerve repair practical exam evaluation sheet

Instructions:

- Please evaluate the video-recorded performance by placing a mark on the 5-point visual Likert scale provided for each step of the surgery.
- 3 required elements are listed below each step
- Score of 1 (poor) is equal to none of the 3 steps performed properly. Range is 1 – 2 based on how close the candidate gets to actually fulfilling one of the steps.
- Score of 2 – 3 is used if 1/3 are achieved
- Score of 3 – 4 is used if 2/3 steps are achieved
- Score of 4 – 5 is used if 3/3 steps are achieved
- Evaluator's judgement is used to choose on the Likert scale where the candidate falls within the range given, based on the ease with which the requirements are met and your assessment of the surgical skill demonstrated by the candidate.

1. Incision planning

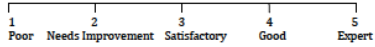
- Incision performed according to Brunner lines
- Adequate zone of injury exposure
- Flaps lifted have adequate vascularization



Comments:

2. Carrying out incision

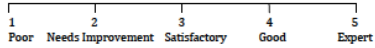
- Incision made without requiring multiple passes of scalpel
- Incision to subcutaneous fat before starting dissection of flap
- No damage to important structures



Comments:

3. Flaps lifting and retraction

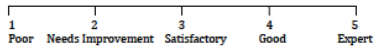
- Flaps lifted in flexor tendon sheath plane
- Gentle handling of flap without crushing
- Adequate flap retraction allowing good exposure of pedicles



Comments:

4. Pedicle identification

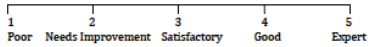
- Dissection performed in plane of flexor tendon sheath
- Dissection started away from zone of injury and progressing toward zone of injury
- Dissection performed with scissors, longitudinal to pedicle



Comments:

5. Transition to microscope

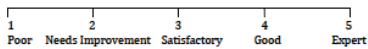
- Transitions to microscope when has achieved
- Pedicle dissection throughout zone of injury
- Adequate exposure
- Adequate flap retraction



Comments:

6. Microscope adjustment

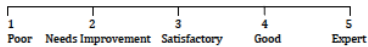
- Does not fiddle with microscope placement once it has been adjusted before beginning the microsurgery
- Surgical field of interest centered in microscope field
- Microscope was placed so that surgeon is in comfortable positioning to carry out the microsurgery procedure



Comments:

7. Digital nerve dissection

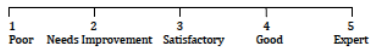
- Isolates digital nerve from artery on a sufficient distance to safely and accurately carry out neurography
- Refreshes nerve ends before nerve coaptation
- Verifies that there is no tension on repair



Comments:

8. Suture placement

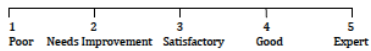
- Handles nerve ends by epineurium only
- Suture placed through epineurium only
- Does not tear epineurium upon needle passing or suture tying



Comments:

9. Suture tying

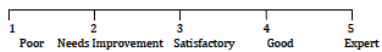
- Places adequate number of sutures
- Knots are square
- No bunching or extrusion of fascicles upon suture tying



Comments:

10. Skin suture tying

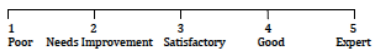
- Knots are square
- Equal bites are taken on either side of the incision
- Needle is passed at an appropriate distance from the skin edge



Comments:

11. Skin closure

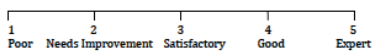
- No skin inversion
- No skin strangulation with too much eversion
- Adequate number of sutures



Comments:

12. Surgical instrument handling

- Handles instruments gently
- Handles instruments with ease
- Holds instruments appropriately



Comments:

Figure 11: Variation de la performance des participants selon le niveau d'expertise

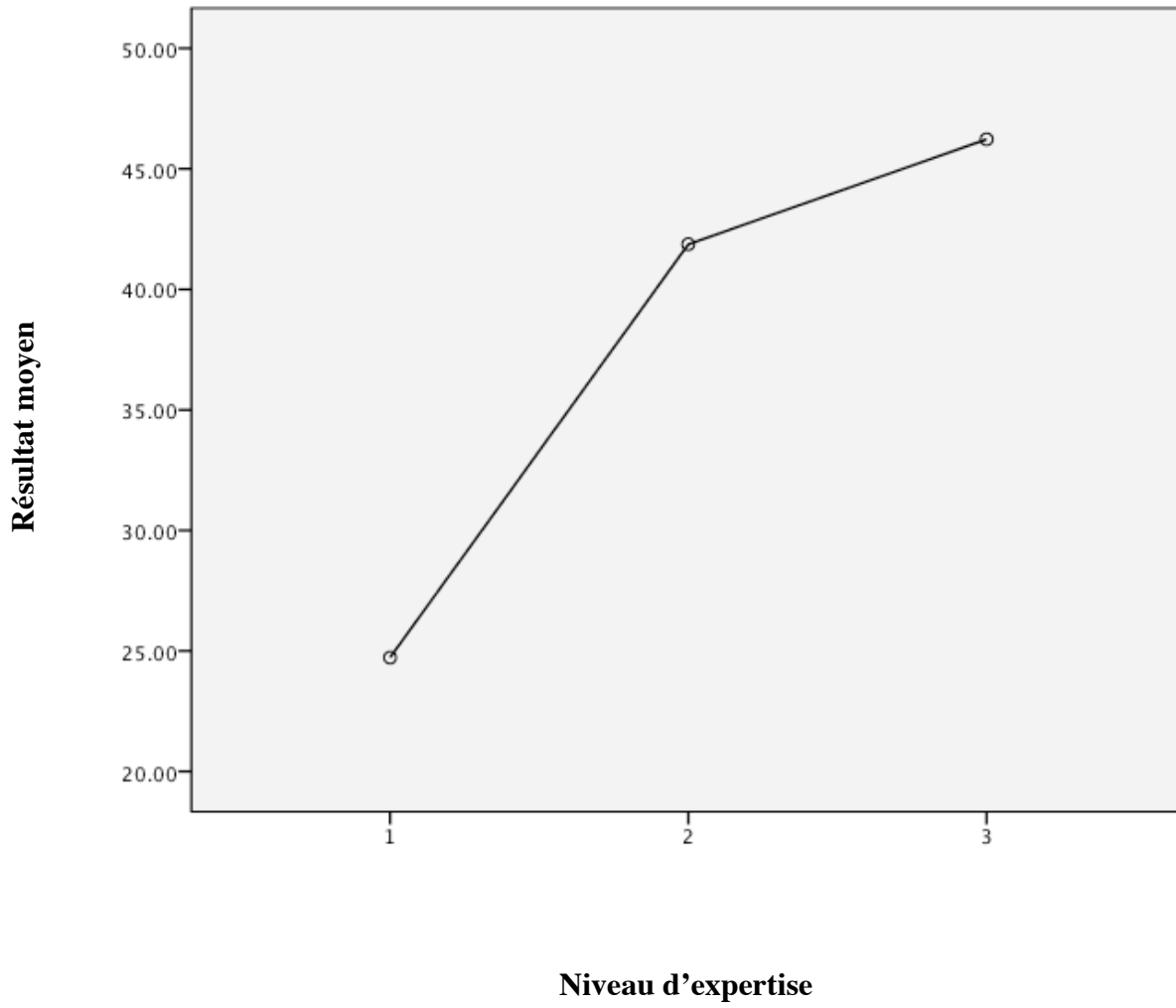
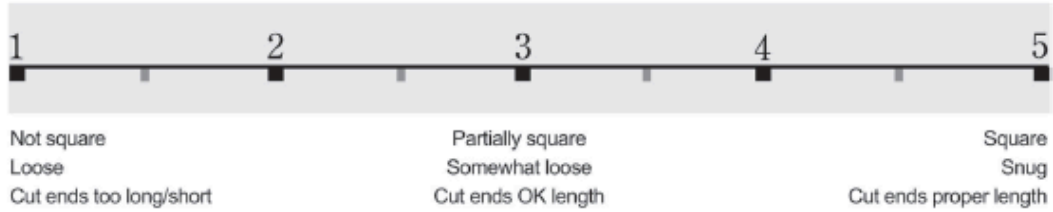


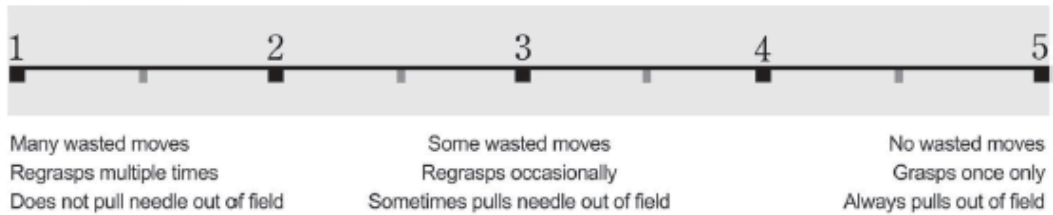
Figure 12 : University of Western Ontario Microsurgical Skills Acquisition instrument :
 UWOMSA (77)

KNOT TYING MODULE

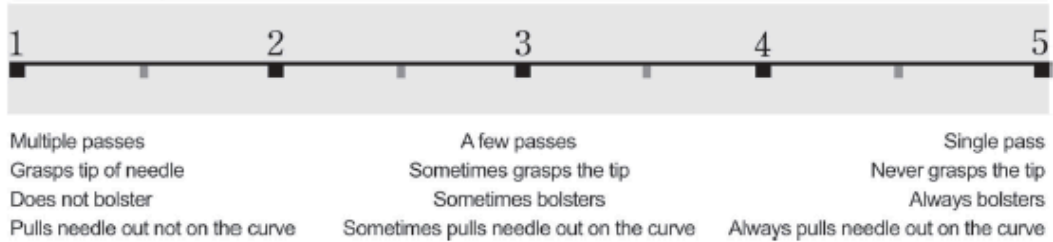
A. QUALITY OF KNOT



B. EFFICIENCY



C. HANDLING



ANASTOMOSIS MODULE

A. PREPARATION

1	2	3	4	5
Forgets background				Background in place
Ends set up poorly in approximating clamp			Approximating clamp applied correctly	
Forgets dilatation		Rough dilatation		Gentle dilatation
No adventitial stripping	Excessive/inadequate adventitial stripping			Clean adventitial stripping

B. SUTURING

1	2	3	4	5
Inaccurate needle placement		Inconsistent needle placement		Accurate needle placement
Pulls needle through roughly		Rough/inconsistent needle passage		Takes needle out on curve
Drops suture end/inefficient knot tying		Knot tying loose/tight/inefficient		Efficient tying
Does not look inside lumen				Always checks inside lumen
Too much movement at anastomosis with tying				Anastomosis stays still with tying

C. FINAL PRODUCT

1	2	3	4	5
Rough outer appearance	Outer appearance inconsistent/partially inverted			Smooth outer appearance
Back wall stitch	Possible back wall stitch			No back wall stitch
Not patent				Patent
Suture ends intraluminal	Some suture ends intraluminal			All suture ends extraluminal
Poor suture spacing	Suture spacing inconsistent			Appropriate suture spacing

Figure 13 : Structured Assessment of Microsurgical Skills (SAMS) (12)

Please tick level (scale 1-5)		1	2	3	4	5
Dexterity	Steadiness	Frequent exhibition of tremor		Occasionally signs of tremor		Fluid movements and obvious control of movements
	Instrument handling	Repeatedly tentative awkward moves through inappropriate use		Competent use, but occasionally stiff/awkward		Fluid movements with no stiffness or awkwardness
	Tissue handling	Frequently unnecessary force on tissue/ damage caused		Careful, but occasionally inadvertent damage		Consistently appropriate handling with minimal damage
Visuo-spatial ability	Dissection	Frequently uncontrolled grasping of tissue or readjustment of view required		Correct use of instruments at right angle, but occasional uncontrolled grasping/ view disturbance		Efficient use of instruments with vessel ends clearly prepared
	Suture placement	Frequently lost sutures & uneven placement		Occasionally uneven sutures placement		Consistent delicately placed sutures with adequate spaces
	Knot technique	Insecure knots/ frequent thread in knots		Secure knots, but occasionally awkward movement		Consistent secure knots placed effortlessly
Operative flow	Steps	Frequently stopped and unsure of next move		Reasonable progression of procedure		Obvious logically planned course of operation with effortless flow from one move to next
	Motion	Many unnecessary or repetitive moves		Efficient, but some unnecessary moves		Economy of movement and maximum efficiency
	Speed	Excessive time used for each step due to insufficient dexterity		Efficient time, but some unnecessary/ repetitive moves		Excellent speed and superior dexterity without awkward moves
Judgement	Irrigation	Poor use of irrigation, frequently desiccated vessels		Fair use of irrigation, but occasionally too wet interfering operative flow		Appropriate use of irrigation
	Patency test	Unsure of result of Acland test with possible thrombosed patency		Correct clamp removal, but repeatedly using Acland test		Appropriate sequence and time of patency test with confidence of unimpeded flow
	Bleeding control	No check for uncompleted anastomosis before clamp removal		Anastomosis checked but requires sutures for anastomotic leak		Places extra sutures as appropriate before clamp removal

Figure 14 : Liste d'erreurs (12)

Errors list: [please tick if observed]

Planning	Inappropriate operative field	
	Inappropriate vessel set up	
	Focus lost	
	Loss of central view	
Dexterity	Vessel clamp re-application	
	Broken sutures/needle off suture	
	Wrong grasp / damage tissue	
	Opposite wall caught	
	Vessel tear	
	Thread in knot	
Visuospatial ability	Insufficient vessel preparation	
	Empty grasp	
	Unequal stitch bites	
	Suture pulled through	
	Suture cut through	
	Loose knot	
Operative flow	Re-do suture	
	Inappropriate magnification	
	Inadequate loop length	
Judgement	Vessel dessication	
	'pooling'	
	Anastomotic leak - extra sutures	
	'crushing' patency test	
	Excessive sutures	
	Impeded flow - re-do anastomosis	

Tableau I : Participants

Participant	Niveau d'expertise	Évaluateur # 1	Évaluateur # 2	Groupe⁺
1	Expert	48.85	35.1	3
2	PGY-3	40.7	35.75	2
3	PGY-5	34.73	25.5	2
4	PGY-4	34.43	39.35	2
5	Expert	44.18	60	3
6	PGY-1	27.7	29.45	1
7	PGY-4	52.25	43.8	2
8*	Expert	45.65	17	3
9	PGY-5	47.25	24.65	2
10	PGY-2	21.75	14.7	1

+Groupe de participants selon leur niveau d'expertise. Le groupe 1 représente les résidents juniors, le groupe 2 celui des résidents séniors et le groupe 3 celui des experts.

* L'évaluation du participant # 8 par l'évaluateur # 2 étant incomplète, celui-ci a été exclus du calcul de la fiabilité inter-évaluateur.

Tableau II : Variance de la cohérence interne de l’outil d’évaluation selon la déletion de chacune de ses étapes

Étape	Cronbach alpha si étape omise
Étape 1	0.955
Étape 2	0.957
Étape 3	0.954
Étape 4	0.955
Étape 5	0.956
Étape 6	0.961
Étape 7	0.953
Étape 8	0.957
Étape 9	0.956
Étape 10	0.957
Étape 11	0.960
Étape 12	0.955

Tableau III : Analyse post-hoc des différences de scores entre les groupes de participants

(I) Niveau	(J) Niveau	Différence moyenne (I-J)	Erreur standard	Signification statistique
1	2	-17.14700*	5.22968	0.032
	3	-21.50167*	5.70605	0.017
2	1	17.14700*	5.22968	0.032
	3	-4.35467	4.56484	0.626
3	1	21.50167*	5.70605	0.017
	2	4.35467	4.56484	0.626

Bibliographie

1. Ahmed K, Jawad M, Dasgupta P, Darzi A, Athanasiou T, Khan MS. Assessment and maintenance of competence in urology. *Nat Rev Urol.* 2010;7(7):403-13.
2. Beard JD, Jolly BC, Newble DI, Thomas WE, Donnelly J, Southgate LJ. Assessing the technical skills of surgical trainees. *Br J Surg.* 2005;92(6):778-82.
3. Goderstad JM, Sandvik L, Fosse E, Lieng M. Assessment of Surgical Competence: Development and Validation of Rating Scales Used for Laparoscopic Supracervical Hysterectomy. *J Surg Educ.* 2016;73(4):600-8.
4. Ahmed K, Miskovic D, Darzi A, Athanasiou T, Hanna GB. Observational tools for assessment of procedural skills: a systematic review. *Am J Surg.* 2011;202(4):469-80 e6.
5. Beard JD. Assessment of surgical skills of trainees in the UK. *Ann R Coll Surg Engl.* 2008;90(4):282-5.
6. Ahmed K, Khan RS, Darzi A, Athanasiou T, Hanna GB. Recertification: what do specialists think about skill assessment? *Surgeon.* 2013;11(3):120-4.
7. Mickelson JJ, Macneily AE. Translational education: tools for implementing the CanMEDS competencies in Canadian urology residency training. *Can Urol Assoc J.* 2008;2(4):395-404.
8. Goff BA, Lentz GM, Lee D, Fenner D, Morris J, Mandel LS. Development of a bench station objective structured assessment of technical skills. *Obstet Gynecol.* 2001;98(3):412-6.
9. Hopmans CJ, den Hoed PT, van der Laan L, van der Harst E, van der Elst M, Mannaerts GH, et al. Assessment of surgery residents' operative skills in the operating

theater using a modified Objective Structured Assessment of Technical Skills (OSATS): a prospective multicenter study. *Surgery*. 2014;156(5):1078-88.

10. Aggarwal R, Grantcharov T, Moorthy K, Milland T, Papasavas P, Dosis A, et al. An evaluation of the feasibility, validity, and reliability of laparoscopic skills assessment in the operating room. *Ann Surg*. 2007;245(6):992-9.

11. Ahmed A, Ishman SL, Laeeq K, Bhatti NI. Assessment of improvement of trainee surgical skills in the operating room for tonsillectomy. *Laryngoscope*. 2013;123(7):1639-44.

12. Chan W, Niranjana N, Ramakrishnan V. Structured assessment of microsurgery skills in the clinical setting. *J Plast Reconstr Aesthet Surg*. 2010;63(8):1329-34.

13. Laeeq K, Bhatti NI, Carey JP, Della Santina CC, Limb CJ, Niparko JK, et al. Pilot testing of an assessment tool for competency in mastoidectomy. *Laryngoscope*. 2009;119(12):2402-10.

14. Balayla J, Abenhaim HA, Martin MC. Does residency training improve cognitive competence in obstetric and gynaecologic surgery? *J Obstet Gynaecol Can*. 2012;34(2):190-6.

15. Larson JL, Williams RG, Ketchum J, Boehler ML, Dunnington GL. Feasibility, reliability and validity of an operative performance rating system for evaluating surgery residents. *Surgery*. 2005;138(4):640-7; discussion 7-9.

16. Davoudi M, Colt HG, Osann KE, Lamb CR, Mullon JJ. Endobronchial ultrasound skills and tasks assessment tool: assessing the validity evidence for a test of endobronchial ultrasound-guided transbronchial needle aspiration operator skill. *Am J Respir Crit Care Med*. 2012;186(8):773-9.

17. Santos EG, Salles GF. Construction and validation of a surgical skills assessment tool for general surgery residency program. *Rev Col Bras Cir.* 2015;42(6):407-12.
18. Marglani O, Alherabi A, Al-Andejani T, Javer A, Al-Zalabani A, Chalmers A. Development of a tool for Global Rating of Endoscopic Surgical Skills (GRESS) for assessment of otolaryngology residents. *B-ENT.* 2012;8(3):191-5.
19. Balasundaram I, Aggarwal R, Darzi LA. Development of a training curriculum for microsurgery. *Br J Oral Maxillofac Surg.* 2010;48(8):598-606.
20. Acosta S, Sevonius D, Beckman A, Swedish steering committee of the Basic Surgical Skills course SSS. Extended score interval in the assessment of basic surgical skills. *Med Educ Online.* 2015;20:25819.
21. Golnik KC, Gauba V, Saleh GM, Collin R, Naik MN, Devoto M, et al. The ophthalmology surgical competency assessment rubric for lateral tarsal strip surgery. *Ophthal Plast Reconstr Surg.* 2012;28(5):350-4.
22. Laeeq K, Waseem R, Weatherly RA, Reh DD, Lin SY, Lane AP, et al. In-training assessment and predictors of competency in endoscopic sinus surgery. *Laryngoscope.* 2010;120(12):2540-5.
23. de Montbrun SL, Roberts PL, Lowry AC, Ault GT, Burnstein MJ, Cataldo PA, et al. A novel approach to assessing technical competence of colorectal surgery residents: the development and evaluation of the Colorectal Objective Structured Assessment of Technical Skill (COSATS). *Ann Surg.* 2013;258(6):1001-6.
24. Fisher JB, Binenbaum G, Tapino P, Volpe NJ. Development and face and content validity of an eye surgical skills assessment test for ophthalmology residents. *Ophthalmology.* 2006;113(12):2364-70.

25. Duran CA, Shames M, Bismuth J, Lee JT, Simulation Committee of the Association of Program Directors in Vascular S. Validated assessment tool paves the way for standardized evaluation of trainees on anastomotic models. *Ann Vasc Surg.* 2014;28(1):115-21.
26. Gofton WT, Dudek NL, Wood TJ, Balaa F, Hamstra SJ. The Ottawa Surgical Competency Operating Room Evaluation (O-SCORE): a tool to assess surgical competence. *Acad Med.* 2012;87(10):1401-7.
27. Doyle JD, Webber EM, Sidhu RS. A universal global rating scale for the evaluation of technical skills in the operating room. *Am J Surg.* 2007;193(5):551-5; discussion 5.
28. Berger AJ, Gaster RS, Lee GK. Development of an affordable system for personalized video-documented surgical skill analysis for surgical residency training. *Ann Plast Surg.* 2013;70(4):442-6.
29. Geoffrion R, Choi JW, Lentz GM. Training surgical residents: the current Canadian perspective. *J Surg Educ.* 2011;68(6):547-59.
30. Jantz MA, McGaghie WC. It's time for a STAT assessment of bronchoscopy skills: the endobronchial ultrasound bronchoscopy (EBUS)-STAT and EBUS-transbronchial needle aspiration skill evaluation. *Am J Respir Crit Care Med.* 2012;186(8):703-5.
31. Fernandez SA, Wiet GJ, Butler NN, Welling B, Jarjoura D. Reliability of surgical skills scores in otolaryngology residents: analysis using generalizability theory. *Eval Health Prof.* 2008;31(4):419-36.
32. Al-Ruzzeh S, Karthik S, O'Regan D. Objective surgical skill assessment: the diagonal operating matrix. *Interact Cardiovasc Thorac Surg.* 2007;6(2):188-91.

33. Hogle NJ, Liu Y, Ogden RT, Fowler DL. Evaluation of surgical fellows' laparoscopic performance using Global Operative Assessment of Laparoscopic Skills (GOALS). *Surg Endosc.* 2014;28(4):1284-90.
34. Dumestre D, Yeung JK, Temple-Oberle C. Evidence-based microsurgical skill-acquisition series part 1: validated microsurgical models--a systematic review. *J Surg Educ.* 2014;71(3):329-38.
35. Ghaderi I, Vaillancourt M, Sroka G, Kaneva PA, Vassiliou MC, Choy I, et al. Evaluation of surgical performance during laparoscopic incisional hernia repair: a multicenter study. *Surg Endosc.* 2011;25(8):2555-63.
36. Elliott MJ, Caprise PA, Henning AE, Kurtz CA, Sekiya JK. Diagnostic knee arthroscopy: a pilot study to evaluate surgical skills. *Arthroscopy.* 2012;28(2):218-24.
37. Ault G, Reznick R, MacRae H, Leadbetter W, DaRosa D, Joehl R, et al. Exporting a technical skills evaluation technology to other sites. *Am J Surg.* 2001;182(3):254-6.
38. Cremers SL, Ciolino JB, Ferrufino-Ponce ZK, Henderson BA. Objective Assessment of Skills in Intraocular Surgery (OASIS). *Ophthalmology.* 2005;112(7):1236-41.
39. Flores RL, Havlik RJ, Choi M, Heidelman JF, Bennett JD, Tholpady S. Measuring surgical competency in facial trauma: the Arch Bar Placement Assessment Scale. *Ann Plast Surg.* 2014;73(3):299-303.
40. Hiemstra E, Kolkman W, Wolterbeek R, Trimbos B, Jansen FW. Value of an objective assessment tool in the operating room. *Can J Surg.* 2011;54(2):116-22.
41. Adrales GL, Donnelly MB, Chu UB, Witzke DB, Hoskins JD, Mastrangelo MJ, Jr., et al. Determinants of competency judgments by experienced laparoscopic surgeons. *Surg Endosc.* 2004;18(2):323-7.

42. Chou B, Bowen CW, Handa VL. Evaluating the competency of gynecology residents in the operating room: validation of a new assessment tool. *Am J Obstet Gynecol.* 2008;199(5):571 e1-5.
43. Cremers SL, Lora AN, Ferrufino-Ponce ZK. Global Rating Assessment of Skills in Intraocular Surgery (GRASIS). *Ophthalmology.* 2005;112(10):1655-60.
44. Carlsen CG, Lindorff-Larsen K, Funch-Jensen P, Lund L, Charles P, Konge L. Reliable and valid assessment of Lichtenstein hernia repair skills. *Hernia.* 2014;18(4):543-8.
45. Bowles PF, Harries M, Young P, Das P, Saunders N, Fleming JC. A validation study on the use of intra-operative video recording as an objective assessment tool for core ENT surgery. *Clin Otolaryngol.* 2014;39(2):102-7.
46. Ghaderi I, Manji F, Park YS, Juul D, Ott M, Harris I, et al. Technical skills assessment toolbox: a review using the unitary framework of validity. *Ann Surg.* 2015;261(2):251-62.
47. Angelo RL, Ryu RK, Pedowitz RA, Gallagher AG. The Bankart Performance Metrics Combined With a Cadaveric Shoulder Create a Precise and Accurate Assessment Tool for Measuring Surgeon Skill. *Arthroscopy.* 2015;31(9):1655-70.
48. Alvand A, Logishetty K, Middleton R, Khan T, Jackson WF, Price AJ, et al. Validating a global rating scale to monitor individual resident learning curves during arthroscopic knee meniscal repair. *Arthroscopy.* 2013;29(5):906-12.
49. Aoun SG, El Ahmadih TY, El Teclé NE, Daou MR, Adel JG, Park CS, et al. A pilot study to assess the construct and face validity of the Northwestern Objective Microanastomosis Assessment Tool. *J Neurosurg.* 2015;123(1):103-9.

50. Gumbs AA, Hogle NJ, Fowler DL. Evaluation of resident laparoscopic performance using global operative assessment of laparoscopic skills. *J Am Coll Surg.* 2007;204(2):308-13.
51. Curry M, Malpani A, Li R, Tantilillo T, Jog A, Blanco R, et al. Objective assessment in residency-based training for transoral robotic surgery. *Laryngoscope.* 2012;122(10):2184-92.
52. Palter VN, Grantcharov TP. A prospective study demonstrating the reliability and validity of two procedure-specific evaluation tools to assess operative competence in laparoscopic colorectal surgery. *Surg Endosc.* 2012;26(9):2489-503.
53. Palter VN, MacRae HM, Grantcharov TP. Development of an objective evaluation tool to assess technical skill in laparoscopic colorectal surgery: a Delphi methodology. *Am J Surg.* 2011;201(2):251-9.
54. Konge L, Larsen KR, Clementsen P, Arendrup H, von Buchwald C, Ringsted C. Reliable and valid assessment of clinical bronchoscopy performance. *Respiration.* 2012;83(1):53-60.
55. Kramp KH, van Det MJ, Hoff C, Lamme B, Veeger NJ, Pierie JP. Validity and reliability of global operative assessment of laparoscopic skills (GOALS) in novice trainees performing a laparoscopic cholecystectomy. *J Surg Educ.* 2015;72(2):351-8.
56. Awad Z, Hayden L, Robson AK, Muthuswamy K, Tolley NS. Reliability and validity of procedure-based assessments in otolaryngology training. *Laryngoscope.* 2015;125(6):1328-35.

57. Eubanks TR, Clements RH, Pohl D, Williams N, Schaad DC, Horgan S, et al. An objective scoring system for laparoscopic cholecystectomy. *J Am Coll Surg.* 1999;189(6):566-74.
58. Sugden C, Aggarwal R. Assessment and feedback in the skills laboratory and operating room. *Surg Clin North Am.* 2010;90(3):519-33.
59. Vassiliou MC, Feldman LS, Andrew CG, Bergman S, Leffondre K, Stanbridge D, et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. *Am J Surg.* 2005;190(1):107-13.
60. Downing SM. Reliability: on the reproducibility of assessment data. *Med Educ.* 2004;38(9):1006-12.
61. Bath AP, Wilson T. Objective assessment of surgical competency--ENT trainees. *Clin Otolaryngol.* 2007;32(6):475-9.
62. Gauba V, Saleh GM, Goel S. Ophthalmic plastic surgical skills assessment tool. *Ophthal Plast Reconstr Surg.* 2008;24(1):43-6.
63. Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg.* 1997;84(2):273-8.
64. Hong JW, Kim YS, Lee WJ, Hong HJ, Roh TS, Song SY. Evaluation of the efficacy of microsurgical practice through time factor added protocol: microsurgical training using nonvital material. *J Craniofac Surg.* 2010;21(3):876-81.
65. Alrasheed T, Liu J, Hanasono MM, Butler CE, Selber JC. Robotic microsurgery: validating an assessment tool and plotting the learning curve. *Plast Reconstr Surg.* 2014;134(4):794-803.

66. Selber JC, Chang EI, Liu J, Suami H, Adelman DM, Garvey P, et al. Tracking the learning curve in microsurgical skill acquisition. *Plast Reconstr Surg.* 2012;130(4):550e-7e.
67. Cheung JJ, Chen EW, Darani R, McCartney CJ, Dubrowski A, Awad IT. The creation of an objective assessment tool for ultrasound-guided regional anesthesia using the Delphi method. *Reg Anesth Pain Med.* 2012;37(3):329-33.
68. Golnik C, Beaver H, Gauba V, Lee AG, Mayorga E, Palis G, et al. Development of a new valid, reliable, and internationally applicable assessment tool of residents' competence in ophthalmic surgery (an American Ophthalmological Society thesis). *Trans Am Ophthalmol Soc.* 2013;111:24-33.
69. Glarner CE, McDonald RJ, Smith AB, Levenson GE, Peyre S, Pugh CM, et al. Utilizing a novel tool for the comprehensive assessment of resident operative performance. *J Surg Educ.* 2013;70(6):813-20.
70. Kurashima Y, Feldman LS, Al-Sabah S, Kaneva PA, Fried GM, Vassiliou MC. A tool for training and evaluation of laparoscopic inguinal hernia repair: the Global Operative Assessment Of Laparoscopic Skills-Groin Hernia (GOALS-GH). *Am J Surg.* 2011;201(1):54-61.
71. Golnik KC, Motley WW, Atilla H, Pilling R, Reddy A, Sharma P, et al. The ophthalmology surgical competency assessment rubric for strabismus surgery. *J AAPOS.* 2012;16(4):318-21.
72. Gwet KL. *Handbook of inter-rater reliability: the definitive guide to measuring the extent of agreement among raters.* 2014:151-6.

73. Gwet KL. Handbook of inter-rater reliability: the definitive guide to measuring the extent of agreement among raters. 2014:241-6.
74. Konge L, Lehnert P, Hansen HJ, Petersen RH, Ringsted C. Reliable and valid assessment of performance in thoracoscopy. *Surg Endosc.* 2012;26(6):1624-8.
75. Smith RJ, McCannel CA, Gordon LK, Hollander DA, Giaconi JA, Stelzner SK, et al. Evaluating teaching methods of cataract surgery: validation of an evaluation tool for assessing surgical technique of capsulorhexis. *J Cataract Refract Surg.* 2012;38(5):799-806.
76. Beard J, Rowley D, Bussey M, Pitts D. Workplace-based assessment: assessing technical skill throughout the continuum of surgical training. *ANZ J Surg.* 2009;79(3):148-53.
77. Temple CL, Ross DC. A new, validated instrument to evaluate competency in microsurgery: the University of Western Ontario Microsurgical Skills Acquisition/Assessment instrument [outcomes article]. *Plast Reconstr Surg.* 2011;127(1):215-22.
78. Francis HW, Masood H, Chaudhry KN, Laeeq K, Carey JP, Della Santina CC, et al. Objective assessment of mastoidectomy skills in the operating room. *Otol Neurotol.* 2010;31(5):759-65.
79. Husslein H, Shirreff L, Shore EM, Lefebvre GG, Grantcharov TP. The Generic Error Rating Tool: A Novel Approach to Assessment of Performance and Surgical Education in Gynecologic Laparoscopy. *J Surg Educ.* 2015;72(6):1259-65.

80. Koehler RJ, Amsdell S, Arendt EA, Bisson LJ, Braman JP, Butler A, et al. The Arthroscopic Surgical Skill Evaluation Tool (ASSET). *Am J Sports Med.* 2013;41(6):1229-37.
81. Konge L, Vilmann P, Clementsen P, Annema JT, Ringsted C. Reliable and valid assessment of competence in endoscopic ultrasonography and fine-needle aspiration for mediastinal staging of non-small cell lung cancer. *Endoscopy.* 2012;44(10):928-33.
82. Herati AS, Montag S, Andonian S, Shapiro EY, Akerman M, Kavoussi LR, et al. Audio and visual analysis of urologic laparoscopic and robotic skills: objective criteria for surgical skill evaluation. *Urology.* 2012;80(6):1277-82.
83. Hadley C, Lam SK, Briceno V, Luerksen TG, Jea A. Use of a formal assessment instrument for evaluation of resident operative skills in pediatric neurosurgery. *J Neurosurg Pediatr.* 2015:1-8.
84. Goh AC, Goldfarb DW, Sander JC, Miles BJ, Dunkin BJ. Global evaluative assessment of robotic skills: validation of a clinical assessment tool to measure robotic surgical skills. *J Urol.* 2012;187(1):247-52.