

Université de Montréal

Évaluation de l'acceptabilité du test de concordance de script : Une étude nationale chez les résidents en oto-rhino-laryngologie.

par Andrée-Anne Leclerc

Programme de cycle supérieur en sciences biomédicales
Faculté de Médecine

Mémoire présenté
en vue de l'obtention du grade de maîtrise
en sciences biomédicales.

mai 2017

© Andrée-Anne Leclerc, 2017

Résumé

OBJECTIFS. Le test de concordance de script (TCS) est une méthode objective d'évaluation du raisonnement clinique, et plus particulièrement de l'habileté à interpréter l'information médicale dans un contexte d'incertitude. Plusieurs études ont prouvé sa validité pour l'évaluation de l'apprentissage. Par contre, nous en connaissons peu sur l'acceptabilité du TCS par les participants et experts. L'objectif était donc de comparer l'acceptabilité du TCS, lorsque administré à un niveau national, entre des résidents en oto-rhino-laryngologie et chirurgie cervico-faciale (ORL-CCF) et un groupe d'experts dans le domaine.

MÉTHODE. En 2013 et 2016, des questions formatives de type TCS ainsi qu'un sondage post-test ont été inclus dans l'examen national *in-training* (NITE) d'ORL-CCF. Cet examen est administré à tous les résidents en ORL-CCF du Canada à partir de la deuxième année de résidence. Les mêmes questions de type TCS et le sondage post-test ont été envoyés à un groupe d'oto-rhino-laryngologistes travaillant dans quatre universités du Canada.

RÉSULTATS. Pour 64% des résidents et médecins en pratique, notre étude était leur première exposition au test de concordance de script. La majorité des résidents a trouvé difficile de s'adapter au TCS. Ils trouvaient que les scénarios cliniques n'étaient pas clairs et ne considéraient pas le TCS comme étant utile pour évaluer le raisonnement clinique. Au contraire, la grande majorité des experts avait l'impression que les cas reflétaient des situations cliniques réalistes et aurait recommandé le TCS comme méthode d'évaluation formative en ORL-CCF.

CONCLUSION. Il y a une différence majeure entre l'opinion qu'ont les résidents en ORL-CCF sur l'évaluation du raisonnement clinique par le TCS, en comparaison avec l'opinion des experts. Pour améliorer l'acceptabilité du TCS par les résidents, un effort conscient d'éducation sera nécessaire, ainsi qu'une augmentation de l'exposition au TCS.

Mots-clés : Script, Concordance, Raisonnement, Acceptabilité, Oto-rhino-laryngologie.

Abstract

OBJECTIVES. Script concordance test (SCT) is an objective assessment method of clinical reasoning which assesses the ability to interpret medical information under conditions of uncertainty. Many studies have supported its validity as a tool to assess higher levels of learning, but little is known regarding its acceptability by major stakeholders. Thus, the objective of our study is to compare the acceptability of SCT when administered at a national level to Otolaryngology – Head and Neck Surgery (OTL-HNS) residents and a reference group of experts.

METHODS. In 2013 and 2016, SCT formative questions, as well as an exit survey, were included in the national in-training exam (NITE) of OTL-HNS. This exam is administered to all OTL-HNS residents across Canada, from second to fifth year of residency. The same SCT questions and survey were then sent to a group of OTL-HNS physicians from five canadian universities.

RESULTS. For 64% of faculty and residents, the study was their first exposure to SCTs. Overall, residents found it difficult to adapt to SCTs, thought that the clinical scenarios were not clear and believed that SCTs were not useful for assessing clinical reasoning. In contrast, the vast majority of experts felt that SCTs reflected real-life clinical situations, and would recommend SCT as an evaluation method in OTL-HNS.

CONCLUSION. There is a great difference in the acceptability of SCT as an assessment tool for clinical reasoning between OTL-HNS residents and expert faculty. To improve the perception of SCT by residents, education and an increased exposure to the SCT would be necessary.

Keywords: Script, Concordance, Reasoning, Acceptability, Otolaryngology

Table des matières

Résumé.....	i
Abstract.....	iii
Table des matières.....	iv
Liste des tableaux.....	vii
Liste des figures.....	viii
Liste des sigles.....	ix
Liste des abréviations.....	x
Remerciements.....	xii
1. Introduction.....	1
2. Le raisonnement clinique.....	3
2.1 La taxonomie de Bloom.....	3
2.2 Les modèles de raisonnement clinique.....	4
2.2.1 Modèle associatif.....	5
2.2.2 Modèle analytique.....	6
2.2.3 Modèle du double processus.....	7
2.3 L'évaluation du raisonnement clinique.....	8
2.3.1 Les questions à choix multiples.....	8
2.3.2 Le <i>patient management problem</i>	9
2.3.3 L'examen par éléments clés.....	10
2.3.4 L'examen clinique objectif structuré.....	11
2.3.5 L'examen oral traditionnel.....	11
3. Le test de concordance de script.....	13
3.1 Les composantes du test de concordance de script.....	14
3.1.1 La mise en situation.....	14
3.1.2 L'hypothèse et l'information additionnelle.....	15
3.1.3 L'échelle de Likert.....	15

3.2 Le système de correction	16
3.2.1 Le groupe d'expert.....	16
3.2.2 Le solutionnaire	17
3.3 La validité du TCS	18
3.4 Les controverses associées au TCS.....	19
4. Objectif de l'étude.....	21
5. Méthodologie	22
5.1 Le cadre de l'étude.....	22
5.2 Le choix des experts.....	23
5.3 Les questionnaires.....	23
5.4 La collecte des données	24
5.5 L'analyse des résultats	25
6. Résultats.....	26
7. Discussion.....	31
7.1 L'acceptabilité du test de concordance de script	31
7.1.1 Les différences entre les résidents et les experts	31
7.1.1 Les différences entre les résidents juniors et séniors.....	33
7.2 L'importance du test de concordance de script.....	34
7.2.1 L'utilisation de multiples formes d'évaluation	34
7.2.2 L'évaluation et l'enseignement du raisonnement clinique	34
7.2.3 L'effet sur l'apprentissage	35
7.3 Les faiblesses de l'étude	35
7.4 L'implication des résultats.....	37
8. Conclusion	39
Bibliographie.....	i
Annexe 1. Exemple d'un questionnaire de type <i>patient management problem</i>	iv
Annexe 2. Exemple d'un questionnaire par éléments clés	v
Annexe 3. Liste des questions de type TCS utilisées dans l'examen NITE de 2013	ix

Annexe 4. Sondage post-test des résidents xvi
Annexe 5. Sondage post-test des experts xviii

Liste des tableaux

Tableau 1. Caractéristiques des résidents ayant participé à l'étude	26
Tableau 2. Caractéristiques des experts ayant participé à l'étude	27
Tableau 3. Réponses au sondage post test. Comparaison entre les résidents et experts.....	28
Tableau 4. Réponses au sondage post test. Comparaison entre les résidents juniors et séniors	29
Tableau 5. Proportion des participants qui recommanderaient d'utiliser le TCS lors d'évaluation en ORL-CCF.....	30

Liste des figures

Figure 1. Taxonomie de Bloom	3
Figure 2. Exemple d'un test de concordance de script.....	14

Liste des sigles

ECOS : Examen clinique objectif structuré

EEC : Examen par éléments clés

NITE : Examen national *in-training*

ORL-CCF : Otorhinolaryngologie et chirurgie cervico-faciale

OTL-HNS : *Otolarygology and Head and Neck Surgery*

PGY : *Post graduate year*

PMP : *Patient management problem*

QCM : Questions à choix multiple

SCT : *Script concordance test*

TCS : Test de concordance de script

Liste des abréviations

Ex. : Exemple

Etc. : etcétera

*« Le véritable enseignement n'est point de te parler mais de te conduire »
Antoine de Saint-Exupéry*

Remerciements

Ce mémoire est le résultat d'un travail qui n'aurait pu voir le jour sans l'aide de plusieurs. Je remercie tous les résidents et experts qui ont pris le temps de répondre aux questionnaires et de donner leur opinion sur le sujet. Sans leur participation, le projet n'aurait jamais pu voir le jour. Aussi, pour leur aide, soutient, encouragements et conseils, je remercie Dr Tareck Ayad et Dr Lily H. P. Nguyen qui ont su me guider au travers de ce projet. J'ai aussi eu la chance de recevoir de précieux conseils des Dr Bernard Charlin et Dr Stuart Lubarsky, qui sont reconnus pour leur expertise sur le test de concordance de script. Finalement, l'aide précieuse de Dr Brian Rotenberg, qui était responsable de l'examen national *in-training* en 2013, nous a permis de concrétiser l'implantation du TCS dans l'examen NITE.

À tous, merci.

1. Introduction

Au cours d'un programme de résidence en médecine, il est attendu du candidat qu'il acquière diverses connaissances, compétences et attitudes. L'une des compétences cruciales à acquérir et à développer est la capacité de raisonnement clinique. Le raisonnement clinique est souvent vu comme une composante essentielle de la compétence d'un médecin.

De nos jours, les programmes de formation préclinique et de résidences sont bien adaptés pour enseigner et évaluer l'acquisition de connaissances et d'habiletés cliniques. En oto-rhino-laryngologie, par exemple, les résidents seront évalués sur leurs connaissances au cours de différents examens, et sur leurs compétences techniques en salle d'opération sous supervision directe. Par exemple, les examens administrés en cours de résidence aux résidents du programme d'oto-rhino-laryngologie et chirurgie cervico-faciale de l'Université de Montréal sont des examens écrits à réponses courtes deux fois par an (dont l'examen national *in-training*) et des examens oraux deux fois par an. Dans ce contexte, il est pertinent de se demander quels types de connaissances ces examens évaluent réellement. Pour ce faire, nous avons revu la composition de l'examen national *in-training* (NITE) de 2012. Dans cet examen, 50% des questions portaient sur les connaissances et la compréhension, 45% sur l'application et l'analyse et seulement 5% sur la synthèse et l'évaluation qui correspondent au raisonnement clinique. Il y a donc place à la mise en œuvre de nouveaux tests permettant l'évaluation du raisonnement clinique.

L'une des options pour évaluer le raisonnement clinique est le test de concordance de script. Le TCS a été développé au début des années 2000 et a depuis fait l'objet de nombreuses publications, dans divers domaines, liés à la santé ou non. En oto-rhino-laryngologie, deux publications portent sur le test de concordance de script à ce jour. La première, celle de Kania et al. en 2011, conclut que le TCS permet de discriminer le niveau d'expérience clinique du participant¹. Puis, Iravani et al. , au début de 2016, a démontré la fiabilité du TCS lorsqu'administré à un groupe de résidents seniors en ORL-CCF. Dans cette étude, 26

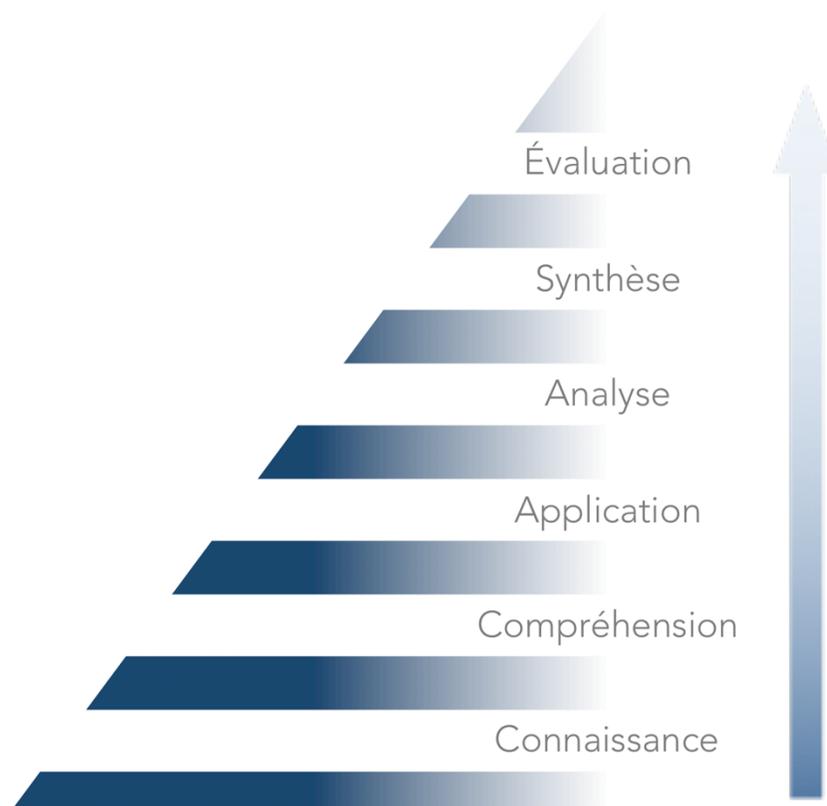
résidents de 3^e et 4^e année en oto-rhino-laryngologie et 9 experts ont répondu à un test comprenant 20 items et 60 questions de type TCS. Il y avait une différence statistiquement significative entre la moyenne des scores des résidents et des experts.

2. Le raisonnement clinique

Le raisonnement clinique est défini comme un processus complexe qui utilise des connaissances spécifiques à une discipline, la cognition, le contexte clinique ainsi que l'analyse des informations d'un patient, dans le but d'évaluer sa signification et de peser les différentes conclusions possibles ou actions alternatives². Le raisonnement clinique est souvent considéré comme une composante majeure de ce qui définit la compétence clinique d'un médecin³, de ce qui définit « un bon médecin ». La compréhension de ce qu'est le raisonnement clinique et la connaissance de ses différents processus est essentielle car ils nous permettent d'améliorer la façon dont le raisonnement clinique est enseigné et évalué⁴ au cours de la formation médicale.

2.1 La taxonomie de Bloom

Figure 1. Taxonomie de Bloom



La taxonomie de Bloom est le résultat d'un consensus de l'*American Psychological Association* qui s'était donné pour mandat d'établir des catégories d'objectifs et de processus mentaux à évaluer. La première publication de la taxonomie de Bloom fut émise en 1956. Elle comprend six niveaux d'apprentissages hiérarchisés, les plus bas étant nécessaires pour l'atteinte des niveaux supérieurs⁵. Ces niveaux sont la connaissance, la compréhension, l'application, l'analyse, la synthèse et l'évaluation. La connaissance réfère à la mémoire, le fait de se rappeler des faits, concepts, etc. La compréhension est démontrée par la capacité de comparer et d'organiser cette information. L'application est l'utilisation des connaissances pour résoudre un problème. L'analyse est la capacité de retirer une partie de l'information pour pouvoir l'utiliser. Finalement nous atteignons les niveaux supérieurs de domaines cognitifs soit la synthèse et l'évaluation, qui correspondent au raisonnement clinique. La synthèse est définie comme la capacité de combiner des éléments séparés pour apporter une réponse nouvelle et l'évaluation comme la capacité de défendre une réponse en émettant un jugement basé sur certaines connaissances ou critères.

La taxonomie de Bloom est à la fois utilisée comme base pour les différentes théories d'éducation mais aussi de façon plus concrète pour aider à balancer l'éducation et les évaluations. En effet, l'utilisation des différentes catégories de la taxonomie de Bloom de façon consciente lors de la mise en place d'un programme d'enseignement ou d'un outil d'évaluation permet de répartir le contenu pour l'adapter au niveau des apprenants ou à ce que l'on désire enseigner ou évaluer.

2.2 Les modèles de raisonnement clinique

Différents modèles ont été développés pour décrire les différents types de processus cognitifs utilisés pour résoudre un problème médical. Les principaux modèles sont le modèle associatif, le modèle analytique et le modèle du double processus.

2.2.1 Modèle associatif

Le modèle de raisonnement associatif (aussi appelé type 1², non-analytique ou intuitif³) est un chemin rapide de raisonnement clinique. Ce processus non-analytique est habituellement employé par les cliniciens d'expérience pour arriver à un diagnostic. En général, ce modèle ne nécessite pas d'effort conscient^{3,6} et c'est l'expérience clinique qui permet de l'acquérir au fil du temps⁶. Le fait que ce processus soit intuitif le rend particulièrement rapide³. Il peut être utilisé sous différentes formes dont la forme de reconnaissance de similarité ou de reconnaissance de forme.

Le modèle associatif utilisé sous sa forme de reconnaissance de similarité implique la comparaison du cas rencontré par le clinicien avec un autre cas similaire que ce dernier a connu dans le passé⁶. Avec l'expérience, les cliniciens accumulent un bagage de situations vécues qu'ils peuvent comparer à la situation actuelle. C'est principalement le temps qui permet d'acquérir plusieurs situations de référence pour pouvoir par la suite utiliser ce type de raisonnement clinique. Dans une moindre mesure, les cas références peuvent parfois être acquis lors de situations de simulation ou par l'exposé de cas rapportés par d'autres cliniciens.

Une autre forme de modèle associatif qui est utilisé est le modèle de la reconnaissance de forme. Ce modèle implique la réorganisation des informations en éléments cliniques clés² qui permettent d'émettre des hypothèses. Le clinicien transforme inconsciemment le discours du patient et les informations cliniques en axes sémantiques² ou paires d'opposés² (ex. : constant vs épisodique, aigu vs chronique). Une réorganisation de cette information en réseaux plus complexes (scripts) est nécessaire pour permettre une utilisation plus rapide de ces données² et d'aussi pour pouvoir émettre un diagnostic.

L'avantage du modèle associatif est qu'il permet au clinicien de générer tôt des hypothèses (solutions entrevues au problème posé⁴). Une étude réalisée en médecine d'urgence nous démontre que les cliniciens vont générer une moyenne de 5 hypothèses

diagnostiques et ce, même avant la rencontre du patient ou dans les premières minutes de la rencontre⁶. Les inconvénients de ce modèle est que le clinicien pourrait être tenté d'ignorer certains éléments pour accommoder la situation clinique avec les hypothèses initialement générées et ainsi générer une hypothèse finale erronée. Au contraire, le clinicien pourrait rester prisonnier de son script et ne pas être capable de prendre une décision par rapport à une situation clinique². Dans cette situation, les éléments cliniques de la situation rencontrée ne concordent pas parfaitement avec les éléments qui composent le script et le clinicien peut alors être incapable d'aller au-delà de ces différences pour émettre un diagnostic final.

2.2.2 Modèle analytique

Le modèle analytique (de type 2², rationnel ou délibéré³) est basé sur des données collectées consciemment par le clinicien et sur l'application de règles³. Le modèle analytique est souvent préféré dans les contextes plus complexes ou d'incertitude, et quand le temps le permet³. Deux formes de modèle analytique sont souvent citées dans la résolution de problèmes dans le domaine médical : le modèle hypothéticodéductif et le modèle de raisonnement en chaînage avant.

Le modèle hypothéticodéductif fait appel dans un premier temps à la génération d'hypothèses diagnostiques sur lesquelles sera basé la suite du raisonnement clinique, soit le recueil d'informations ou tests complémentaires⁴. La démarche analytique sera alors nécessaire pour vérifier si la réponse initialement choisie est compatible ou non avec les différentes valeurs contenues dans le script de la réponse. Le processus analytique se poursuit alors jusqu'à ce qu'assez de valeurs compatibles soient trouvées. Le but n'est donc pas de trouver toutes les valeurs possibles mais suffisamment de valeur pour confirmer l'hypothèse⁶.

Un second modèle analytique utilisé dans la démarche de raisonnement clinique, qui peut être considéré comme d'opposé du raisonnement hypothético-déductif², est le raisonnement en chaînage avant. Celui-ci est généralement utilisé lorsqu'une hypothèse

préalable ne peut être générée. Plusieurs facteurs peuvent empêcher la formulation d'une hypothèse précoce comme la complexité d'un cas, la rareté d'un cas ou le manque d'expérience⁶. En l'absence de prémisse, des règles causales ou conditionnelles seront mises en œuvres successivement afin de prioriser ou d'éliminer une hypothèse^{4,6}.

2.2.3 Modèle du double processus

Lorsque les cliniciens utilisent uniquement les modèles associatifs ou analytiques séparément, ils produisent un nombre d'erreur plus élevé, peu importe leur expérience³. Le modèle du double processus implique en général l'application du modèle de reconnaissance de forme suivi du modèle hypothético-déductif³.

La génération d'hypothèses précoces est faite de façon naturelle et irréfléchie. Surtout chez le clinicien expérimenté, il est pratiquement impossible de supprimer cette émission d'hypothèses⁴, même lorsque une demande explicite de ne pas les générer est faite⁷. Chez les moins expérimentés, la génération d'hypothèses précoces peut être supprimée mais conduit souvent à un nombre d'erreurs plus élevé⁷. Le système analytique permet dans un deuxième temps de confirmer ou non l'hypothèse³ générée par le modèle de reconnaissance de forme. Il se poursuivra jusqu'à l'obtention d'une réponse concordante. L'importance de l'enclenchement du système analytique se révèle surtout dans des situations complexes ou inhabituelles. Par contre, ce système n'est pas à toutes épreuves. En effet, des situations où la vigilance est réduite (manque de temps, fatigue), où le clinicien est trop confiant et où le clinicien manque de motivation³ peuvent résulter en une diminution de l'efficacité du système analytique et donc en une erreur.

Finalement, ce modèle ne délaisse pas complètement le savoir de sciences de base que le clinicien a acquis dans les premières années de sa formation. Ces connaissances restent disponibles et placent des contraintes sur ce qui est cohérent ou non avec une hypothèse donnée².

2.3 L'évaluation du raisonnement clinique

Plusieurs outils d'évaluation du raisonnement clinique existent, tant à l'écrit qu'à l'oral. Par contre, aucun test disponible permet d'évaluer tous les éléments qui composent le raisonnement clinique. Le plus souvent, une combinaison de différents tests sera nécessaire pour évaluer la capacité d'un candidat à avoir un raisonnement clinique juste. Les principaux tests écrits évaluant le raisonnement clinique sont les questions à choix multiples (QCM), le « Patient Management Problem » (PMP), l'examen par éléments clés (EEC) et le test de concordance de script (TCS). À l'oral, les examens oraux traditionnels ainsi que les examens cliniques objectifs structurés (ECOS) sont utilisés pour évaluer le raisonnement clinique.

2.3.1 Les questions à choix multiples

Le format de question à choix multiples est assez simple et connu par la majorité. Il comprend une question et une liste de choix de réponses qui se veulent plausibles. Dépendamment de la façon dont la question est écrite, les questions à choix multiples peuvent évaluer tous les niveaux de connaissance décrits dans la taxonomie de Bloom. Pour tester les niveaux de connaissances supérieurs, la question comprendra généralement un cas clinique et le participant devra utiliser à la fois ses connaissances de base sur la matière et sa capacité de synthèse et d'évaluation pour y répondre. Ce type de question est communément appelé QCM à contexte riche⁸. Les QCM à contexte riche semblent déclencher des processus de réflexion différents que les questions à choix multiples traditionnelles. En effet, l'étude de Schuwirth et al. suggère que les QCM contenant un cas clinique élicitent une réorganisation des informations données ainsi qu'un processus de raisonnement plus élaboré⁹. À l'opposé, les QCM traditionnelles vont plus souvent déclencher une analyse de « vrai ou faux »⁹ par rapport aux différentes réponses proposées plutôt qu'un processus de raisonnement.

Traditionnellement, les QCM n'étaient pas conçues pour évaluer le raisonnement clinique. Par contre, l'ajout d'un cas clinique dans la question peut permettre de l'évaluer. Bien que l'écriture de ce genre de question semble facile, il faut y consacrer du temps pour éviter les erreurs. Des erreurs communes sont la création d'une question avec trop d'ambiguïté, qui rend impossible d'arriver à une conclusion unique ou, au contraire, de donner trop d'indices vers la bonne réponse¹⁰. Aussi, beaucoup de médecins évoluant dans les milieux d'enseignement ont déjà eu à écrire des QCM traditionnelles. Pour certains d'entre eux, il pourrait être facile de tomber dans le piège d'écrire une QCM à contexte riche mais avec une formulation qui n'éluciderait pas l'utilisation du raisonnement clinique mais qui, au contraire, ne ferait appel qu'à la mémoire.

La principale limite du QCM à contexte riche est donc la difficulté de créer une question et des choix de réponses valides qui permettent réellement d'évaluer le raisonnement clinique. Par contre, cette difficulté dans la création des questions et le temps nécessaire pour le faire s'applique aussi au TCS.

2.3.2 Le *patient management problem*

Développé dans les années 60 et 70, le *patient management problem* (PMP) a pour objectif de simuler le processus d'évaluation d'une situation clinique, de choix du diagnostic le plus plausible et de prise de décisions thérapeutiques¹⁰. En 1978, Hubbard avait décrit le PMP comme un outil d'évaluation des compétences nécessaires pour identifier, résoudre et prendre en charge un problème médical¹¹. L'évaluation porte sur le raisonnement qu'a suivi le participant tout au long des questions, et est comparé à celui d'un groupe d'experts. Un exemple de ce type de question est disponible à l'annexe 1.

La principale critique du PMP est que le chemin de raisonnement suivi pour résoudre un problème diffère d'un expert à l'autre et peut donc différer d'un participant à l'autre sans toutefois être erroné. De plus, les participants moins expérimentés vont souvent utiliser une

démarche plus exhaustive, qui peut donc donner plus de point en cochant plus d'items, sans toutefois être le reflet d'un meilleur raisonnement clinique.

Comme le TCS, le PMP a été développé dans le but d'évaluer plus que les simples connaissances factuelles. Le but était de pouvoir tester le cheminement de pensée nécessaire lors de l'évaluation et de la prise en charge d'un patient¹². Par contre, le format du PMP est plus complexe. En effet, les questions auxquelles les participants auront à répondre dépendent des réponses données dans les questions précédentes. C'est donc dire que chacun des participants n'aura pas répondu aux mêmes questions au final, ce qui rend la notation du test plus complexe.

2.3.3 L'examen par éléments clés

L'examen par éléments clés est conçu à partir d'une série de cas cliniques qui ont pour but d'évaluer les éléments clés de la prise en charge d'un cas¹⁰. Les questions qui s'en suivent peuvent être de différentes natures (QCM, réponse courte, réponse longue). Le concept d'élément-clé vient de la prémisse que tous les éléments d'un problème n'ont pas la même importance et que certains éléments font une différence majeure dans la résolution du problème¹³. L'EEC permet donc de simuler des tâches réelles et d'évaluer le raisonnement clinique si la question est bien posée. Un exemple de ce type de question est disponible à l'annexe 2.

Bien que le temps nécessaire pour préparer ce genre d'examen soit long, il reste plus court que pour d'autres types de questionnaires tels que le PMP ou le TCS. La principale faiblesse de l'examen par éléments clés provient du fait que le solutionnaire est fait à l'aide d'un consensus d'experts. Pour obtenir un consensus, les examinateurs auront tendance à choisir des questions plus simples et qui font l'unanimité, ce qui peut rendre l'examen trop facile. Au contraire, le format du solutionnaire du TCS ne nécessite pas que les experts soient arrivés à un consensus. En effet, le test prend en compte que pour les cas avec un certain degré

d'incertitude, le consensus n'est pas possible autant pour les experts que dans les réponses des résidents.

2.3.4 L'examen clinique objectif structuré

L'ECOS est une méthode d'évaluation utilisée dans le domaine médical depuis 1975¹⁴ et est toujours utilisé comme évaluation sommative par le Collège Royal des médecins et chirurgiens du Canada. L'évaluation se fait sous forme de stations standardisées d'une durée de 10 à 15 minutes généralement. La station comprend habituellement un patient standardisé avec lequel le candidat évalué devra interagir. Une grille d'évaluation formelle est remplie par un évaluateur selon des critères bien précis. L'ECOS peut être modifié pour permettre d'évaluer des compétences de raisonnement clinique, de relations interpersonnelles ou des compétences techniques. Il a été démontré que l'ECOS permet d'évaluer la capacité d'interprétation des données ainsi que le manque de connaissances cliniques¹⁵.

Selon certains auteurs, l'un des avantages de l'ECOS serait qu'il entraînerait moins de stress que l'examen oral traditionnel¹⁶. Les inconvénients d'implanter l'ECOS comme outil d'évaluation des résidents sont le temps nécessaire et les coûts reliés à la mise en place de l'examen, ainsi que le manque de réalisme des scénarios qui sont souvent des situations cliniques classiques qui ne reflètent pas la réalité¹⁴.

2.3.5 L'examen oral traditionnel

L'examen oral traditionnel est aussi utilisé de façon routinière dans les différents programmes de résidence, tant sous forme formative que sommative. Le principal avantage de l'examen oral est qu'il permet d'évaluer plusieurs niveaux de connaissance et permet au participant de justifier ses réponses par des explications. Par contre, ce type d'examen est limité à la fois par des caractéristiques du participant (anxiété, stress, attitude) et de l'évaluateur (clémence, sévérité, conflit interpersonnel)¹⁰. Même si certaines caractéristiques

interpersonnelles peuvent jouer un rôle dans l'objectivité du test, les questions sont généralement standardisées et les réponses doivent faire l'unanimité pour que la question soit à l'examen.

Au contraire, étant un examen écrit, le TCS génèrerait moins de stress et d'anxiété que les examens oraux traditionnels. De plus, tous les facteurs humains qui peuvent changer l'attitude du participant ou celle de l'évaluateur sont éliminés. L'autre avantage du TCS par rapport à l'examen oral traditionnel est sa plus grande flexibilité d'administration. En effet, un examen écrit ne demande pas autant de ressources de temps et de personnel. Il permet aussi d'évaluer plus de candidats à la fois.

3. Le test de concordance de script

Le concept de script a été introduit pour la première fois dans la littérature médicale en 1983 par Clancey¹⁷ et est défini comme une structure dans laquelle est organisé les connaissances médicales ainsi que les liens entre ces connaissances¹⁸. La sélection d'un script permet de débiter l'analyse de la situation par un seul élément ou un petit groupe d'éléments de l'histoire clinique. Le clinicien compare ensuite la situation clinique à différents éléments du script pour trouver la meilleure hypothèse. Les scripts sont le reflet de la maîtrise qu'a un clinicien d'un domaine particulier. Plus les scripts sont complexes et précis, plus le clinicien sera maître dans ce domaine². Le clinicien accumule au fil du temps plusieurs scripts comprenant plusieurs formes nuancées et organisés d'une même condition².

Le test de concordance de script est défini comme un test permettant d'évaluer la capacité d'interpréter l'information médicale dans des conditions d'incertitude¹⁸. Il reflète donc la pratique clinique dans laquelle des décisions importantes doivent être prises, dans des contextes comportant un certain degré d'incertitude, d'imprécision ou avec des données incomplètes. Le test de concordance de script part du principe que le participant a déjà acquis un certain nombre de connaissances de base dans la matière évaluée.

L'introduction du test de concordance de script dans la littérature médicale comme outil d'évaluation date de la fin des années 90, alors appelé test de script diagnostique¹⁹. Depuis, plusieurs études se sont penchées sur la validité et fiabilité du TCS, sur l'évaluation de sa forme et de sa méthode de correction et sur les moyens d'optimiser les caractéristiques du test.

3.1 Les composantes du test de concordance de script

Un questionnaire de type TCS est un ensemble de cas cliniques auxquels sont reliées des questions. Idéalement, pour avoir la meilleure fiabilité, le test devrait comprendre entre 20 et 25 cas, ayant chacun entre 2 et 4 questions²⁰.

Figure 2. Exemple d'un test de concordance de script. A - Mise en situation. B - Hypothèse. C - Information additionnelle. D - Format de réponse.

Vous devez répondre aux questions suivantes concernant la probabilité du diagnostic en utilisant l'échelle suivante. Le diagnostic est :				
-2	-1	0	+1	+2
Beaucoup moins probable	Moins probable	Ne change pas le diagnostic	Plus probable	Beaucoup plus probable

A 1- Une fillette de 1 an est référée à votre clinique pour toux sèche chronique par son médecin de famille. Elle ne présente aucun symptôme infectieux présentement, ni dans le dernier mois.

B Si vous pensez à :	C Et vous apprenez que :	D Le diagnostic devient
a. Reflux gastro oesophagien	Elle prend Ranitidine (8 mg/kg/jour) depuis 8 semaines et il n'y a aucune amélioration.	-2 <input type="checkbox"/> -1 <input type="checkbox"/> 0 <input type="checkbox"/> +1 <input type="checkbox"/> +2 <input type="checkbox"/>
b. Écoulement nasal postérieur	Elle est capable de respirer par le nez et il n'y a pas de rhinorrhée.	-2 <input type="checkbox"/> -1 <input type="checkbox"/> 0 <input type="checkbox"/> +1 <input type="checkbox"/> +2 <input type="checkbox"/>
c. Fibrose kystique	Elle tousse seulement la nuit.	-2 <input type="checkbox"/> -1 <input type="checkbox"/> 0 <input type="checkbox"/> +1 <input type="checkbox"/> +2 <input type="checkbox"/>

3.1.1 La mise en situation

La mise en situation qui débute chaque groupe de questions, aussi appelée stimulus¹⁸, est souvent tirée de cas cliniques rencontrés en pratique. La description du cas se veut courte et réaliste, et permet de placer la question dans un contexte clinique. De plus, la situation doit contenir un niveau d'incertitude ou de données contradictoires pour pouvoir favoriser le

raisonnement clinique. Ce degré de contradiction ne doit cependant pas rendre la situation clinique irréaliste et improbable, et c'est pour cette raison que les cas sont souvent basés sur des situations cliniques réelles. L'incertitude et les données contradictoires se veulent un reflet de la pratique clinique où les patients se présentent parfois avec une historique incomplète ou difficile à interpréter.

3.1.2 L'hypothèse et l'information additionnelle

Une hypothèse se rapportant au diagnostic, à l'évaluation clinique ou au traitement est ensuite fournie au participant. Cette hypothèse est une hypothèse plausible dans le contexte de ce cas. Le but de cette hypothèse est d'activer un script chez le répondant¹⁸. Comme le but est de simuler le raisonnement clinique normal, il n'y aurait pas d'avantage à mettre une hypothèse improbable par rapport au cas clinique car cela ne ferait que générer de la confusion chez le répondant. Par la suite, une information additionnelle est apportée au cas. Cette nouvelle information se veut la simulation du processus de recherche d'informations nouvelles pour confirmer ou infirmer l'hypothèse, comme le fait le clinicien lorsqu'il pose des questions supplémentaires au patient pour confirmer ou infirmer son hypothèse diagnostique. Dans l'exemple de la Figure 2, l'hypothèse correspond à la section « Si vous pensez à » et l'information additionnelle à la section « Et vous apprenez que ».

3.1.3 L'échelle de Likert

La réponse du participant est donnée sur une échelle de Likert. La réponse du participant correspond au changement de probabilité qu'apporte la nouvelle information sur la probabilité de l'hypothèse initiale fournie. Le TCS utilise habituellement une échelle de Likert sur 5 points, de -2, -1, 0, +1, +2, signifiant que l'hypothèse est soit beaucoup moins probable, moins probable, inchangée, plus probable ou beaucoup plus probable suite à l'information ajoutée. Initialement, le TCS avait été décrit avec une échelle de Likert à 7 points¹⁹, mais qui ne s'est pas montrée supérieure à l'échelle de 5 points. L'échelle de 5 points, étant plus simple, a alors été adoptée.

D'autres auteurs suggèrent que peu d'information est perdue en passant d'une échelle de 5 points à une échelle de 3 points^{21,22}. Dans l'échelle de 3 points, le participant doit seulement choisir si la nouvelle information diminue ou augmente la probabilité de l'hypothèse. Par contre, d'autres études démontrent un avantage de l'échelle à 5 points et arrivent à la conclusion que l'échelle de 5 points a une fiabilité et un coefficient de corrélation plus forts que l'échelle de 3 points²³.

3.2 Le système de correction

3.2.1 Le groupe d'expert

Un groupe d'expert est formé pour répondre aux questions du TCS. Les réponses de ce groupe serviront à former le solutionnaire pour la correction des réponses au test. Le nombre de membres du groupe d'expert nécessaire pour obtenir une fiabilité acceptable a été évalué à un minimum de dix experts. Par contre, le nombre idéal serait de vingt experts. Au delà de vingt experts, il y aurait peu de bénéfices ajoutés²⁴ surtout lorsque l'on prend en compte la difficulté de recruter les experts.

Une critique rapportée du TCS est dans la définition même de ce que l'on considère un expert. Pour former le groupe, différents critères sont utilisés mais aucune mesure véritable de l'expertise n'est disponible²⁵. En général, le choix des experts est fait parmi des médecins ayant terminé leur formation dans la spécialité évaluée et pratiquant toujours. Le choix d'experts évoluant dans un milieu d'enseignement est souhaitable car la concordance entre les résultats des résidents et ceux d'experts en milieu académique était plus élevé qu'entre ceux des résidents et ceux des experts pratiquant en milieu non académique. Par contre, les deux groupes d'experts ont produit un classement similaire des résidents²⁶. Donc, dépendamment de ce qui est évalué et de l'importance de l'évaluation, l'utilisation d'experts qui ne travaillent pas dans des centres d'enseignement pourrait être acceptable.

Enfin, l'étude de Petrucci de 2013²⁷ a apporté une nouvelle notion, soit celle d'utiliser plusieurs petits groupes d'experts ayant des sur-spécialisations au sein d'une même spécialité. Ces sous-groupes d'experts ont pour mandat de répondre aux questions se rapportant à leur sur-spécialité et non à tout le questionnaire. Cette idée a été développée suite à une étude du même groupe en 2007, qui a démontré une moyenne de score plus faible pour les résidents finissants que pour les résidents de 4^e année²⁸. L'hypothèse a alors été avancée que les résidents de dernière année, qui étudient pour leurs examens finaux, ont les connaissances les plus à jour, se rapprochant donc plus de l'opinion des sur-spécialistes dans certains domaines. Comme les connaissances médicales évoluent très rapidement, un médecin se spécialisant dans une sur-spécialité précise pourrait ne plus être à jour en ce qui a trait au développement des autres sur-spécialités de sa discipline principale. C'est pour éviter ce problème que l'option d'utiliser différents sous-groupes d'experts a été développée. L'un des désavantages d'utiliser des sur-spécialistes est qu'il devient alors plus difficile de recruter un nombre suffisant d'experts par groupe. L'étude de Petrucci comptait 5 experts par sous-catégorie pour une totalité de 30 experts. Cette étude a tout de même démontré un coefficient de fiabilité élevé (0.81)²⁷. Nous pouvons alors probablement déduire que 5 experts par sous-groupe serait un nombre adéquat.

3.2.2 Le solutionnaire

Le solutionnaire est formé à partir des réponses du groupe de référence. La méthode de correction traditionnelle du TCS implique une correction sur une échelle de Likert à 5 points sous forme d'agrégat, c'est à dire que la réponse la plus populaire au sein du groupe d'expert donne 1 point. Les autres réponses choisies par une minorité d'experts donne une fraction des points. Cette fraction correspond au nombre d'experts ayant choisi la réponse divisé par le nombre d'experts ayant choisi la meilleure réponse. Les réponses non choisies donnent 0.

Lors de la construction du solutionnaire, il peut arriver que les réponses de certains experts dévient beaucoup de celles données par la majorité du groupe. La question s'est alors posée à savoir si les réponses divergentes devraient être retirées du solutionnaire. Une étude de Gagnon en 2011 a comparé les résultats obtenus au test lorsque corrigé avec quatre solutionnaires différents, trois solutionnaires obtenus avec différentes méthodes pour retirer les réponses déviantes et un sans le retrait des réponses déviantes²⁹. Les méthodes de retrait des réponses déviantes étaient : de retirer les experts avec un score total trop bas, de retirer les réponses jugées déviantes par un groupe et de retirer les réponses trop loin du mode (valeur la plus fréquente dans un échantillon). Avec un grand groupe d'expert, le fait de retirer les réponses déviantes n'avait aucun impact sur les résultats. Par contre, si le retrait des réponses déviantes est jugé nécessaire, le groupe suggérait la méthode de distance par rapport au mode²⁹. Au contraire, certains pensent qu'il faut éviter de retirer les réponses déviantes³⁰. En effet, nous savons que différents chemins de réflexion sont empruntés par différents individus pour résoudre le même problème. Un chemin différent n'est pas nécessairement une erreur et un participant pourrait avoir le même chemin de pensée qu'un expert et donc recevoir une partie des points pour sa réponse.

3.3 La validité du TCS

On peut définir la validité d'un test comme étant sa capacité à mesurer ce qu'il est sensé mesurer. De nombreuses études ont démontré la validité du TCS dans plusieurs domaines différents dont la pédiatrie³¹, l'urologie³², la médecine d'urgence³³, l'oto-rhino-laryngologie³⁴, l'optométrie³⁵, les sciences infirmières³⁶, etc., à différents niveaux de formation (étudiants³⁷, résidents, experts) et dans différents buts (auto-évaluation³⁸, évaluation, formation continue).

La validité prédictive du TCS a été évaluée en comparant les résultats d'étudiants en médecine avec leurs résultats de fin de résidence deux ans plus tard. Le TCS a une bonne valeur prédictive lorsque comparé aux questions de type problèmes à réponse courte et

examen oral, qui évaluent aussi le raisonnement clinique. Par contre, lorsque comparé aux ECOS, le TCS prédit moins bien la performance des résidents. Cette différence peut être expliquée par le fait que les ECOS évaluent aussi les compétences techniques en plus du raisonnement clinique³⁹.

Les scores au TCS ont aussi été comparés au National Board of Medical Examiners et montraient une bonne validité⁴⁰. Lorsque comparés aux scores de QCM à contexte riche, qui évaluent aussi le raisonnement clinique, les résultats des répondants vont dans le même sens que le TCS mais sans corrélation significative⁸. Cela suggère donc que le QCM et le TCS n'évaluent pas la même composante du raisonnement clinique et sont donc complémentaires.

3.4 Les controverses associées au TCS

Peu d'études sur le test de concordance de script ont questionné la validité de ce test. Dans ces études, trois points de discordance sont mis de l'avant : L'utilisation du coefficient α comme mesure de validité, le modèle de notation utilisant les points partiels et l'agrégat, et l'influence du style de réponse du participant sur le pointage.

En 2013, Lineberry et al. a révisé la littérature sur la validité du TCS et en est venu à la conclusion que pratiquement toutes les études utilisaient le coefficient α comme mesure de validité⁴¹. Bien que ce soit une mesure acceptée pour évaluer la cohérence des réponses du participant, cette mesure ne teste pas la corrélation entre les groupes d'experts et la corrélation entre les experts. Ces deux dernières variables nous permettraient d'évaluer si les experts choisis représentent réellement un échantillon des experts du domaine et si les réponses des experts sont en générales cohérentes inter-experts⁴¹. C'est donc dans ce contexte que Lineberry et al. n'étaient pas convaincus de la validité du TCS, ce qui a été réitéré dans deux autres articles des mêmes auteurs^{42,43}.

Un autre point de discordance est l'utilisation de points partiels. Cette forme de notation est basée sur le fait qu'il peut y avoir des divergences de réponse qui sont tout de même valides. Par contre, il est pertinent de se questionner à savoir si cette divergence n'est pas plutôt causée par une moins bonne performance de certains experts. Certains auteurs ont plutôt suggéré d'utiliser un consensus d'experts ou de se référer à des données empiriques pour créer le solutionnaire⁴¹.

Finalement, le style de réponse du répondant a été évalué et celui-ci peut grandement influencer le résultat du participant. En effet, le fait d'éviter les extrêmes (réponses +2 et -2) améliorerait le résultat au test^{43,44}. Ceci est expliqué, entre autre, par le fait que lorsque les questions sont écrites, il y a moins de chance d'écrire des questions dont la réponse modale serait à l'un des extrêmes⁴¹. Comme il y a moins de chance que le cas clinique présente un contexte d'incertitude lorsque la réponse est +2 ou -2, la question ne sera pas écrite ou ne sera pas choisie pour le solutionnaire. De plus, les questions ayant des réponses aux extrêmes testent plus souvent les connaissances factuelles que le raisonnement clinique. Finalement, en cas de doute, le participant qui donne une réponse partielle (+1 ou -1) a plus de chance d'obtenir des points partiels vu la distribution en agrégat des réponses, que s'il répond aux extrêmes. Donc, c'est dire qu'un participant pourrait être entraîné pour répondre aux TCS afin d'améliorer son score sans améliorer ses connaissances ou son raisonnement clinique⁴⁴.

4. Objectif de l'étude

Depuis le début des années 2000, le test de concordance de script a fait l'objet de plusieurs dizaines d'études. Ces études nous ont convaincu de la validité de cet outil d'évaluation à plusieurs niveaux de formation et dans plusieurs domaines de la santé. Par contre, ce test demeure peu utilisé au sein des programmes de résidence, partout au Canada. Le but de faire cette étude est d'évaluer à quel degré les gens connaissent le TCS ainsi que les différents facteurs qui peuvent freiner son acceptation, dans le but de pouvoir plus facilement l'intégrer aux évaluations du programme de résidence.

Notre hypothèse est que le test de concordance de script est difficile à intégrer dans les programmes de résidence car il est encore méconnu. Après des professeurs, cette méconnaissance rend l'utilisation du TCS difficile pour les évaluations formatives et sommatives des résidents. Le temps supplémentaire nécessaire pour se familiariser et comprendre le TCS, ainsi que pour s'adapter à la méthode d'élaboration du questionnaire et à la production du solutionnaire peut affecter l'acceptabilité du TCS par les médecins responsables de l'enseignement aux résidents. Après des résidents, cela entraîne plutôt une difficulté supplémentaire pour répondre aux questions vu le format inhabituel de celles-ci. En effet, le résident qui ne connaît pas le TCS doit prendre plus de temps pour comprendre le processus pour répondre aux questions et pourrait répondre de façon inexacte s'il n'a pas pris le temps de le faire. Ces obstacles peuvent donc nuire à l'acceptabilité du TCS par les résidents.

L'objectif de notre étude était donc d'évaluer l'acceptabilité actuelle du TCS à un niveau national auprès des résidents en ORL-CCF et d'un groupe de référence composé d'experts en pratique universitaire.

5. Méthodologie

5.1 Le cadre de l'étude

L'examen national *in-training* (NITE) en oto-rhino-laryngologie et chirurgie cervico-faciale est un examen formatif national annuel qui sert à évaluer les connaissances des résidents en ORL-CCF. Cet examen a pour objectif de préparer les résidents à leur examen final de spécialité et de permettre d'identifier précocement les résidents qui pourraient avoir besoin d'aide supplémentaire. Habituellement, l'examen NITE ne comporte que des questions à réponses ouvertes et courtes ou à développement court. L'examen est d'une durée de trois heures. Pour les années de 2013 et 2016, le comité de l'examen NITE avait inclus des questions de type TCS à leur examen. Le TCS comptait 8 cas clinique pour un total de 28 questions. Les questions étaient séparées du reste de l'examen et il était mentionné explicitement que ces questions n'étaient pas comptabilisées dans le score total des résidents. Un questionnaire post-test était ensuite fourni aux résidents pour recueillir leurs impressions. Un ajout de 20 minutes était alloué pour répondre au TCS et au questionnaire post-test.

Dans le cadre de notre étude, nous avons eu l'accord du comité d'éthique (#13-120-CERES-D) de l'Université de Montréal pour une utilisation secondaire des données obtenues par le comité du NITE. Avec le consentement implicite des participants, le comité nous a fourni de façon anonyme les résultats au sondage post-test des résidents. Seul le niveau de formation des résidents était connu. Nous avons aussi obtenu la liste des questions de type TCS utilisées dans le NITE que nous avons ensuite subdivisé en catégories soit : neuro-otologie, rhinologie, oncologie, pédiatrie. Ces questions ont été fournies aux experts pour la formation du TCS. Un questionnaire post-test semblable à celui des résidents a été réalisé pour obtenir les impressions des experts sur le TCS.

5.2 Le choix des experts

Le recrutement des experts s'est fait au sein des 4 universités québécoises ayant un programme d'ORL-CCF soit l'Université de Montréal, l'Université McGill, l'Université de Sherbrooke et l'Université Laval. Au Québec, une entente entre les comités d'éthique des différentes universités existe, ce qui permet d'obtenir une seule approbation éthique valide dans toutes les universités par la suite. Si nous avions étendu notre étude au reste du Canada, comme nous l'avons fait pour les résidents, nous aurions dû obtenir une approbation éthique de chacune des 9 autres universités canadiennes ayant un programme d'ORL-CCF et inclure un collaborateur dans chacune de ces universités. Nous avons donc choisi d'inclure seulement le Québec, en s'appuyant sur l'hypothèse que comme les résidents canadiens passent tous le même examen de certification en fin de résidence et que les différents cours de révision sont donnés et suivis par des médecins de tous le Canada, la formation et la pratique devait être assez uniforme pour pouvoir extrapoler les résultats obtenus des universités québécoises seulement.

Pour cette étude, nous avons choisi de définir les experts comme des oto-rhino-laryngologistes travaillant en milieu universitaire et ayant complété une sur-spécialité en oto-rhino-laryngologie pédiatrique, en oto-rhino-laryngologie oncologique, en rhinologie ou en otologie. Comme mentionné dans l'étude de Lurie en 2011, différents critères peuvent être utilisés pour former le groupe d'experts mais aucune mesure de l'expertise n'est disponible²⁵. Nous avons donc choisi d'inclure des médecins qui sont toujours en pratique, pratiquant dans un hôpital universitaire où des résidents en ORL-CCF sont formés, ayant complété une année ou plus dans une sur-spécialité de l'ORL-CCF et qui pratiquent dans une des 4 universités québécoises offrant une formation en ORL-CCF.

5.3 Les questionnaires

Les questions de type TCS ont été créées pour l'examen national in-training de 2013. Le questionnaire a été construit en suivant les lignes directrices publiées en 2008 par Fournier

sur la création d'un test de concordance de script⁴⁵. Une équipe comprenant un résident et deux médecins ayant une pratique universitaire en ORL-CCF ont produit les questions. Ces questions furent ensuite révisées par deux experts dans le domaine du TCS soit Dr Bernard Charlin et Dr Stuart Lubarsky. Dr Charlin est considéré comme l'un des pères du test de concordance de script et a publié plus d'une quarantaine d'articles sur le test de concordance de script. Dr Lubarsky a lui aussi un intérêt sur le test de concordance de script depuis plusieurs années et a plus d'une dizaine de publications sur le sujet. Après la révision des questions, le questionnaire a été donné aux responsables du NITE qui l'ont approuvé. À notre connaissance, aucune validation des questions n'a été faite ni de test préalable.

Les post-tests présentés aux résidents et experts ont été créés pour cette étude. Les auteurs de l'étude ont chacun créé des questions qu'ils croyaient pertinentes pour l'étude après avoir fait la revue de littérature. Les questions ont ensuite été discutées et celles qui faisaient l'unanimité entre les 3 auteurs principaux ont été gardées. Les questions ont ensuite été révisées par Dr Bernard Charlin.

Le test de concordance de script ainsi que les questionnaires post-test ont été créés en anglais et en français par l'équipe de rédaction. L'un des membres de l'équipe a une pratique principalement en Anglais et c'est assuré de l'exactitude de la traduction. Nous nous sommes aussi assurés que le sens des questions reste le même lors de la traduction.

5.4 La collecte des données

La collecte des données des résidents s'est faite en collaboration avec le comité du NITE. Ils nous ont fourni les réponses des résidents au TCS ainsi que le sondage post-test anonyme pour les années 2013 et 2016. Seul le niveau de formation du résident était connu et s'il avait répondu au questionnaire de 2013. Le seul critère d'exclusion utilisé pour le groupe des résidents était d'avoir déjà répondu au questionnaire en 2013 pour les gens ayant participé au NITE de 2013 et 2016. Dans ce cas, nous avons exclu le post-test de 2016.

Pour la collecte des données provenant des experts, des courriels ont été envoyés aux directeurs de programme des universités québécoises pour que l'information soit ensuite redistribuée aux professeurs de leur université. Le courriel était alors redirigé à tous les médecins ayant une formation dans l'une des sur-spécialités représentées dans notre questionnaire et pratiquant dans l'un des centres hospitaliers des universités mentionnées précédemment. Un courriel explicatif ainsi que le formulaire de consentement, la section du TCS correspondant à la sur-spécialité du destinataire ainsi que le sondage post-test ont été envoyés à deux reprises en 2013 et une fois en 2016. Les participants pouvaient nous renvoyer les documents remplis par courriel ou par fax. Le nom du participant était retiré avant de compiler les réponses au sondage post-test.

5.5 L'analyse des résultats

Pour l'analyse des résultats aux sondages post-test, nous avons fait une comparaison des résultats entre les deux groupes avec le test U de Mann-Whitney, avec une valeur p significative à 0,05 à l'aide du logiciel IBM SPSS Statistics 24. Aussi, une analyse qualitative des résultats aux questions à développement a été effectuée pour les deux groupes.

Nous avons choisi de ne pas analyser les résultats du TCS pour 2 raisons. La première raison est que les questions du TCS n'ont pas été validées et qu'aucune étude pilote n'a été faite avec les questions. La deuxième étant que l'objectif de notre étude n'était pas de valider le TCS mais plutôt d'en évaluer l'acceptabilité.

6. Résultats

En combinant la participation au NITE de 2013 et 2016, 203 résidents étaient éligibles pour la participation à cette étude. De ce nombre, 36 résidents (18%) ont participé à l'étude soit 20 résidents juniors (deuxième et troisième année de résidence) et 16 résidents seniors (quatrième et cinquième année de résidence) (Tableau 1). Pour ce qui est des experts, un total de 23 experts ont participé à l'étude sur un total de 66 (35%). Tel qu'indiqué dans le Tableau 2, la majorité (78%) des experts était affiliée à l'Université de Montréal.

Pour ce qui est du niveau de connaissance sur le TCS, 72% des résidents et 52% des experts n'avaient jamais entendu parler du test de concordance de script avant cette étude. Parmi tous les participants ayant des connaissances sur le TCS, 68% affirment que leurs connaissances sur ce sujet sont faibles. Donc, un total de 86% des participants ont une connaissance nulle ou faible du TCS lorsque l'on combine les résidents et les experts.

Tableau 1. Caractéristiques des résidents ayant participé à l'étude

		N (36)	Pourcentage
Niveau de résidence	R2	6	16,7%
	R3	14	38,9%
	R4	10	27,8%
	R5	6	16,7%
	Total	36	100%
Connaissance du TCS avant l'étude	Oui	9	25,0%
	Non	26	72,2%
	Incertain	1	2,8%
Année de participation	2013	15	41,7%
	2016	21	58,3%
Connaissance du TCS selon l'année de participation	2013	3	20,0%
	2016	6	28,6%

Tableau 2. Caractéristiques des experts ayant participé à l'étude

		N (23)	Pourcentage
Sur-spécialité	Pédiatrie	5	21,7%
	Otologie	5	21,7%
	Oncologie	8	34,8%
	Rhinologie	5	21,7%
Université d'appartenance	Université de Montréal	18	78,3%
	Université Sherbrooke	4	17,4%
	Université Laval	1	4,3%
	Université McGill	0	0%
Connaissance du TCS avant l'étude	Oui	10	43,5%
	Non	12	52,2%
	Incertain	2	8,7%
Année de participation	2013	20	87,0%
	2016	3	13,0%
Connaissance du TCS selon l'année de participation	2013	9	45,0%
	2016	1	33,3%

Le tableau 3, représentant la comparaison entre les réponses au sondage post-test des résidents et experts, présente le pourcentage des répondants étant en accord ou fortement en accord avec les différentes énoncées. Lorsque le tableau est regardé dans sa globalité, les résultats qui y sont présentés suggèrent que les experts ont une opinion générale plus positive du TCS que les résidents. Pour toutes les questions du post-test, une différence statistiquement significative était présente entre les résidents et les experts.

Les experts, par rapport aux résidents, trouvaient dans une plus grande proportion que les explications fournies au début du questionnaire étaient claires, que les exemples de cas étaient utiles, que les questions étaient claires, qu'il était facile de s'adapter à cette méthode d'évaluation, que les cas reflétaient des situations cliniques réelles, que les questions reflétaient des situations qui pourraient être rencontrées dans la pratique et que le TCS est utile pour tester la capacité de raisonnement clinique en ORL.

Tableau 3. Réponses au sondage post test. Comparaison entre les résidents et experts

	En accord / Fortement en accord (%)		<i>P</i> ($<0,05$)	N	%
	Résidents	Experts			
Les explications fournies au début du questionnaire étaient claires	65,7%	87,0%	0,014	58	98,3%
Les exemples de cas présentés avant le questionnaire étaient utiles	62,9%	87,0%	0,007	58	98,3%
Les questions se rapportant aux cas étaient claires	44,4%	91,3%	0,000	59	100%
Il était facile de s'adapter à cette méthode d'évaluation	30,5%	73,9%	0,001	59	100%
Les questions reflètent des situations cliniques réelles	69,4%	100%	0,003	59	100%
Les questions reflètent des situations que j'ai déjà rencontré ou que je pourrais rencontrer durant ma résidence ou ma pratique future	72,2%	100%	0,005	59	100%
Cette évaluation est utile pour tester la capacité de raisonnement clinique en ORL	50,0%	91,3%	0,002	59	100%

La comparaison des réponses au sondage post-test entre les résidents juniors et séniors (Tableau 4) ne démontre pas de différence statistiquement significative à l'exception d'une seule question. En effet, les résidents juniors ont, dans une plus grande proportion, trouvé que le temps accordé pour répondre aux questions était adéquat par rapport aux résidents séniors, ce qui était statistiquement significatif.

Tableau 4. Réponses au sondage post test. Comparaison entre les résidents juniors et séniors

	En accord / Fortement en accord (%)		<i>p</i> (<0,05)	N	%
	Résidents juniors	Résidents séniors			
Les explications fournies au début du questionnaire étaient claires	65,0%	62,5%	0,799	35	97,2%
Les exemples de cas présentés avant le questionnaire étaient utiles	65,0%	56,3%	0,397	35	97,2%
Les questions se rapportant aux cas étaient claires	55,0%	31,3%	0,530	36	100%
Il était facile de s'adapter à cette méthode d'évaluation	35,0%	25,0%	0,503	36	100%
Le temps accordé pour répondre aux cas était adéquat	75,0%	37,5%	0,036	35	97,7%
Les questions reflètent des situations cliniques réelles	75,0%	62,5%	0,340	36	100%
Les questions reflètent des situations que j'ai déjà rencontré ou que je pourrais rencontrer durant ma résidence ou ma pratique future	70,0%	75%	0,557	36	100%
Cette évaluation est utile pour tester la capacité de raisonnement clinique en ORL	50,0%	56,3%	0,801	36	100%

Une revue a aussi été faite des réponses aux questions ouvertes du post-test. Lorsque questionnés sur ce qu'ils avaient le moins aimé du TCS, les résidents mentionnaient que le format les rendait parfois confus et qu'ils n'étaient pas habitués de répondre et de penser dans un format semblable. Aussi, plusieurs mentionnaient leur inquiétude en regard à la subjectivité des réponses et de l'évaluation négative que cela pourrait leur apporter. Finalement, certains mentionnaient qu'ils ne comprenaient pas ce que les questions évaluaient. Pour les experts, les éléments moins appréciés du TCS étaient le manque d'information pour prendre une décision éclairée, la simplicité de certaines questions et l'absence d'une réponse parfaite. L'autre élément mentionné par les deux groupes est qu'il était difficile de départager les réponses beaucoup moins probables des réponses moins probables (-2 et -1) et les réponses plus probables des réponses beaucoup plus probables (+1 et +2). Pour ce qui est des points

positifs, les résidents ont apprécié le réalisme des scénarios, le fait que le test n'évalue pas les connaissances mais plutôt le cheminement pour arriver à la bonne réponse et le raisonnement clinique. Certains ont même mentionné qu'ils appréciaient l'échelle de Likert à 5 points car cela leur permettait de « donner de la force » à leur raisonnement.

Finalement, pour ce qui est des réponses à la question d'inclure le TCS comme méthode d'évaluation en ORL-CCF, 49% de tous les répondants pensent que le TCS devrait être inclus sous une forme ou une autre dans l'évaluation des résidents en ORL. La comparaison de différents groupes dans le Tableau 6 nous démontre que les experts sont en faveur de l'implantation du TCS dans une plus grande proportion (74%) que les résidents (33%). L'opinion des résidents juniors est semblable à celle des résidents séniors à 30% par rapport à 38%. Finalement, l'opinion favorable à l'ajout du TCS dans les évaluations était plus présente dans le groupe qui connaissait déjà le TCS (68%) par rapport à ceux qui ne le connaissaient pas (40%).

Tableau 5. Proportion des participants qui recommanderaient d'utiliser le TCS lors d'évaluation en ORL-CCF.

		Recommandent le TCS (%)	N
Comparaison de l'opinion des résidents par rapport à celle des experts.	Résidents	33,3%	12
	Experts	73,9%	17
Comparaison de l'opinion des résidents juniors par rapport à celle des résidents séniors.	Résidents juniors (R2, R3)	30,0%	6
	Résidents séniors (R4, R5)	37,5%	6
Avez-vous déjà entendu parlé du test de concordance de script (TCS) ?	Non	39,5%	15
	Oui	68,4%	13

7. Discussion

7.1 L'acceptabilité du test de concordance de script

7.1.1 Les différences entre les résidents et les experts

Comme nous nous y attendions, notre étude a démontré une différence statistiquement significative entre l'opinion des résidents et celle des experts en ce qui a trait à la clarté des questions, au fait que les questions reflètent des situations cliniques réelles, qu'il est facile de s'adapter à ce type de questions et que le TCS est utile pour évaluer le raisonnement clinique. D'autres études avaient déjà recueilli l'opinion des résidents sur le TCS sans l'évaluer formellement. Dans l'étude de Kelly et al.⁴⁰ de 2012, les étudiants rapportaient que les questions de type TCS n'étaient pas plus difficiles que les QCM. Par contre, les résidents préféraient tout de même les QCM. En ORL, l'étude de Kania et al. de 2011¹ n'a pas démontré une différence d'appréciation sur le test de concordance de script entre le groupe des experts et des étudiants. Les faits rapportés dans ces deux études semblent différer de ce que nous avons recueilli. La différence majeure vient, à notre avis, du fait que dans leurs études, la participation se faisait dans le cadre d'une étude seulement, sans aucun lien avec une évaluation réelle des participants. Au contraire, dans notre étude les résidents ont répondu au questionnaire dans le contexte de leur examen formatif national annuel. Bien que les résidents n'étaient pas évalués sur les questions de type TCS, leur perception du TCS a pu être biaisée par le contexte dans lequel l'étude a été faite.

Dans notre cas, bien qu'il était clairement spécifié que les réponses TCS ne seraient pas analysées, nous croyons que d'avoir utilisé des questions incluses dans un examen officiel comme le NITE, bien que formatif, a pu influencer négativement l'opinion des résidents. Par contre, cette opinion se veut probablement plus près de la réalité car le TCS a été introduit dans un contexte bien réel. En effet, il y a toujours une crainte des résidents de se faire évaluer défavorablement par un test qu'ils ne maîtrisent pas bien ou qui est inhabituel. Les résidents peuvent alors avoir la crainte que leurs résultats ne soient pas un réel reflet de leurs

compétences et connaissances. Pour évaluer le raisonnement clinique, les résidents sont plutôt habitués à devoir répondre à des questions à choix multiples ou à des questions orales. Les QCM semblent plus faciles pour les résidents, car elles ne comprennent pas d'élément d'incertitude. Pour ce qui est des examens oraux, les résidents sont souvent formés au cours de leur résidence sur la façon de répondre à ce genre de questions. De plus, lors des examens oraux, les résidents ont l'opportunité d'expliquer leur réponse et leur démarche, en plus de pouvoir demander des explications supplémentaires si la question n'est pas claire pour eux. Par contre, les examens oraux sont beaucoup plus subjectifs qu'on peut le croire, car le résultat peut dépendre de l'attitude de la personne qui évalue et de la personne qui est évaluée.

Les résultats peuvent aussi en partie être expliqués par le manque de connaissances sur le test de concordance de script. En effet, lorsqu'on regarde les questions du post-test ayant reçu les plus faibles pourcentages d'appréciation par les résidents (Tableau 3), on peut supposer qu'une meilleure connaissance du format et des bases théoriques derrière le TCS aurait probablement pu améliorer l'opinion des résidents. En effet, seulement 30% des résidents ont trouvé facile de s'adapter au TCS et 44% trouvaient que les questions étaient claires. Une meilleure connaissance du format ou le fait d'avoir pu répondre à ce genre de question auparavant auraient probablement donné plus de confiance aux résidents pour y répondre. Par ailleurs, le fait que seulement 50% des résidents pensent que le TCS est utile pour tester la capacité de raisonnement clinique, malgré le fait que de très nombreuses études nous l'ont démontré, confirme le manque de connaissance sur le TCS.

Finalement, pour ce qui est d'inclure le TCS dans les évaluations des résidents en ORL-CCF, la majorité mentionne que cela devrait être ajouté comme évaluation formative au cours de la résidence. Effectivement, il serait probablement prématuré d'inclure un nouveau test dans l'évaluation de fin de formation des résidents sans l'avoir préalablement ajouté au cours du curriculum pour des évaluations formatives. Le TCS apporte plusieurs avantages lorsqu'utilisé comme évaluation formative. Il permet de faire un dépistage des résidents ayant des problèmes au niveau du raisonnement clinique et d'ainsi pouvoir utiliser des outils

pédagogiques pour les aider. En plus d'être utilisé comme outil d'évaluation, le TCS peut être utilisé comme outil d'apprentissage en fournissant, après l'examen, les réponses et explications des différents experts sur la résolution de chacune des questions.

7.1.1 Les différences entre les résidents juniors et séniors

Il a été intéressant de comparer l'opinion des résidents juniors par rapport à celle des résidents séniors (Tableau 4). Le groupe des résidents juniors comptait 20 participants et celui des résidents séniors 16, ce qui en fait des groupes comparables. Pour toutes les questions post-test sauf une, il n'y avait aucune différence statistiquement significative entre les deux groupes. La seule différence qui était significative était l'opinion sur le temps accordé pour répondre aux cas. À cette question, 75% des résidents de deuxième et troisième année s'entendaient pour dire que le temps accordé pour répondre aux cas était adéquat. Pour les résidents séniors, le temps accordé était adéquat dans seulement 38% des cas. Pour expliquer ce résultat, nous pouvons émettre comme hypothèse que les résidents séniors avaient l'impression qu'ils manquaient de temps parce qu'avec plus de connaissances théoriques, ils passaient plus de temps à réfléchir sur leurs nombreuses hypothèses et à peser la force de chacune de celles-ci. Ce résultat est donc intéressant mais il est difficile d'en tirer des conclusions autres que celle que les résidents séniors auraient aimé avoir plus de temps pour répondre aux questions que les résidents juniors.

Finalement, les résidents qui recommanderaient le TCS comme méthode d'évaluation en ORL-CCF étaient de 30% pour les résidents juniors et 37% pour les résidents séniors, donc qu'une très faible différence entre les 2 groupes. Nous ne sommes pas surpris de cette absence de différence puisque les réponses au sondage post-test étaient aussi semblables. Donc, comme les 2 groupes avaient des opinions semblables sur l'ensemble des éléments du test de concordance de script, il n'est pas surprenant qu'ils aient la même opinion sur le fait de l'utiliser comme méthode d'évaluation ou non.

7.2 L'importance du test de concordance de script

7.2.1 L'utilisation de multiples formes d'évaluation

L'utilisation de multiples formes d'évaluation au cours de la résidence permet d'évaluer différents aspects des connaissances et aptitudes nécessaires pour devenir un médecin compétent. Pour les débutants, il est acceptable d'évaluer la simple mémorisation de fait, mais ceci n'est pas suffisant, surtout lorsque les résidents avancent dans leur formation. Pour l'évaluation du raisonnement clinique, plusieurs outils d'évaluation sont disponibles et l'avantage de les combiner vient du fait qu'ils n'évaluent pas exactement la même chose. De plus, chaque format d'évaluation comporte des avantages et des inconvénients. Lorsque l'on combine plusieurs formes d'évaluation, nous nous assurons d'évaluer de façon juste et équitable tous les résidents, et de déceler ceux qui nécessiteraient plus d'aide et d'encadrement. Par exemple, le TCS permet, grâce à sa forme qui ne comprend que quelques questions sur de nombreux cas cliniques, une évaluation de plusieurs sujets, contrairement aux examens oraux qui souvent n'évaluent que quelques sujets avec de nombreuses questions. La théorie de spécificité de contenu¹⁰ explique que de mesurer la compétence à résoudre un problème portant sur un sujet spécifique ne prédit pas la performance aux autres problèmes. Ceci met en évidence que de questionner le résident sur seulement quelques sujets ne prédit pas sa capacité de raisonnement clinique dans d'autres situations, sur d'autres sujets. Dans ce contexte, il est préférable de ne poser que quelques questions par sujet sur une grande diversité de sujets, pour s'assurer d'avoir un reflet juste des compétences du résident. Le TCS, par sa forme, permet précisément de questionner sur une grande diversité de sujet et d'évaluer le raisonnement clinique dans ces différents contextes.

7.2.2 L'évaluation et l'enseignement du raisonnement clinique

Les connaissances de sciences de base ainsi que les connaissances cliniques sont essentielles et servent de fondation au raisonnement clinique. Par contre, ce n'est pas en lisant dans les livres ou en écoutant des conférences que l'étudiant perfectionne le raisonnement clinique⁴. Le raisonnement clinique s'acquière au fil du temps, à l'aide d'activités ciblées sur

la résolution de problème^{3,4} et d'activités cliniques supervisées suivies d'une rétroaction rapide et pertinente.

Certains auraient tendance à penser que le raisonnement clinique ne s'enseigne pas et qu'il est plutôt le résultat d'une acquisition au fil du temps et qu'il repose sur les traits de personnalité du candidat. Bien au contraire, le raisonnement clinique s'enseigne et doit être enseigné³. La qualité de la rétroaction est probablement ce qui est déterminant dans l'acquisition des apprentissages. De plus, l'apprenant devrait être encouragé à formuler des hypothèses précocement³ pour pouvoir recevoir du feedback sur ses hypothèses. Finalement, l'apprenant doit pouvoir être exposé à de multiples situations variées pour développer son intuition³ et construire ses scripts. Il faut donc, au cours de la résidence, évaluer la progression du raisonnement clinique des résidents pour s'assurer qu'ils évoluent normalement et pouvoir modifier, au besoin, l'enseignement qui est fait en conséquence.

7.2.3 L'effet sur l'apprentissage

« L'évaluation détermine les apprentissages. »¹⁰ Il est reconnu que les étudiants modifient leur façon d'étudier selon la manière dont ils seront évalués. Il est donc important de réaliser l'impact du choix des types d'évaluations au cours de la résidence. En ayant des évaluations qui évaluent spécifiquement le raisonnement clinique, les résidents ne peuvent pas qu'apprendre par cœur des données. Cela oblige l'apprenant à établir consciemment des liens entre les différents aspects d'un même problème, d'une même condition et de conditions différentes.

7.3 Les faiblesses de l'étude

La principale faiblesse de notre étude est le faible taux de participation (18% des résidents). Ce faible taux de participation s'explique par plusieurs facteurs. Premièrement, l'examen du NITE est un examen qui a une durée de trois heures. Le fait d'ajouter 30 minutes

de questions, à la fin de l'examen, a pu en décourager plusieurs à participer volontairement à cette évaluation. Aussi, la majorité des résidents ne connaissaient pas le format du TCS. Comme ce format de réponse est souvent considéré comme difficile d'adaptation, plusieurs ont possiblement commencé à y répondre pour ensuite se décourager. Finalement, bien qu'il était clair que le TCS ne faisait pas parti du NITE et que cela n'influencerait pas leurs résultats, certains résidents ont pu craindre que leurs réponses pourraient influencer leur note ou l'opinion que les correcteurs auraient d'eux. Même si l'examen NITE est un examen formatif, les résultats sont fournis aux directeurs de programme et les résidents y accordent beaucoup d'importance.

Une autre faiblesse de l'étude est l'absence de validation des questions de TCS. Bien que la validité du TCS n'était pas évalué dans cette étude et que les résultats au TCS n'ont pas été calculé, la validation des questions nous aurait permis d'être assuré que le format des questions était adéquat. En effet, dans la situation où les questions seraient trop faciles ou difficiles, ou même mal construites, le résultat au post-test pourrait être biaisé. Par contre, le fait que les questions aient été révisées par Dr Charlin, un expert du TCS qui est aussi oto-rhino-laryngologiste, nous apporte une certaine confiance que la construction et le niveau des questions étaient adéquats.

Malheureusement, aucun calcul de taille d'échantillon n'a été effectué avant le début de l'étude. Par conséquent, le recrutement des participants a dû s'effectuer en deux temps (soit en 2013 et 2016) pour s'assurer d'avoir un nombre suffisamment élevé de répondants. Un calcul à postériori d'une taille d'échantillon serait évidemment artificiel. Nous ne pouvons donc exclure que certains résultats pourraient être différents si une taille d'échantillon avait été calculé et qu'un recrutement d'un nombre différent de participants en avait résulté.

Finalement, il aurait bien évidemment été préférable de faire le recrutement des experts à un niveau national, tel que fait pour les résidents. Par contre, pour des raisons techniques et d'approbation éthique, nous avons choisi de n'inclure que le Québec. Nous croyons que la

pratique de l'ORL-CCF au Québec est représentative de la pratique au Canada en général. Par contre, nous sommes conscients que la démographie des experts, telle que décrite dans le Tableau 2, est plutôt représentative de la pratique à l'Université de Montréal. En effet, 75% des experts recrutés dans cette étude sont affiliés à l'Université de Montréal. Nous croyons que le fait que 2 des 3 chercheurs de l'étude y sont affiliés ait pu entraîner une motivation supplémentaire à participer à l'étude. Nous croyons tout de même que la pratique des différents spécialistes de l'Université de Montréal est similaire et comparable à la pratique des autres centres universitaires du Canada.

7.4 L'implication des résultats

Les résultats obtenus nous suggèrent que les médecins pratiquant en milieu universitaire seraient déjà beaucoup plus enclins à utiliser le test de concordance de script comme outils d'évaluation durant la résidence que les résidents. Pour pouvoir aller de l'avant avec l'ajout du TCS comme outil d'évaluation, il faudrait donc convaincre les résidents de son utilité, de sa capacité à tester le raisonnement clinique et de son objectivité.

Les résidents ont eu l'impression que les explications fournies, les exemples de cas et les cas étaient moins clairs comparativement aux experts. Les résidents ont donc eu beaucoup plus de difficulté à s'adapter à cette méthode d'évaluation que les experts. L'une des options envisagées pour améliorer l'acceptation du TCS auprès des résidents pourrait être d'améliorer les explications fournies au début du questionnaire en les rendant plus détaillées, en les rendant disponibles avant l'évaluation ou en les fournissant dans un format différent (power point, vidéo, présentation orale). Aussi, plus d'exemples pourraient être fournis aux résidents ou des cas tests pourraient être fait en groupe avec les résidents. Finalement, simplement le fait d'exposer les résidents plus souvent au TCS améliorerait la compréhension qu'ils ont du test et du fait même, son acceptabilité selon nous. En effet, depuis l'école primaire, ils sont exposés aux QCM, aux questions à développement et aux examens oraux. Il est donc normal que ces méthodes d'évaluations soient plus naturelles pour eux.

Dans le but d'améliorer l'acceptabilité du TCS et de familiariser les résidents avec son format, l'ajout d'examens avec des questions de type TCS devrait se faire graduellement. Aussi, la façon d'étudier pour pouvoir répondre à ce genre de question est différente et les résidents devraient avoir la chance de s'y habituer graduellement. Pour y arriver, l'ajout progressif de questions TCS à visée formative ou d'enseignement serait selon nous à privilégier. Dans ce contexte, s'assurer d'offrir une formation adéquate sur ce format de question est pour nous essentiel. Cela permettrait d'éviter plusieurs frustrations et incompréhensions en période d'examen. Une fois établi auprès des résidents et des médecins des milieux universitaires, il serait plus facile de l'utiliser comme outil d'évaluation sommative, en ajout aux formats actuels, pour permettre d'évaluer le raisonnement clinique des résidents en oto-rhino-laryngologie.

8. Conclusion

Il n'est plus à démontrer que le raisonnement clinique est une composante essentielle des compétences que doit acquérir un résident pour devenir un médecin compétent. Cette acquisition se fait de façon progressive au cours des études médicales et de la résidence. Par contre, cette acquisition ne doit pas se faire de façon passive mais plutôt via un effort conscient d'enseignement et d'évaluation du raisonnement clinique. Pour l'évaluer, plusieurs outils sont disponibles, dont le test de concordance de script qui est décrit dans la littérature depuis la fin des années 90.

Des dizaines d'études ont rapporté une validité adéquate du TCS dans différents domaines et à différents niveaux de formation pour sa capacité à évaluer le raisonnement clinique. Présentement, peu de programmes de formation médicale ou de résidence utilisent le TCS comme outil d'évaluation, tant formatif que sommatif. Le but de cette étude était donc d'évaluer l'acceptabilité du test de concordance de script auprès des résidents et des médecins en pratique universitaire. Pour ce faire, cette étude s'est faite dans les programmes canadiens d'oto-rhino-laryngologie du Canada.

Notre étude suggère que présentement, les résidents et experts en ORL-CCF en savent peu sur le TCS. L'acceptabilité auprès des résidents s'est avérée plus faible qu'auprès des experts. En effet, les résidents qui ont participé à l'étude ont trouvé plus difficile de s'adapter au TCS et n'étaient pas convaincus que le TCS représentait des situations cliniques réelles. De plus, les résidents participants étaient réticents à rajouter le TCS comme méthode d'évaluation au cours de la résidence.

Ces résultats nous suggèrent bien que l'acceptabilité du TCS, surtout auprès des résidents est encore faible. Ces résultats reflètent probablement le manque de connaissance et d'expérience qu'ont les résidents avec l'outil qu'est le TCS. Cependant, l'acceptabilité auprès

des experts nous laisse penser que ces derniers pourraient participer volontairement à l'introduction de ce type d'évaluation. Néanmoins, nous pensons que l'ajout du TCS à des fins de formation, et non d'évaluation, pourrait être une solution utile pour familiariser les résidents avec le TCS, tout en ayant l'avantage de donner un outil de formation de plus aux résidents. L'introduction du TCS en évaluation formative permettrait une transition adéquate vers son implantation à des fins d'évaluation sommative. Les études en cours et futures sur le TCS comme outil de formation plutôt que comme outil d'évaluation pourra nous en dire plus sur les avantages qu'il apporte dans ce domaine.

Bibliographie

1. Kania RE, Verillaud B, Tran H, et al. Online script concordance test for clinical reasoning assessment in otorhinolaryngology: the association between performance and clinical experience. *Archives of otolaryngology--head & neck surgery*. 2011;137(8):751-755.
2. Fisher B, Rourke L. Developing, Maintaining, and Teaching Clinical Diagnostic Expertise. *Canadian Journal of General Internal Medicine*. 2016;11(1):19-24.
3. Pelaccia T, Tardif J, Tribby E, Charlin B. An analysis of clinical reasoning through a recent and comprehensive approach: the dual-process theory. *Medical Education Online*. 2011(16).
4. Nendaz M, Charlin B, LeBlanc V, Bordage G. Le raisonnement clinique: données issues de la recherche et implications pour l'enseignement. *Pédagogie Médicale*. 2005;6(4):235-254.
5. Leclercq D. Une taxonomie des Objectifs Instrumentée au Service de l'Évaluation (TOISE). In: Liège EdIUd, ed. *Édumétrie et docimologie pour praticiens chercheurs*2005:1-34.
6. Pelaccia T, Tardif J, Tribby E, Ammirati C, Bertrand C, Charlin B. How do physicians make diagnoses and decisions? Key issues in emergency medicine. *Ann Fr Med Urgences*. 2011;1:77-84.
7. Norman GR, Brook LR, Colle CL, Hatala RM. The Benefit of Diagnostic Hypotheses in Clinical Reasoning: Experimental Study of an Instructional Intervention for Forward and Backward Reasoning. *Cognition and Instruction*. 1999;17(4):433-448.
8. Charlin B. Assessing the uncertainty component of clinical reasoning. *Pédagogie Médicale*. 2006;7(1):5-6.
9. Schuwirth LW VM, van der Vleuten CP, Boshuizen HP, Dinant GJ. Do short cases elicit different thinking processes than factual knowledge questions do? *Med Educ*. 2001;35(4):348-356.
10. Charlin B, Bordage G, Van der Vleuten C. L'évaluation du raisonnement clinique. *Pédagogie Médicale*. 2003;4(1):42-52.
11. Hubbard JP. *Measuring medical education*. Philadelphia: Lea and Febiger; 1978.
12. Vaughan VC. The Patient Management Problem as an Evaluative Instrument. *Pediatrics in Review*. 1979;1(3):67-76.
13. Lignes directrices pour l'élaboration de problèmes à éléments-clés et dénoncés de cas. Le conseil médical du Canada; 2012:9.
14. Zayyan M. Objective structured clinical examination: the assessment of choice. *Oman Med J*. 2011;26(4):219-222.
15. Morag E, Lieberman G, Volkan K, Shaffer K, Novelline R, Lang EV. Clinical Competence assessment in radiology: introduction of an objective structured clinical examination in the medical school curriculum. *Academic radiology*. 2001;8(1):74-81.
16. Peterson R. Examining the examination: canadian versus US certification exam. *Can Assoc Radiol*. 2000;51(2):209.
17. Custers EJ. Thirty years of illness scripts: Theoretical origins and practical applications. *Medical teacher*. 2015;37(5):457-462.

18. Lubarsky S, Dory V, Duggan P, Gagnon R, Charlin B. Script concordance testing: From theory to practice: AMEE Guide No. 75. *Medical teacher*. 2013;35(3):184-193.
19. Charlin B, Brailovsky C, Leduc C, Blouini D. The Diagnosis Script Questionnaire: A New Tool to Assess a Specific Dimension of Clinical Competence. *Advances in Health Sciences Educations*. 1998;3:51-58.
20. Gagnon R, Charlin B, Lambert C, Carriere B, Van der Vleuten C. Script concordance testing: more cases or more questions? *Advances in health sciences education : theory and practice*. 2009;14(3):367-375.
21. Charlin B. Reasoning in the context of uncertainty. *Academic Medecine*. 2005.
22. Bland AC, Kreiter CD, Gordon JA. The psychometric properties of five scoring methods applied to the script concordance test. *Academic medicine : journal of the Association of American Medical Colleges*. 2005;80(4):395-399.
23. Wilson AB, Pike GR, Humbert AJ. Analyzing script concordance test scoring methods and items by difficulty and type. *Teaching and learning in medicine*. 2014;26(2):135-145.
24. Gagnon R, Charlin B, Coletti M, Sauve E, van der Vleuten C. Assessment in the context of uncertainty: how many members are needed on the panel of reference of a script concordance test? *Medical education*. 2005;39(3):284-291.
25. Lurie S. Towards greater clarity in the role of ambiguity in clinical reasoning. *Medical education*. 2011(45):326-328.
26. Charlin B, Gagnon R, Sauve E, Coletti M. Composition of the panel of reference for concordance tests: do teaching functions have an impact on examinees' ranks and absolute scores? *Medical teacher*. 2007;29(1):49-53.
27. Petrucci AM, Nouh T, Boutros M, Gagnon R, Meterissian SH. Assessing clinical judgment using the Script Concordance test: the importance of using specialty-specific experts to develop the scoring key. *American journal of surgery*. 2013;205(2):137-140.
28. Meterissian S, Zabolotny B, Gagnon R, Charlin B. Is the script concordance test a valid instrument for assessment of intraoperative decision-making skills? *American journal of surgery*. 2007;193(2):248-251.
29. Gagnon R, Lubarsky S, Lambert C, Charlin B. Optimization of answer keys for script concordance testing: should we exclude deviant panelists, deviant responses, or neither? *Advances in health sciences education : theory and practice*. 2011;16(5):601-608.
30. Charlin B, Desaulniers M, Gagnon R, Blouin D, Van der Vleuten C. Comparison of an Aggregate Scoring Method With a Consensus Scoring Method in a Measure of Clinical Reasoning Capacity. *Teaching and learning in medicine*. 202;14(3):150-156.
31. Carriere B, Gagnon R, Charlin B, Downing S, Bordage G. Assessing clinical reasoning in pediatric emergency medicine: validity evidence for a Script Concordance Test. *Annals of emergency medicine*. 2009;53(5):647-652.
32. Sibert L, Charlin B, Gagnon R, Corcos J, Khalaf A, Grise P. [Evaluation of clinical reasoning in urology: contribution of the Script Concordance Test]. *Progres en urologie : journal de l'Association francaise d'urologie et de la Societe francaise d'urologie*. 2001;11(6):1213-1219.
33. Fournier JP, Thiercelin D, Pulcini C, et al. Évaluation du raisonnement clinique en médecine d'urgence: les tests de concordance des scripts décèlent mieux l'expérience

- clinique que les questions à choix multiples à contexte riche. *Pédagogie Médicale*. 2006;7(1):20-30.
34. Iravani K, Amini M, Doostkam A, Dehbozorgian M. The validity and reliability of script concordance test in otolaryngology residency training. *J Adv Med Educ Prof*. 2016;4(2):93-96.
 35. Faucher C, Dufour-Guindon MP, Lapointe G, Gagnon R, Charlin B. Assessing clinical reasoning in optometry using the script concordance test. *Clin Exp Optom*. 2016;99(3):280-286.
 36. Cote S, Tanguay A, Gagnon R, Charlin B, Michaud C. Élaboration et validation d'un test de concordance de script pour évaluer le raisonnement clinique des infirmières en contexte de détérioration de la condition clinique. *Pédagogie Médicale*. 2014;15(1):7-20.
 37. Humbert AJ, Johnson MT, Miech E, Friedberg F, Grackin JA, Seidman PA. Assessment of clinical reasoning: A Script Concordance test designed for pre-clinical medical students. *Medical teacher*. 2011;33(6):472-477.
 38. Caire F, Sol JC, Moreau JJ, Isidori P, Charlin B. [Self-assessment for neurosurgery residents by script concordance test (SCT). The process of test elaboration]. *Neuro-Chirurgie*. 2004;50(1):66-72.
 39. Brailovsky C, Charlin B, Beausoleil S, Cote S, Van der Vleuten C. Measurement of clinical reflective capacity early in training as a predictor of clinical reasoning performance at the end of residency: an experimental study on the script concordance test. *Medical education*. 2001;35(5):430-436.
 40. Kelly W, Durning S, Denton G. Comparing a script concordance examination to a multiple-choice examination on a core internal medicine clerkship. *Teaching and learning in medicine*. 2012;24(3):187-193.
 41. Lineberry M, Kreiter CD, Bordage G. Threats to validity in the use and interpretation of script concordance test scores. *Medical education*. 2013;47(12):1175-1183.
 42. Lineberry M. Missing the mark: the faulty logic of aggregate scoring in script concordance tests. *Medical education*. 2014;48(11):1038-1040.
 43. Lineberry M, Kreiter CD, Bordage G. Script concordance tests: strong inferences about examinees require stronger evidence. *Medical education*. 2014;48(4):452-453.
 44. See KC, Tan KL, Lim TK. The script concordance test for clinical reasoning: re-examining its utility and potential weakness. *Medical education*. 2014;48(11):1069-1077.
 45. Fournier JP, Demeester A, Charlin B. Script concordance tests: guidelines for construction. *BMC medical informatics and decision making*. 2008;8:18.

Annexe 1. Exemple d'un questionnaire de type *patient management problem*

Madame Roy est une dame de 58 ans, sans antécédent particulier, qui consulte pour des vertiges. Ses vertiges ont débuté il y a 2 jours, sont constants et ne sont pas associés à un mouvement en particulier. Elle n'a pas noté de baisse d'audition ni d'autre symptôme accompagnateur.

Question 1. Quel est votre diagnostic différentiel?

Question 2. Comment allez-vous établir le diagnostic?

Question 3. Quelle est la cause la plus probable de son vertige?

Question 4. Quels sont les éléments les plus importants dans la prise en charge d'un vertige?

Annexe 2. Exemple d'un questionnaire par éléments clés

Exemple tiré des lignes directrices pour l'élaboration de problèmes à éléments-clés et d'énoncés de cas du conseil médical du Canada¹¹.

Un homme âgé de 36 ans est transporté à l'urgence en ambulance après être tombé sur le trottoir et avoir perdu connaissance alors qu'il attendait l'autobus. Le témoin, qui a immédiatement appelé l'ambulance, a signalé à l'équipe d'ambulanciers qu'avant de tomber, le patient semblait confus et agité et se disputait avec une personne invisible. Après la chute, le corps de l'homme a été saisi de spasmes durant un court moment, et son visage est devenu bleu. Par la suite, les spasmes se sont transformés en mouvements saccadés parcourant tout le corps pendant environ une minute. L'homme n'a pas repris connaissance après cet épisode. Pendant les 10 minutes du déplacement en ambulance, il a eu deux autres épisodes similaires, sans reprendre connaissance, suivis d'un troisième dont vous avez été témoin à son arrivée à l'urgence.

Sa température est de 37,8 °C. Son apparence est négligée, et il n'a toujours pas repris connaissance. Aucun membre de la famille ou ami du patient ne l'a accompagné à l'hôpital.

Question 1

Quel est votre principal diagnostic ou quels sont vos principaux diagnostics à ce stade-ci? Vous pouvez en indiquer jusqu'à deux.

1. _____
2. _____

Élément-clé

Dans le cas d'un adulte transporté à l'urgence en raison de crises d'épilepsie multiples et qui n'a pas repris connaissance, le candidat devra poser un diagnostic provisoire d'état de mal épileptique

Clés de notation

Note : Critères

1,0 : État de mal épileptique

0 : Plus de 2 réponses

Question 2

Quelle est votre prise en charge à ce stade-ci (soyez précis)? Énumérez autant de mesures que vous le jugez approprié.

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____
7. _____
8. _____
9. _____
10. _____

Éléments-clés

Dans le cas d'un adulte transporté à l'urgence en raison de crises d'épilepsie multiples et qui n'a pas repris connaissance, le candidat devra instaurer le traitement initial, y compris l'administration d'un complexe vitaminique B (thiamine) par voie intraveineuse, d'un bolus de glucose hypertonique (dextrose à 50 %), de lorazépam ou de diazépam ou de clonazépam ou de phénobarbital ET de phénytoïne.

Clés de notation

Note : Critères

0,25 : Thiamine par voie intraveineuse (doit mentionner « par voie intraveineuse »)

0,25: Glucose hypertonique par voie intraveineuse (doit mentionner « par voie intraveineuse »)

0,5: Lorazépam par voie intraveineuse (doit mentionner « par voie intraveineuse »)

ET Phénytoïne par voie intraveineuse (doit mentionner « par voie intraveineuse »)

0 : Plus de 7 réponses

Question 3

Dix minutes après son arrivée à l'urgence, le patient n'a toujours pas repris connaissance. L'infirmière a trouvé un numéro de téléphone dans son portefeuille. Quelles questions sur l'état du patient poseriez-vous à la personne qui répondra au téléphone, dans l'hypothèse où celle-ci connaît le patient? Sélectionnez jusqu'à six réponses ou sélectionnez le no 33 si vous pensez qu'il n'y a pas lieu de téléphoner à ce moment précis.

1. Douleur abdominale
2. Antécédents de consommation d'alcool
3. Antécédents de maux de dos
4. Antécédents de cancer
5. Abus de cocaïne
6. Antécédents de pontage coronarien
7. Antécédents de diabète
8. Diarrhée
9. Étourdissements
10. Allergies médicamenteuses
11. Antécédents familiaux
12. Allergies alimentaires
13. Maux de tête
14. Déficience auditive
15. Abus d'héroïne
16. Douleurs articulaires
17. Infection pulmonaire
18. Antécédents pharmaceutiques
19. Affection musculaire
20. Nausée
21. Antécédents de palpitations
22. Présence d'un animal de compagnie à la maison
23. Antécédents d'un problème similaire
24. Profession
25. Antécédents sexuels
26. Antécédents de tabagisme
27. Difficultés d'intégration sociale
28. Intervention chirurgicale
29. Antécédents en matière de voyages
30. Infection virale
31. Déficience visuelle
32. Vomissements
33. Il n'est pas pertinent de téléphoner à ce moment précis

Élément-clé

Dans le cas d'un adulte transporté à l'urgence en raison de crises d'épilepsie multiples et qui n'a pas repris connaissance, le candidat recueillera l'anamnèse pour déterminer les causes possibles des crises d'épilepsie (y compris la consommation d'alcool, la prise de médicaments, la consommation de drogues et les antécédents de diabète).

Clés de notation

Note : Critères

0,25 : 2

0,25 : 5 ou 15

0,25 : 7

0,25 : 18

0 : 33

0 : Plus de 6 réponses

Annexe 3. Liste des questions de type TCS utilisées dans l'examen NITE de 2013

Test de concordance de script

Comment répondre aux questions de type TCS

Chaque question comporte trois parties : 1) une vignette clinique, 2) une hypothèse diagnostique en lien avec la vignette, et 3) une nouvelle information qui pourrait ou non changer la probabilité de l'hypothèse diagnostique. Parfois, la deuxième partie pourrait être une investigation ou une option de traitement au lieu d'une hypothèse diagnostique. Plus d'une question se rapporte à chaque vignette clinique. Chaque question doit être considérée indépendamment (ne tenez pas compte de l'hypothèse ou option proposée dans les questions précédentes). Voici un exemple :

Vous devez répondre aux questions suivantes concernant la probabilité de l'intervention diagnostique en utilisant l'échelle suivante :

-2	-1	0	+1	+2
Beaucoup moins pertinent	Moins pertinent	Ne change pas le plan de traitement	Plus pertinent	Beaucoup plus pertinent

1- Un adolescent de 15 ans souffre de douleur à la gorge depuis 5 jours, pire depuis 3 jours. C'est le premier épisode du genre. Vous lui faites une culture de gorge qui revient positive pour le Strep groupe A. Vous apprenez que cet adolescent a aussi un frère. Vous vous demandez si vous devriez lui faire aussi une culture de gorge.

Si vous pensez à :	Mais vous apprenez que :	L'intervention diagnostique devient :				
a. Faire une culture de gorge à son frère.	Son frère a 4 ans et n'a éprouvé aucun symptômes.	-2	-1	0	+1	+2
b. Ne pas faire de culture de gorge	Son frère a 2 ans et est symptomatique.	-2	-1	0	+1	+2

À la question **1a**, si vous croyez que le fait que son frère ait 4 ans et soit asymptomatique rend l'intervention diagnostique plus pertinente, vous devez encercler « +1 ». Si vous croyez que cela rend le l'intervention diagnostique **beaucoup moins** pertinente, vous encercler « -2 ». Si vous croyez que cette information ne change rien à la pertinence de l'intervention diagnostique, vous encercler le 0.

À la question **1b**, ne considérez pas l'information contenue dans la question **1a** concernant l'intervention diagnostique de faire une culture de gorge et l'information que son frère de 4 ans est asymptomatique.

Vous aurez 20 minutes pour compléter le texte. Après avoir répondu aux questions de type TCS, veuillez s'il vous plait remplir le questionnaire d'appréciation.

Test de concordance de script

Questions en ORL

Vous devez répondre aux questions suivantes concernant la probabilité du diagnostic en utilisant l'échelle suivante

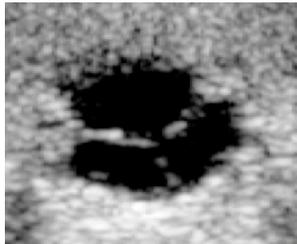
-2	-1	0	+1	+2
Beaucoup moins probable	Moins probable	Ne change pas le diagnostic	Plus probable	Beaucoup plus probable

1- Une fillette de 1 an est référée à votre clinique pour toux sèche chronique par son médecin de famille. Elle ne présente aucun symptôme infectieux présentement, ni dans le dernier mois.

Si vous pensez à :	Et vous apprenez que :	Le diagnostic devient				
a. Reflux gastro oesophagien	Elle prend Ranitidine (8mg/kg/jour) depuis 8 semaines et il n'y a aucune amélioration.	-2	-1	0	+1	+2
		<input type="checkbox"/>				
b. Écoulement nasal postérieur	Elle est capable de respirer par le nez et il n'y a pas de rhinorrhée.	-2	-1	0	+1	+2
		<input type="checkbox"/>				
c. Fibrose kystique	Elle tousse seulement la nuit.	-2	-1	0	+1	+2
		<input type="checkbox"/>				

2- Une pré-adolescente de 12 ans consulte pour une masse non douloureuse au niveau du cou en latéral gauche (niveau 2) qu'elle a remarqué il y a 2 mois. La masse grossit lentement. Elle aurait eu un IVRS il y a environ 3 mois, d'une durée de 6 jours. À l'examen physique, la masse mesure 3 cm de diamètre, elle est légèrement ferme et plus ou moins mobile et non érythémateuse.

Si vous pensez à :	Et vous apprenez que :	Le diagnostic devient				
a. Rhabdomyosarcome	La jeune fille se sent plus faible et a une diminution de l'appétit depuis quelques semaines	-2	-1	0	+1	+2
		<input type="checkbox"/>				
b. Lymphome	Vous n'avez palpé aucune autre adénopathie lors de l'examen physique.	-2	-1	0	+1	+2
		<input type="checkbox"/>				
c. Kyste branchial	Voici l'image à l'échographie	-2	-1	0	+1	+2
		<input type="checkbox"/>				



-2	-1	0	+1	+2
Beaucoup moins probable	Moins probable	Ne change pas le diagnostic	Plus probable	Beaucoup plus probable

3- Une femme de 48 ans vous consulte en clinique pour une congestion nasale intermittente depuis plusieurs années. À l'examen physique, elle a tout au plus une légère déviation de son septum à droite. À la laryngoscopie flexible, vous découvrez une masse au niveau du nasopharynx qui l'obstrue partiellement.

Si vous pensez à :	Et vous apprenez que :	Le diagnostic devient				
a. Carcinome à cellules squameuses du nasopharynx	Elle est née en Chine, mais vit au Canada depuis 18 ans.	-2	-1	0	+1	+2
		<input type="checkbox"/>				
b. Lymphome	Elle a une atteinte ipsilatérale du nerf trijumeau à l'examen physique.	-2	-1	0	+1	+2
		<input type="checkbox"/>				
c. Kyste de Tornwaldt	La masse semble légèrement translucide et un peu déviée vers la gauche.	-2	-1	0	+1	+2
		<input type="checkbox"/>				

4- Vous voyez un homme de 66 ans à votre clinique pour la première fois. Ce patient souffre d'un troisième épisode de vertiges rotatoires intermittent depuis les 18 derniers mois. Les deux premiers épisodes ont durés plusieurs minutes même jusqu'à quelques heures. Il a eu un test d'audition normal il y a 2 ans, avant le début de ses symptômes. Son examen neurologique est normal (absence de nystagmus, nerfs crâniens normaux et fonctions cérébelleuses normales)

Si vous pensez à :	Et vous apprenez que :	Le diagnostic devient				
a. Labyrinthite	Le patient a des céphalées intermittentes et a été diagnostiqué avec des migraines il y a 8 mois.	-2	-1	0	+1	+2
		<input type="checkbox"/>				
b. Maladie de Ménière	Il refait un audiogramme qui est toujours normal.	-2	-1	0	+1	+2
		<input type="checkbox"/>				
c. Maladie de Ménière	Il est fumeur et hypertendu. Lors des 2 derniers épisodes, il a ressenti une légère faiblesse au bras droit, qui est complètement résolue.	-2	-1	0	+1	+2
		<input type="checkbox"/>				
d. Neuronite vestibulaire	Il a remarqué que son audition a légèrement diminuée dans les deux dernières années, mais il pense que cela est seulement relié à son âge.	-2	-1	0	+1	+2
		<input type="checkbox"/>				

-2	-1	0	+1	+2
Beaucoup moins probable	Moins probable	Ne change pas le diagnostic	Plus probable	Beaucoup plus probable

5- Un collègue omnipraticien vous envoie la photo de la bouche d'une patiente de 55 ans qu'il a vu en consultation ce jour. Les taches ne sont pas surélevées et sont brunes-noires. Elle est asymptomatique. Elle a remarqué l'apparition de ces lésions dans la dernière année mais elle ne peut dire quand exactement.



Si vous pensez à :

a. Macule mélanotique

b. Syndrome de Peutz-Jehgers

c. Mélanome

Et vous apprenez que :

Elle est en investigation pour un possible cancer de l'ovaire droit.

Elle a d'autres taches similaires au niveau des mains.

Elle a une autre lésion similaire à l'intérieur de la joue à droite.

Le diagnostic devient

-2	-1	0	+1	+2
<input type="checkbox"/>				

-2	-1	0	+1	+2
<input type="checkbox"/>				

-2	-1	0	+1	+2
<input type="checkbox"/>				

Vous devez répondre aux questions suivantes concernant la probabilité de l'intervention diagnostique ou d'un plan de traitement en utilisant l'échelle suivante:

-2	-1	0	+1	+2
Beaucoup moins pertinente	Moins pertinente	N'a pas d'influence	Plus pertinent	Beaucoup plus pertinent

1- Vous êtes demandé en consultation à l'urgence pour une femme de 48 ans qui a reçu un coup de couteau lors d'une altercation. La plaie semble profonde et mesure 7 cm de long, dans la zone II du cou à droite. Ses signes vitaux sont stables pour le moment, il n'y a pas de saignement actif et pas d'Évidence de pneumothorax.

Si vous décidez de:

Et vous apprenez que :

Votre intervention devient :

a. De l'observer

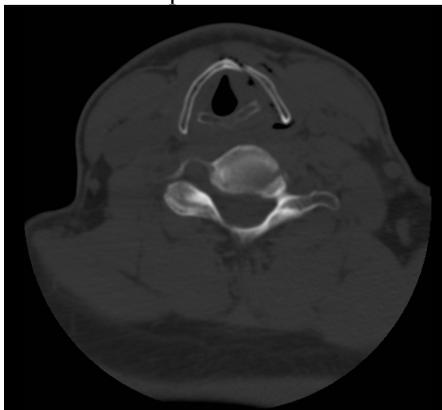
De l'emphysème sous-cutané est présent de façon modérée au niveau du cou.

-2	-1	0	+1	+2
<input type="checkbox"/>				

b. De prévoir une salle d'opération.

Elle a eu une laryngoscopie normale. Voici une coupe du scan :

-2	-1	0	+1	+2
<input type="checkbox"/>				



c. De l'observer

Elle n'a pas de fracture au scan mais se plaint d'une douleur importante à la déglutition.

-2	-1	0	+1	+2
<input type="checkbox"/>				

-2	-1	0	+1	+2
Beaucoup moins pertinente	Moins pertinente	N'a pas d'influence	Plus pertinent	Beaucoup plus pertinent

2- Vous évaluez une patiente de 52 ans, sans antécédents significatifs, pour un nodule thyroïdien gauche de 1.5 cm dont la cytoponction a démontré la présence de cellules suspectes de carcinome papillaire. Vous demandez une échographie cervicale afin d'évaluer les chaînes ganglionnaires.

Si vous pensez à :	Et vous apprenez que :	Votre intervention devient :				
		-2	-1	0	+1	+2
a. Faire un évidement cervical du compartiment central	A l'échographie on note la présence de 3 ganglions dans la zone 6 ipsilatérale (de 4 à 13mm), avec des portions hypoéchogènes, des zones punctiformes hyperéchogènes et une vascularisation périphérique augmentée.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b. Faire un évidement cervical du compartiment central et latéral.	A l'échographie on note la présence d'un ganglion de 1.2cm dans la zone 4 ipsilatérale, avec une vascularisation périphérique augmentée. La zone 6 est normale.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c. Faire un évidement cervical du compartiment central et latéral.	La cytoponction du ganglion de la zone 4 est compatible avec la présence de carcinome papillaire.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d. Faire un évidement cervical du compartiment latéral seulement.	La cytoponction du ganglion de la zone 4 est compatible avec la présence de carcinome papillaire.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e. Faire un évidement cervical des compartiments central et latéral gauche.	A l'échographie on note la présence d'un ganglion suspect de 1.5 cm dans la zone 6 dont la cytoponction est compatible avec un carcinome papillaire. Le compartiment latéral est normal.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

-2	-1	0	+1	+2
Beaucoup moins pertinente	Moins pertinente	N'a pas d'influence	Plus pertinent	Beaucoup plus pertinent

3- Une fille de 17 ans vous est référée par son médecin de famille pour « amygdalite » récurrente. À l'examen physique, les amygdales sont de grade III. La patiente vous apprend que, dans la dernière année, elle a fait 8 amygdalites, toutes diagnostiquées par le même médecins

Si vous pensez à :	Et vous apprenez que :	Le plan de traitement devient :										
a. D'offrir une amygdalectomie.	Elle a eu au moins 3 autres épisodes d'amygdalite pour lesquels elle n'a pas consulté son médecin de famille.	<table border="1"> <thead> <tr> <th>-2</th> <th>-1</th> <th>0</th> <th>+1</th> <th>+2</th> </tr> </thead> <tbody> <tr> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> </tr> </tbody> </table>	-2	-1	0	+1	+2	<input type="checkbox"/>				
-2	-1	0	+1	+2								
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>								
b. D'offrir une amygdalectomie.	Lorsqu'elle n'a pas d'infection, son amygdale droite est clairement plus grosse que la gauche.	<table border="1"> <thead> <tr> <th>-2</th> <th>-1</th> <th>0</th> <th>+1</th> <th>+2</th> </tr> </thead> <tbody> <tr> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> </tr> </tbody> </table>	-2	-1	0	+1	+2	<input type="checkbox"/>				
-2	-1	0	+1	+2								
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>								
c. D'offrir une amygdalectomie.	Elle ronfle et a des symptômes d'apnée du sommeil, qui s'aggravent tous lors de ses épisodes d'amygdalites.	<table border="1"> <thead> <tr> <th>-2</th> <th>-1</th> <th>0</th> <th>+1</th> <th>+2</th> </tr> </thead> <tbody> <tr> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> </tr> </tbody> </table>	-2	-1	0	+1	+2	<input type="checkbox"/>				
-2	-1	0	+1	+2								
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>								
d. D'offrir une amygdalectomie.	Son médecin de famille lui a parlé d'infection de gorge mais n'a jamais utilisé le terme amygdalite. Elle n'a jamais eu de culture de gorge pour streptocoque.	<table border="1"> <thead> <tr> <th>-2</th> <th>-1</th> <th>0</th> <th>+1</th> <th>+2</th> </tr> </thead> <tbody> <tr> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> </tr> </tbody> </table>	-2	-1	0	+1	+2	<input type="checkbox"/>				
-2	-1	0	+1	+2								
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>								

Annexe 4. Sondage post-test des résidents

Questionnaire sur le test de concordance de script - Répondant

Niveau de formation: R1 R2 R3 R4 R5

1- Aviez-vous déjà entendu parlé du test de concordance de script (TCS) ? *Oui* *Non* *Incertain*

Si oui, vous diriez que votre niveau de connaissance sur le test de concordance de script est :

- Faible
- Modéré
- Important

2 – Aviez-vous déjà répondu à un test comprenant des questions de type TCS avant? _____

Si oui, a) à quel moment durant votre formation _____

b) ans quel but?

- Examen final
- Test formatif
- Recherche
- Autre _____

3- Pour les questions suivantes, choisissez la réponse qui correspond le mieux à votre opinion.

	<i>Fortement en désaccord (1)</i>	<i>En désaccord (2)</i>	<i>Neutre (3)</i>	<i>En accord (4)</i>	<i>Fortement en accord (5)</i>
• Les explications fournies au début du questionnaire étaient claires.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Les exemples de cas (avec réponses) présentés avant le questionnaire étaient utiles.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Les questions se rapportant aux cas étaient claires.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Il était facile de s'adapter à cette méthode d'évaluation.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Le temps accordé pour répondre aux cas était adéquat.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Les questions reflétaient des situations cliniques réelles.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	<i>Fortement en désaccord (1)</i>	<i>En désaccord (2)</i>	<i>Neutre (3)</i>	<i>En accord (4)</i>	<i>Fortement en accord (5)</i>
• Les questions reflètent des situations que j'ai déjà rencontré ou que je pourrai rencontrer durant ma résidence ou dans ma pratique future.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Cette évaluation est utile pour tester la capacité de raisonnement clinique en ORL.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

4- Qu'avez-vous aimé le plus en répondant au questionnaire de TCS?

5- Qu'avez-vous aimé le moins en répondant au questionnaire de TCS?

6- Recommanderiez vous d'utiliser le TCS lors d'évaluations en ORL?

Oui *Non*

Et pourquoi? _____

7- Autres commentaires?

Annexe 5. Sondage post-test des experts

Questionnaire sur le test de concordance de script - membres du groupe de référence

1- Aviez-vous déjà entendu parler du test de concordance de script (TCS) ?

Oui *Non* *Incertain*

Si oui, vous diriez que votre niveau de connaissance sur le test de concordance de script est :

- Faible
- Modéré
- Important

2 – Aviez-vous déjà répondu à un test comprenant des questions de type TCS avant? _____

Si oui, dans quel contexte?

- Examen final
- Test formatif
- Recherche
- Autre

3- Pour les questions suivantes, choisissez la réponse qui correspond le mieux à votre opinion.

	<i>Fortement en désaccord</i> (1)	<i>En désaccord</i> (2)	<i>Neutre</i> (3)	<i>En accord</i> (4)	<i>Fortement en accord</i> (5)
• Les explications fournies au début du questionnaire étaient claires.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• L'exemple de cas (avec réponse) présenté avant le questionnaire était utile.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Les questions se rapportant aux cas étaient claires.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Il était facile de s'adapter à cette méthode d'évaluation.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Les questions reflétaient des situations cliniques réelles.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	<i>Fortement en désaccord (1)</i>	<i>En désaccord (2)</i>	<i>Neutre (3)</i>	<i>En accord (4)</i>	<i>Fortement en accord (5)</i>
• Les questions reflètent des situations que les résidents pourraient avoir déjà rencontré ou rencontrer dans leur pratique future.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Cette évaluation est utile pour tester la capacité de raisonnement clinique en ORL.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

4- Qu'avez-vous aimé le plus en répondant au questionnaire de TCS?

5- Qu'avez-vous aimé le moins en répondant au questionnaire de TCS?

6- Recommanderiez vous d'utiliser le TCS lors d'évaluations en ORL?

Oui *Non*

Et pourquoi? _____

7- Autres commentaires?
