

Assessing the acceptability of script concordance testing: a nationwide study in otolaryngology

Andrée-Anne Leclerc, MD, MSc
Lily H.P. Nguyen, MD CM, MSc
Bernard Charlin, MD, PhD
Stuart Lubarsky, MD, MHPE
Tareck Ayad, MD

Presented at the Annual Meeting of the Quebec Association of Otolaryngology and Head and Neck Surgery, Oct. 16–18, 2015, Québec, Que.; at the 69th Annual Meeting of the Canadian Society of Otolaryngology – Head and Neck Surgery, June 6–9, 2015, Winnipeg, Man.; and at the 2nd International Conference on Clinical Reasoning, Oct. 28–31, 2014, Montréal, Que.

Accepted July 28, 2020

Correspondence to:

A.-A. Leclerc
Department of Otolaryngology
Hôpital Maisonneuve-Rosemont
5415 Assumption Blvd
Montréal QC H1T 2M4
andree-anne.leclerc.1@umontreal.ca

DOI: 10.1503/cjs.014919

Background: Script concordance testing (SCT) is an objective method to evaluate clinical reasoning that 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 about its acceptability to major stakeholders. The aim of this study was to determine the acceptability of SCT to residents in otolaryngology – head and neck surgery (OTL-HNS) and a reference group of experts.

Methods: In 2013 and 2016, a set of SCT questions, as well a post-test exit survey, were included in the National In-Training Examination (NITE) for OTL-HNS. This examination is administered to all OTL-HNS residents across Canada who are in the second to fifth year of residency. The same SCT questions and survey were then sent to a group of OTL-HNS surgeons from 4 Canadian universities.

Results: For 64.4% of faculty and residents, the study was their first exposure to SCT. Overall, residents found it difficult to adapt to this form of testing, thought that the clinical scenarios were not clear and believed that SCT was not useful for assessing clinical reasoning. In contrast, the vast majority of experts felt that the test questions reflected real-life clinical situations and would recommend SCT as an evaluation method in OTL-HNS.

Conclusion: Views about the acceptability of SCT as an assessment tool for clinical reasoning differed between OTL-HNS residents and experts. Education about SCT and increased exposure to this testing method are necessary to improve residents' perceptions of SCT.

Contexte : Le test de concordance de script (TCS) est une méthode objective d'évaluation du raisonnement clinique qui mesure la capacité d'interpréter les renseignements médicaux en contexte d'incertitude. Beaucoup d'études en appuient la validité en tant qu'outil pour évaluer l'enseignement supérieur, mais on en sait peu sur son acceptabilité auprès des principales parties prenantes. Le but de cette étude était de déterminer l'acceptabilité du TCS chez les résidents en otorhinolaryngologie – chirurgie de la tête et du cou (ORL – chirurgie tête et cou) et un groupe de référence composé d'experts.

Méthodes : En 2013 et 2016, une série de questions de TCS, de même qu'un questionnaire post-test, ont été inclus dans l'examen national en cours de formation NITE (National In-Training Examination) pour l'ORL – chirurgie tête et cou. Cet examen est administré à tous les résidents en ORL – chirurgie tête et cou au Canada qui sont entre leurs deuxième et cinquième années de résidence. Les mêmes questions de TCS ont été envoyées à un groupe de chirurgiens en ORL – chirurgie tête et cou de 4 université canadiennes.

Résultats : Pour 64,4 % des membres facultaires et des résidents, l'étude était leur première exposition au TCS. Dans l'ensemble, les résidents ont trouvé difficile de s'adapter à cette forme de test, même si les scénarios cliniques étaient clairs, et ils ont estimé que le TCS était peu utile pour évaluer le raisonnement clinique. En revanche, la grande majorité des experts ont jugé que les questions du test reflétaient la réalité des cas cliniques et recommanderaient le TCS comme méthode d'évaluation en ORL – chirurgie tête et cou.

Conclusion : Entre les résidents et les experts en ORL – chirurgie tête et cou, les points de vue quant à l'acceptabilité du TCS comme outil d'évaluation du raisonnement clinique ont différé et il faudrait y exposer les résidents davantage pour améliorer leur perception du TCS.

To deliver safe and effective patient care, clinicians must develop and apply sound clinical reasoning skills. Although definitions of clinical reasoning differ, they generally include the idea that clinical reasoning entails cognitive operations allowing clinicians to observe, collect and analyze information, resulting in decisions and actions that take into account a patient's specific circumstances and preferences.^{1,2} Explicit teaching and formal assessment of these essential clinical reasoning skills are crucial during residency, as they represent the formative training years of a burgeoning physician.

In Canada, clinical reasoning skills are currently assessed mainly by oral examinations. However, oral examinations are time consuming, with only 1 resident being evaluated at a time, and they require substantial human and financial resources. Also, human factors (anxiety, stress, attitude, interpersonal conflict) can potentially limit the objectivity of the evaluation.³

Clinical reasoning is often underrepresented in written examinations; for example, in the 2012 National In-Training Examination (NITE), an annual formative examination written by all Canadian residents in otolaryngology – head and neck surgery (OTL-HNS), only 5% of questions were designed to evaluate clinical reasoning. The remainder of the questions were designed to assess either factual knowledge (50%) or the ability to apply knowledge (45%).

Given that there does not exist a single assessment method that adequately addresses all elements of clinical reasoning, a combination of different assessment methods is often required. One such assessment method is script concordance testing (SCT).

Script concordance testing was developed in 1998 by Charlin and colleagues.⁴ It has the unique feature of assessing the interpretation of clinical data under conditions of uncertainty. It has its roots in the concept of the illness script, a specialized knowledge structure, different for each clinician, in which medical knowledge is organized. It creates links between information such as illnesses, clinical features and management options.⁵ Typically, when assessing a specific clinical situation, a physician will subconsciously choose 1 of his or her scripts on the basis of a key element of the encounter. The physician will then compare the clinical situation with different elements of the script to determine the best hypothesis. He or she will actively use his or her knowledge networks (scripts) to constantly make judgments on the effect that each new piece of information has on the status of the hypothesis. Script concordance tests are designed with the assumption that the participant already has some factual knowledge about the subject being tested. As such, the more experienced the clinician, the more refined the scripts.²

The format of SCT differs from that of other evaluation tools (Figure 1). Each question starts with a clinical

scenario, usually inspired by a real patient encounter. Next, a hypothesis is added to activate a script in the participant's mind.⁵ Finally, new information is provided to recreate the process by which the physician searches for new information to confirm or rule out a hypothesis. The answer choices are formatted in a way that invites the participant to reflect on whether the probability of the hypothesis being true has changed in light of the new information.

Script concordance testing is based on the principle that the multiple judgments made in these clinical reasoning processes can be probed and their concordance with those of a panel of reference experts can be measured. The scoring key is thus based on the answers given by a group of experts to the same questions. The traditional scoring method for SCT is an aggregate one (1 point is given for the most popular answer among the experts, other answers chosen by a minority of experts are given partial points, and answers not chosen by any expert are given 0 points).

In the last decade or so, SCT has been widely studied, with validity evidence gathered in various contexts, such as different disciplines (e.g., pediatrics,⁶ urology,⁷ emergency medicine,⁸ optometry,⁹ nursing¹⁰), different levels of expertise (medical students,¹¹ residents, attending physicians) and different purposes (self-evaluation,¹² formative or summative assessments). Specific to OTL-HNS, Kania and colleagues and Iravani and colleagues compared 2 groups and concluded that SCT could discriminate between levels of expertise.^{13,14} Despite this emerging knowledge, little is known about the acceptability of SCT to the major stakeholders.

The acceptability of a test can be defined as the degree to which it appears practical, pertinent and related to the purpose of the test,¹⁵ on the basis of one's personal experiences, beliefs and (mis)conceptions.¹⁶ Acceptability includes the concept of face validity, but it also includes other considerations that will affect the person taking the test (time to complete, burden, etc.). Acceptability, along with reliability, validity, cost-effectiveness and educational impact, is an essential component of a test's usefulness (Van der Vleuten and colleagues introduced this concept of test usefulness in 1996¹⁶). This model defines reliability and validity as the main components of a test's usefulness, and acceptability, educational impact and costs as other important factors. The weight of acceptability relative to other components depends on the context in which the test is used.¹⁷ The degree of acceptability, by residents or experts, will influence the efficiency and implementation of the test. Therefore, the aim of our study was to assess and compare the acceptability of the SCT among OTL-HNS residents and faculty members and to identify potential factors that can affect its acceptability.

Please answer the following questions regarding the probability of the **diagnosis** using the scale below

-2	-1	0	+1	+2
Much less likely	Less likely	Does not change the diagnosis	More likely	Much more likely

A 66-year-old man presents with 3 episodes of rotary intermittent vertigo in the past 18 months. The first 2 episodes lasted minutes to hours. He had a normal hearing test 2 years ago, before the onset of symptoms. The neurologic exam is normal (no nystagmus, normal cranial nerve test and normal cerebellar function).*

If you are thinking†	But you learn that‡	The diagnosis becomes										
a. Labyrinthitis	The patient had intermittent headaches and was diagnosed with migraines 8 months earlier.	<table style="margin: auto;"> <tr> <td style="text-align: center;">-2</td> <td style="text-align: center;">-1</td> <td style="text-align: center;">0</td> <td style="text-align: center;">+1</td> <td style="text-align: center;">+2</td> </tr> <tr> <td style="text-align: center;"><input type="checkbox"/></td> <td style="text-align: center;"><input type="checkbox"/></td> <td style="text-align: center;"><input type="checkbox"/></td> <td style="text-align: center;"><input type="checkbox"/></td> <td style="text-align: center;"><input type="checkbox"/></td> </tr> </table>	-2	-1	0	+1	+2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
-2	-1	0	+1	+2								
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>								
b. Menière disease	A current audiogram is normal.	<table style="margin: auto;"> <tr> <td style="text-align: center;">-2</td> <td style="text-align: center;">-1</td> <td style="text-align: center;">0</td> <td style="text-align: center;">+1</td> <td style="text-align: center;">+2</td> </tr> <tr> <td style="text-align: center;"><input type="checkbox"/></td> <td style="text-align: center;"><input type="checkbox"/></td> <td style="text-align: center;"><input type="checkbox"/></td> <td style="text-align: center;"><input type="checkbox"/></td> <td style="text-align: center;"><input type="checkbox"/></td> </tr> </table>	-2	-1	0	+1	+2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
-2	-1	0	+1	+2								
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>								
c. Menière disease	The patient noticed that his hearing has decreased slightly in the last 2 years. According to him, this is explained by his advancing age.	<table style="margin: auto;"> <tr> <td style="text-align: center;">-2</td> <td style="text-align: center;">-1</td> <td style="text-align: center;">0</td> <td style="text-align: center;">+1</td> <td style="text-align: center;">+2</td> </tr> <tr> <td style="text-align: center;"><input type="checkbox"/></td> <td style="text-align: center;"><input type="checkbox"/></td> <td style="text-align: center;"><input type="checkbox"/></td> <td style="text-align: center;"><input type="checkbox"/></td> <td style="text-align: center;"><input type="checkbox"/></td> </tr> </table>	-2	-1	0	+1	+2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
-2	-1	0	+1	+2								
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>								

Fig. 1. Each question starts with a clinical scenario (*). The description is short, realistic and deliberately includes uncertainty or conflicting information, as in real-life clinical practice when patients have an incomplete history or provide ambiguous information. After the description, a hypothesis (†) about the diagnosis, investigations or treatment is provided. Finally, additional information (‡) is added to the scenario. This new information is provided to try to recreate the process whereby the physician searches for new information to confirm or rule out a hypothesis. The task of the participants is to specify the effect, positive or negative, of the new information on the status of the hypothesis. The answer is given on a 5-point Likert scale. The traditional scoring method gives 1 point for the most popular answer among the experts. Other answers chosen by a minority of experts are given partial points (no. of experts with a different answer divided by no. of experts with the most popular answer). Answers not chosen by any expert are given 0 points.

METHODS

Study overview

We created SCT questions specific to the field of OTL-HNS and then distributed the questions both to OTL-HNS residents (via the NITE) and to staff at 4 hospitals (via email). After answering the SCT questions, both groups responded to a post-test survey about the acceptability of SCT. This study received ethical approval from the Université de Montréal research ethics board (Comité d'éthique de la recherche en santé: 13-120-CERES-D).

Creation of SCT questions, scoring key and post-test survey

An SCT test was created with 8 clinical cases (3 in pediatrics, 3 in oncology, 3 in otology and 1 in rhinology), with a total of 28 questions. All questions were constructed using Fournier and colleagues' guidelines on how to write SCT tests.¹⁸ A team consisting of 1 OTL-HNS resident (A.-A.L.) and 2 OTL-HNS content experts (L.N., T.A.) wrote the SCT questions, and then 2 SCT experts (B.C., S.L.) reviewed and modified them. The SCT questions were then sent to the NITE committee.

Informed by a literature review, 3 of the authors (A.-A.L., L.N., T.A.) created a post-test survey for this

study. In the survey, participants were asked about their level of training (for residents), university affiliation, area of expertise (for faculty) and prior knowledge of SCT. Participants also rated their agreement with a series of statements regarding SCT on a 5-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree). Finally, participants were asked if they recommended SCT as an assessment tool for OTL-HNS residency and were asked to answer 2 open-ended questions (What did you like best about responding to SCT questions? What did you like least about responding to SCT questions?).

Participants

Residents

The SCT questions we created for this study and the post-test survey were added to the end of the annual NITE, with an explicit statement to participants that this was a separate, optional and confidential section and that their test scores for this section would not be forwarded to their program directors. All residents of a Canadian OTL-HNS program in 2013 and 2016, from postgraduate year 2 (PGY-2) to postgraduate year 5 (PGY-5), were eligible to participate. In 2016, residents were excluded if they had participated in the study in 2013.

Experts

There are different criteria that can be used to choose the members of an expert group, and none are universally accepted.¹⁹ We chose to include OTL-HNS surgeons who practised in an academic hospital, who were regularly involved in the teaching and assessment of OTL-HNS residents and who had a minimum of 1 year of fellowship training. The expert group was recruited from 4 universities (Université de Montréal, Université Laval, Université de Sherbrooke, McGill University). Emails were sent in 2013 and 2016 to the OTL-HNS program directors of those universities, who then forwarded the messages to all eligible faculty members. Respondents were sent an email with a consent form, questions from the SCT test we created that were related to their clinical area of expertise and the post-test survey.

Statistical analysis

For the analysis of post-test survey results, we compared the results between the 2 participating groups using the Mann-Whitney *U* test using IBM SPSS Statistics 24, with a *p* value less than 0.05 considered statistically significant. We calculated the median score for each group for the questions probing agreement with statements on a 5-point Likert scale and then compared these values between the groups using the Mann-Whitney *U* test. Subgroup analysis was done with the same tests. Of note, we did not analyze the SCT test

scores because our objective in this study was not to assess the validity of SCT. Thematic analysis was used for the open-ended questions.

RESULTS

Participant characteristics

Forty-one residents (response rate 21.0%) participated in this study. The 2016 results of 5 residents were excluded because they had participated in the study in 2013. A total of 36 residents were included in the final analysis, 20 of whom were junior residents (PGY-2 and PGY-3) and 16 of whom were senior residents (PGY-4 and PGY-5). We received 23 responses from experts (response rate 34.8%). The post-test survey was incomplete for 3 participants in the resident group and 2 in the expert group. The characteristics of the 2 groups of participants are described in Table 1 and Table 2. The level of knowledge regarding SCT was low in both groups.

Acceptability

Results from the post-test acceptability survey of both residents and experts are presented in Table 3. For each statement, experts had, on average, a more positive opinion of the SCT test than the residents. They were more likely to find the questions clear, to find this type of test easy to adapt, and to think that SCT was a useful tool to evaluate clinical reasoning. Despite these differences, both experts and residents agreed that the clinical scenarios were realistic and were likely to be encountered in everyday practice.

Our results did not show a significant difference between the opinions of junior and senior residents (Table 4) for the statements, except for 1 statement. Interestingly, senior residents found that the time allotted to answer all the SCT questions was insufficient, while junior residents found the same amount of time to be adequate.

The majority of the experts (73.9%) were in favour of integrating SCT into OTL-HNS residency programs as a formal assessment tool, while only 33.3% of residents held this view. There was no difference between the views of junior and senior residents (30.0% and 37.5%). The combined group of experts and residents with prior knowledge of SCT were more inclined to recommend SCT (68.4%) than the group with no prior knowledge (39.5%) ($p < 0.05$).

DISCUSSION

To our knowledge, this is the first study to examine the acceptability of SCT as a formal assessment tool in residency among major stakeholders. We found that experts

Table 1. Characteristics of resident participants

Characteristic	No. (%) of participants <i>n</i> = 36
Level of training	
PGY-2	6 (16.7)
PGY-3	14 (38.9)
PGY-4	10 (27.8)
PGY-5	6 (16.7)
Knowledge of SCT before this study	
Yes	9 (25.0)
No	26 (72.2)
Year of participation	
2013	15 (41.7)
2016	21 (58.3)

PGY = postgraduate year; SCT = script concordance testing.

Table 2. Characteristics of expert participants

Characteristic	No. (%) of participants <i>n</i> = 23
Subspecialty	
Oncology	8 (34.8)
Otology	5 (21.7)
Pediatrics	5 (21.7)
Rhinology	5 (21.7)
University affiliation	
Université de Montréal	18 (78.3)
Université de Sherbrooke	4 (17.4)
Université Laval	1 (4.3)
McGill University	0 (0)
Knowledge of SCT before this study	
Yes	10 (43.5)
No	12 (52.2)
Year of participation	
2013	20 (87.0)
2016	3 (13.0)

SCT = script concordance testing.

were significantly more inclined to accept SCT than residents. We hypothesize that the discrepancy between residents' expectations of a test and the reality of being a practising physician may have contributed to the difference of opinion. Residents probably expect certainty in testing conditions, especially in high-stakes examinations. Experts are probably more used to uncertainty. In response to our open-ended questions, experts mentioned that they liked the SCT for its global approach to a case and for its ability to assess knowledge and clinical reasoning at the same time. They were also aware that different evaluation tools must be employed to create an evaluation that truly reflects a resident's capacity.¹⁷

The use of multiple forms of assessment ensures that different aspects of OTL-HNS residents' knowledge and aptitude are evaluated. It may be adequate to focus on assessing factual knowledge among very junior residents, but as training progresses, more complex skills such as decision-making, particularly in difficult clinical

situations, need to be taught and assessed. Various assessment methods have been described to evaluate clinical reasoning, such as multiple-choice questions (MCQs), patient management problems, key features problems, oral examinations and objective structured clinical examinations (OSCEs). In relation to these methods, the strength of SCT is that you can administer more items per unit of testing time. The more content domains you can probe in a given amount of time, the more accurate your assessment of a learner's overall performance will be, because performance on 1 item does not necessarily predict performance on other items probing different content domains (content specificity concept).³ For example, an average SCT examinee can complete about 60–90 questions per hour, whereas an average OSCE examinee can work through only 8–12 items (i.e., stations) per hour. Also, MCQs are good for combating the content specificity problem, but it is easier to develop MCQs to probe pure factual knowledge than to evaluate clinical reasoning. Oral examinations are also a good option for testing multiple levels of knowledge, but SCT needs fewer resources in terms of cost, people and time.

In our study, residents and experts had a different opinion of SCT. A larger percentage of residents than experts found that the test questions were unclear, that they were not useful for evaluating clinical reasoning, that they were not representative of real clinical situations and that it would be hard to adapt to this kind of test.

Other studies have collected participants' opinions about SCT as a secondary objective. In 2015, Cobb and colleagues used a focus group of 18 final-year undergraduates veterinarian students to assess the acceptability of SCT. The students concluded that SCT is confusing but that it is probably more relevant to decision-making in clinical practice and encourages participants to reflect on their own experience and knowledge.²⁰ In a study by Kelly and colleagues, students reported that they did not find SCT questions harder than regular MCQ, but they still preferred MCQs.²¹ Kania and colleagues reported no difference in opinion on SCT between students and experts.¹³ The study setting may partially explain why those results differed from ours. Our study took place immediately after a formal annual NITE and was not limited to the research setting. The perceived connection to an examination possibly influenced residents' views on the acceptability of SCT. Also, the possibility that SCT might be used as a tool in a formal examination was more tangible and thus residents may have viewed this testing method as more threatening.

The majority of the residents in our study were not familiar with SCT, probably because clinical reasoning has traditionally been assessed with oral questions.³ The responses to our open-ended questions showed that some residents felt that oral questions are a better way to evaluate clinical reasoning abilities. The residents'

Table 3. Rates of agreement with statements in post-test survey: comparison of residents and experts

Statement	No. (%) of participants;* group		p value
	Residents n = 36	Experts n = 23	
The instructions at the beginning of the test were clear.	23 (65.7)	20 (87.0)	0.014
The SCT examples (with answers) presented before the test were helpful.	22 (62.9)	20 (87.0)	0.007
The test questions were clear.	16 (44.4)	21 (91.3)	< 0.001
It was easy to adapt to this kind of test.	11 (30.5)	17 (73.9)	0.001
The questions reflected authentic clinical situations.	25 (69.4)	23 (100)	0.003
The questions reflected situations that I have encountered or I am likely to encounter during my training/practice.	26 (72.2)	23 (100)	0.005
This evaluation was useful for testing the ability to reason through cases in otolaryngology.	18 (50.0)	21 (91.3)	0.002

SCT = script concordance testing.
*Participants who answered "agree" or "strongly agree."

Table 4. Rates of agreement with statements in post-test survey: comparison of junior and senior residents

Statement	No. (%) of participants;* stage of training		p value
	Junior n = 20	Senior n = 16	
The instructions at the beginning of the test were clear.	13 (65.0)	10 (62.5)	0.80
The SCT examples (with answers) presented before the test were helpful.	13 (65.0)	9 (56.3)	0.40
The test questions were clear.	11 (55.0)	5 (31.3)	0.53
It was easy to adapt to this kind of test.	7 (35.0)	4 (25.0)	0.50
The time allotted to complete the test was adequate.	15 (75.0)	6 (37.5)	0.036
The questions reflected authentic clinical situations.	15 (75.0)	10 (62.5)	0.34
The questions reflected situations that I have encountered or I am likely to encounter during my training/practice.	14 (70.0)	12 (75.0)	0.56
This evaluation was useful for testing the ability to reason through cases in otolaryngology.	10 (50.0)	9 (56.3)	0.80

SCT = script concordance testing.
*Participants who answered "agree" or "strongly agree."

relative lack of knowledge about the principles of SCT, and their lack of practice with them, could have negatively affected their opinion of SCT. This may explain why experts and residents with prior knowledge of SCT were more inclined to recommend the use of SCT to evaluate residents. Interestingly, when we compared the opinions of junior and senior residents, the only divergence seen was related to the time allotted to answer SCT questions. Senior residents expressed the desire for more time, while junior residents felt that they had enough. Given that senior residents have more experience and knowledge, and probably more elaborate scripts, it is possible that they spent more time thinking about their answers and going through their scripts.

Limitations

The main limitation of our study was the low participation rate (21.0% for residents, 34.8% for experts). It is hard to know if the people who chose to participate were different than those who did not participate. Almost equivalent numbers of junior and senior residents participated in this study, so we believe that the

participants were still representative of the population studied. Also, experts were recruited from only 4 Canadian universities, all in Quebec, and most of those who chose to participate were from Université de Montréal (78.3%). As otolaryngology program accreditation and final board examinations are administered by national bodies, we have no reason to think that the responses from our expert participants would differ from those of experts in the rest of Canada. Other limitations of this study are that the questionnaires were administered at the end of the NITE and thus fatigue could have influenced the results from the residents and that the surveys were not validated.

CONCLUSION

Our study results suggest that OTL-HNS faculty members and residents have differing views on the acceptability of SCT as an assessment tool. Most of the experts and residents who participated in our study had never heard of SCT. We think that using SCT more frequently, in an informal learning setting, may help residents get familiar with SCT and enhance its acceptability as an

assessment tool for clinical reasoning. Use of SCT as a learning tool could also be a way to familiarize residents and faculty members with the method. Different groups are currently working on making this method more accessible for teaching (e.g., in a Web-based format). We believe that SCT could be a useful and powerful addition to our testing methods for residents once it is more widely known and understood.

Affiliations: From the Department of Otolaryngology – Head and Neck Surgery, Université de Montréal, Montreal, Que. (Leclerc, Charlin, Ayad); the Department of Otolaryngology – Head and Neck Surgery, McGill University, Montréal, Que. (Nguyen); the Centre for Medical Education, Université de Montréal, Montreal, Que. (Charlin); the Centre for Medical Education, McGill University, Montreal, Que. (Lubarsky); and the Department of Neurology, McGill University, Montreal, Que. (Lubarsky).

Competing interests: None declared.

Contributors: A.-A. Leclerc, L. Nguyen, B. Charlin, S. Lubarsky and T. Ayad designed the study. A.-A. Leclerc acquired the data, which A.-A. Leclerc and T. Ayad analyzed. A.-A. Leclerc and T. Ayad drafted the manuscript, which A.-A. Leclerc, L. Nguyen, B. Charlin, S. Lubarsky and T. Ayad critically reviewed. All authors provided final approval of the version to be published.

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

Data sharing: The data sets analyzed in this study are not publicly available as they were provided by the National In-Training Examination (NITE) committee, but they are available from the corresponding author on reasonable request.

References

- Durning SJ, Artino AR. Situativity theory: a perspective on how participants and the environment can interact: AMEE Guide No. 52. *Med Teach* 2011;33:188-99.
- Eva KW, Hatala RM, Leblanc VR, et al. Teaching from the clinical reasoning literature: combined reasoning strategies help novice diagnosticians overcome misleading information. *Med Educ* 2007;41:1152-8.
- Charlin B, Bordage G, Van der Vleuten C. L'évaluation du raisonnement clinique. *Pédagogie Médicale* 2003;4:42-52.
- Charlin B, Brailovsky C, Leduc C, et al. The Diagnosis Script Questionnaire: a new tool to assess a specific dimension of clinical competence. *Adv Health Sci Educ Theory Pract* 1998;3:51-8.
- Lubarsky S, Dory V, Duggan P, et al. Script concordance testing: from theory to practice: AMEE Guide No. 75. *Med Teach* 2013;35:184-93.
- Carriere B, Gagnon R, Charlin B, et al. Assessing clinical reasoning in pediatric emergency medicine: validity evidence for a Script Concordance Test. *Ann Emerg Med* 2009;53:647-52.
- Sibert L, Charlin B, Gagnon R, et al. [Evaluation of clinical reasoning in urology: contribution of the Script Concordance Test]. *Prog Urol* 2001;11:1213-9.
- 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;7:20-30.
- Faucher C, Dufour-Guindon MP, Lapointe G, et al. Assessing clinical reasoning in optometry using the Script Concordance Test. *Clin Exp Optom* 2016;99:280-6.
- Cote S, Tanguay A, Gagnon R, et al. 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;15:7-20.
- Humbert AJ, Johnson MT, Miech E, et al. Assessment of clinical reasoning: a Script Concordance Test designed for pre-clinical medical students. *Med Teach* 2011;33:472-7.
- Caire F, Sol JC, Moreau JJ, et al. [Self-assessment for neurosurgery residents by script concordance test (SCT). The process of test elaboration]. *Neurochirurgie* 2004;50:66-72.
- 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. *Arch Otolaryngol Head Neck Surg* 2011;137:751-5.
- Iravani K, Amini M, Doostkam A, et al. The validity and reliability of Script Concordance Test in otolaryngology residency training. *J Adv Med Educ Prof* 2016;4:93-6.
- Mosier CI. A critical examination of the concepts of face validity. *Educ Psychol Meas* 1947;7:191-205.
- Van Der Vleuten CP. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ Theory Pract* 1996;1:41-67.
- Van der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. *Med Educ* 2005;39:309-17.
- Fournier JP, Demeester A, Charlin B. Script concordance tests: guidelines for construction. *BMC Med Inform Decis Mak* 2008;8:18.
- Lurie S. Towards greater clarity in the role of ambiguity in clinical reasoning. *Med Educ* 2011;45:326-8.
- Cobb KA, Brown G, Hammond R, et al. Students' perceptions of the Script Concordance Test and its impact on their learning behavior: a mixed methods study. *J Vet Med Educ* 2015;42:45-52.
- Kelly W, Durning S, Denton G. Comparing a script concordance examination to a multiple-choice examination on a core internal medicine clerkship. *Teach Learn Med* 2012;24:187-93.