Online Script Concordance Test for Clinical Reasoning Assessment in Otorhinolaryngology

The Association Between Performance and Clinical Experience

Romain E. Kania, MD, PhD; Benjamin Verillaud, MD; Hugo Tran, MD; Robert Gagnon, MSc; Driss Kazitani, MD; Patrice Tran Ba Huy, MD; Philippe Herman, MD, PhD; Bernard Charlin, MD, PhD

Objective: To report on the creation and administration of an online Script Concordance Test (SCT) for ear, nose, and throat (ENT), the ENT-SCT.

Design: Prospective study.

Setting: Two tertiary care university centers.

Participants: In total, 132 individuals were asked to test an ENT-SCT of 20 cases and 94 questions based on the major educational objectives of the ENT residency program.

Main Outcome Measures: Three levels of experience were tested: medical students, ENT residents, and board-certified otorhinolaryngologists as the expert panel. The test's construct validity—whether scores were related to clinical experience—was statistically analyzed. Reliability was estimated by the Cronbach α internal consistency coefficient. Participants' perception of the test was assessed with the use of a questionnaire.

Results: The 65 respondents with usable data were medical students (n=21), ENT residents (n=22), and experts (n=22). Total mean (SD) test scores differed significantly: 76.81 (3.31) for the expert panel, 69.05 (4.35) for residents, and 58.29 (5.86) for students. The Cronbach α coefficient was 0.95. More than two-thirds of the participants found the test to be realistic and relevant for assessing clinical reasoning. The test was also considered fun, interesting, and intuitive.

Conclusions: The Web-based ENT-SCT is feasible, reliable, and useful for assessing clinical reasoning. This online assessment tool may have applications for residency programs and continuing medical education.

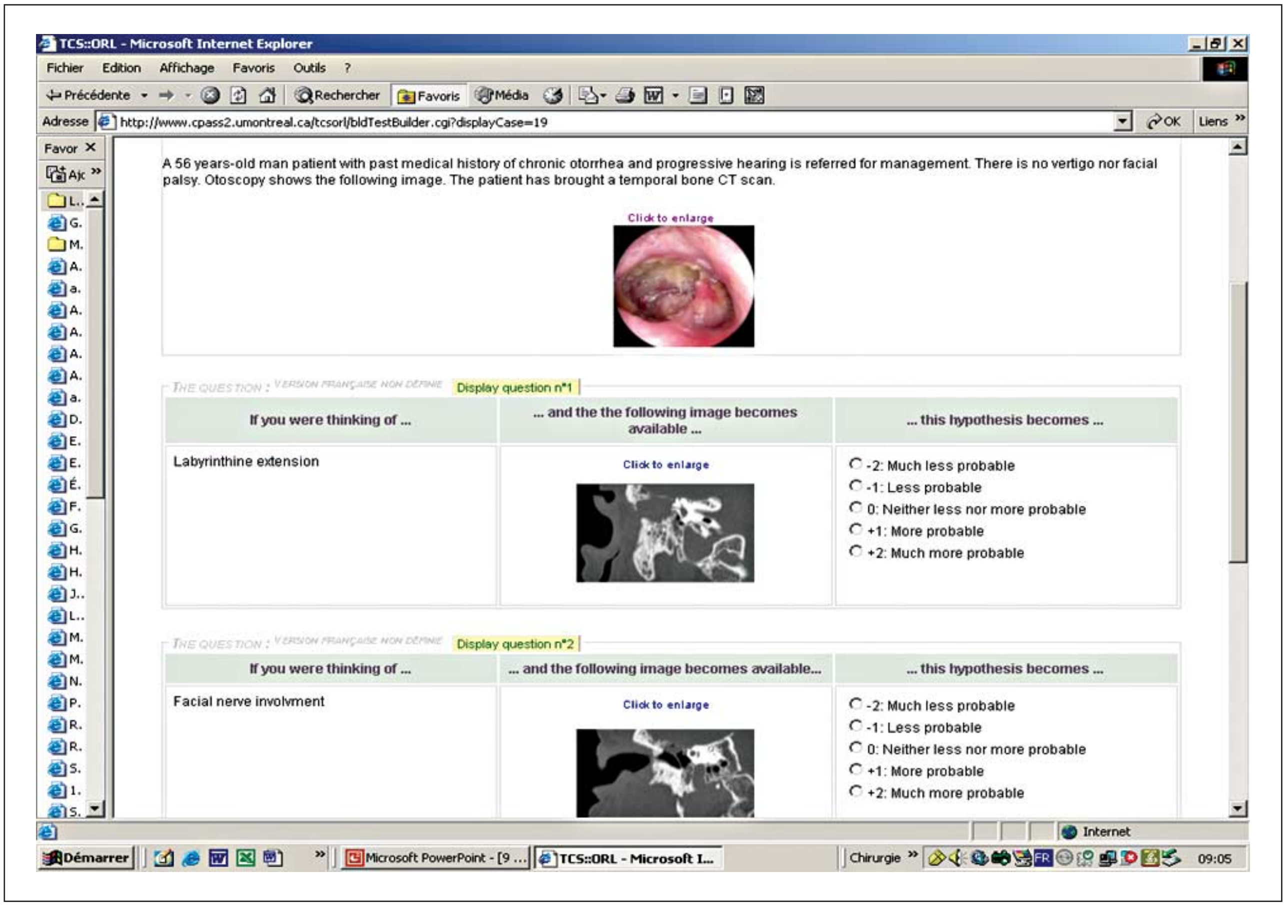
Arch Otolaryngol Head Neck Surg. 2011;137(8):751-755

Author Affiliations: Department of Otorhinolaryngology, Head and Neck Surgery, Center for Neurosensorial Head & Neck Diseases, Lariboisière Hospital, Assistance Publique des Hôpitaux de Paris, University Paris 7 Paris Diderot and CESEM (Centre d'Etude de la SensoriMotricité), UMR (Unité Mixte de Recherche) 8194, CNRS (Centre National de Recherche Scientifique)–University Paris 5 Paris Descartes, Paris, France (Drs Kania, Verillaud, Tran, Tran Ba Huy, and Herman); and CPASS (Centre de pédagogie appliquée en sciences de la santé), Faculty of Medicine, University of Montreal, Montreal, Quebec, Canada (Mr Gagnon and Dr Kazitani).

SCRIPT THEORY[1] ASSERTS THAT experienced practitioners possess elaborate networks of knowledge for the regular tasks they perform, called *illness scripts*.[2] Scripts are made of links between illnesses, their clinical features, and their treatment. They allow for the efficient use of knowledge for diagnosis and the choice of investigation and treatment options. For instance, for an experienced practitioner, the association between Claude Bernard-Horner syndrome and periorbital pain will suggest carotid artery dissection and appropriate investigation and treatment.

According to script theory, each new element of clinical information is examined in light of expected values related to the relevant activated scripts. Clinical reasoning therefore results from multiple interpretations of data.[3] Scripts begin to appear when students are faced with their first clinical cases and are developed and refined during their entire clinical career.[4] The Script Concordance Test (SCT) is based on the principle that each interpretation of clinical data can be captured and compared with that of a panel of experts, therefore providing a measure of clinical reasoning quality. The format of the SCT[5] allows for incorporating the uncertainty that characterizes practice,[6] and physicians consider that the required cognitive tasks are close to the reality of clinical reasoning.[5] Studies[2,7-9] of the test in domains as diverse as neurology, radio-oncology, pediatrics, and perioperative reasoning have shown a positive relationship between test scores and clinical experience. These findings contrast with observations[10] from more traditional ways of testing, in which those being tested achieve higher scores at the end of formal training (end of residency) compared with seasoned professionals.

The number of new technologies available for medical training is growing. They are designed to allow easy test administration and scoring.[11] Tested situations are enriched with clinical images, videos, or radiographic images. To our knowledge, the


TCS::ORL - Microsoft Internet Explorer
Fichier Edition Affichage Favoris Outils ?
Précédente Rechercher Favoris Média
Adresse http://www.cpass2.umontreal.ca/tcsorl/bldTestBuilder.cgi?displayCase=19
OK Liens
Favor
Click to enlarge
Internet
Démarrer
Microsoft PowerPoint - [9 ...
TCS::ORL - Microsoft I...
Chirurgie
FR
09:05

A 56 years-old man patient with past medical history of chronic otorrhea and progressive hearing is referred for management. There is no vertigo nor facial palsy. Otoscopy shows the following image. The patient has brought a temporal bone CT scan.

THE QUESTION : VERSION FRANÇAISE NON DÉFINIE Display question n°1

If you were thinking of ...	... and the the following image becomes available ...	... this hypothesis becomes ...
Labyrinthine extension	Click to enlarge	○ -2: Much less probable ○ -1: Less probable ○ 0: Neither less nor more probable ○ +1: More probable ○ +2: Much more probable

THE QUESTION : VERSION FRANÇAISE NON DÉFINIE Display question n°2

If you were thinking of ...	... and the following image becomes available...	... this hypothesis becomes ...
Facial nerve involvment	Click to enlarge	○ -2: Much less probable ○ -1: Less probable ○ 0: Neither less nor more probable ○ +1: More probable ○ +2: Much more probable

Figure 1. Representative image of an online clinical situation described in a short, written vignette illustrated with an otoscopy image. Questions are displayed below the case presentation. A relevant hypothesis is given in the left column; new information is presented in the middle column (computed tomographic [CT] scan, coronal view); and in the right column, the test participant uses the 5-point Likert scale to evaluate the probability of the hypothesis. The test is available at http://www.cpass4.umontreal.ca/tcsorl/.

SCT format has not been used in the ear, nose, and throat (ENT) discipline. This study reports on the creation and administration of an online SCT for ENT, the ENT-SCT. It presents evidence of score reliability and examines score validity with the hypothesis that experienced physicians will have higher scores than will residents, who will have higher scores than medical students.

METHODS

DEVELOPMENT OF ENT-SCT

Using the educational objectives of the training program for ENT residents, 2 faculty members (R.E.K. and B.C.), senior otorhinolaryngologists, identified 25 clinical situations representative of ENT practice. Situations were described in short, written vignettes, as illustrated in **Figure 1**. All clinical situations contained uncertainty, with 2 or more relevant diagnosis or management options.[12] No situation was unusual, but all were complex enough to be challenging for junior or senior residents.

Each case scenario was followed by a series of 3 to 5 questions, each with 3 parts (Figure 1). The first part ("If you were thinking of . . . ") consists of a relevant option for diagnosis or management. The second part (" . . . and then the following information becomes available . . . ") presents new information in text, imaging, or video format. This new information could be a physical sign, an imaging study, or a laboratory test result that might have an effect on clinical reasoning in diagnosis or management. The third part (" . . . this option becomes . . . ") consists of the participant recording a decision on a 5-point Likert scale in terms of the probability of the hypothesis presented in the first part after having the information in the second part—clinical reasoning and decision-making. Use of the Likert scale allowed for gauging the direction (positive or negative) and intensity of the effect.[12]

The second part of each question presents features (positive or negative) that a clinician would find useful in solving the clinical problem.[12] To separately test clinical reasoning for each question, consecutive questions do not present the same option. Two senior faculty members (P.H. and P.T.B.H.) experienced in otorhinolaryngology practice and teaching were asked to check the pertinence and feasibility of the SCT. They deleted or modified some questions and determined, by consensus, the difficulty level suitable for ENT residents. The final test consists of 20 cases and 94 questions (3-6 per case), including 34 images (15 clinical, 12 radiologic, and 7 complementary examination data) and 1 video. The test is administered online from the Web site of the University of Montreal (http://www.cpass4.umontreal.ca/tcsorl/).

PARTICIPANTS

Potential participants, representing 3 groups of clinical experience, from universities in Paris, France, and the surrounding region were asked to test the SCT. The first group consisted of fourth- to sixth-year medical students who had received educa-

Table. Example of the Scoring Grid Showing the Answer Key Construction and Credit Provided for a Given Question[a]

	Five-Point Likert Scale Recording the Decision				
	–2	–1	0	+1	+2
No. of panel members for each response	0	0	2	9	4
Proportional calculation	0/9	0/9	2/9	9/9	4/9
Credit provided	0	0	0.22	1	0.44

[a] The credit is calculated according to the proportion of expert panel members who selected a decision. The proportional calculation is obtained by dividing the number of panel members who provided an answer on the Likert scale by the mode; here, 9. Therefore, the maximum credit of 1 is given for the response chosen by most of the experts (ie, the mode), other responses are given fractional scores, and responses not chosen by experts are given a score of zero.

tion in otorhinolaryngology but no clinical experience. The second group included residents in otorhinolaryngology with a minimum of 1 year of clinical training in the specialty. The third group was composed of board-certified otorhinolaryngologists. Potential participants received a letter containing information on the format of the SCT, the objective of the project, advantages and disadvantages that could result from participation, criteria for inclusion in the study, how to withdraw from the study, and online access to the test. Anonymity was guaranteed.

SCORING

According to guidelines for construction of an SCT,[12] scoring is obtained by comparing participants' answers with those of a reference panel of experts (ie, physicians with experience in the domain; otorhinolaryngologists in this study). Experts completed the test under the same conditions as those for other participants. The aggregated responses of the experts to each item formed the test's answer grid.[7] Previous work[13] has shown that the reference panel of experts should consist of 15 to 20 members to achieve optimal reliability. The scoring scheme for each question is determined by the frequency with which each response is chosen by the experts. To ensure that each question is given equal weight, the value assigned to each response for a given question is transformed proportionally to give a maximum score of 1 for the response chosen by most of the experts (ie, the mode); other responses are given fractional scores depending on the number of experts choosing them. Responses not chosen by experts receive a score of zero. An example of the scoring grid is shown in the **Table**. The total score for the test was calculated by adding scores for each item. The final score was the total score divided by the number of items (ie, 4) expressed as a percentage. A theoretical score of 100% would mean that the participant had recorded the same decisions as most of the members of the expert panel for all questions. The 2 faculty members who developed the test were not part of the panel of reference, nor were the 2 who validated the test cases and questions.

QUALITATIVE ASSESSMENT OF THE TEST

Once the test was completed, participants were asked to answer a questionnaire including the 6 questions on a semiquantitative 5-point Likert scale of agreement, from 1 (not at all) to 5 (enormously). The first 2 questions focused on whether this ENT-SCT required cognitive tasks close to the reality of clinical reasoning, and the next 4 questions focused on the interest of the participants in the test. "Question 1: Did you find the test realistic? Did you find that clinical situations agreed with the reality of your clinical practice? Question 2: Did you find the test relevant? Do you think this method can help in evaluating clinical reasoning? Question 3: Did you find the test self-instructional? Was the test able to broaden your knowledge of otorhinolaryngology? Question 4: Did you find the test similar to a game? Did you like this method of assessment? Question 5: Did you find the test intuitive? Did you easily adapt to this method of assessment? Question 6: Did you find the test interesting? Would you like to use this method of assessment in the future?" At the end of the questionnaire, participants were invited to provide comments.

STATISTICAL ANALYSIS

Item scores and total scores for each participant were computed with the software provided on the University of Montreal's Web site (http://www.sctmed.ca). For members of the expert panel, scores were computed by a key that excluded their own responses. Test results were treated anonymously. Statistical analysis involved use of commercial software (Statview; SAS Institute, Inc, Cary, North Carolina). Descriptive statistics of the participants' scores were calculated. The normality of score distributions was evaluated by the Kolmogorov-Smirnov test. Analysis of variance was used to evaluate differences between mean scores for groups. For non-normally distributed variables, nonparametric tests were used, and the nonparametric Mann-Whitney test was used to assess differences between mean scores for groups. Reliability was estimated by the Cronbach α coefficient for internal consistency. The χ^2 test was used for analysis of the qualitative assessment between groups. All tests were 2-sided, and $P < .05$ was considered statistically significant. Data for participants with more than 3 missing values were excluded from analyses.

RESULTS

PARTICIPANTS

Letters were sent to 132 potential participants (47 students, 46 residents, and 39 board-certified ENT physicians). Eighty individuals agreed to participate and gave informed consent (response rate, 61%). Fifteen participants had more than 3 missing answers, and their data were excluded. Analyses were therefore conducted for 65 participants: 21 students, 22 residents, and 22 board-certified ENT physicians.

SCORES

Global mean scores and their distribution for the groups of participants according to their level of experience are illustrated in **Figure 2**. Total mean (SD) scores differed significantly among the groups: 76.81 (3.31; range, 72-83) for the expert panel, 69.05 (4.35; range, 61-75) for residents, and 58.29 (5.86; range, 47-68) for students ($P < .001$). The capacity of the test to discriminate according to clinical experience was demonstrated by the significant difference in mean scores between students and residents, residents and the expert panel, and students and

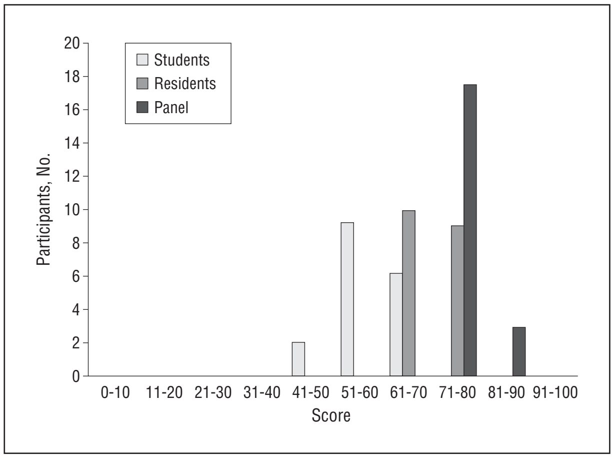


Figure 2. Graphic representation of distribution of total mean scores (in groups of 10) for medical students, residents, and the expert panel. The data represent the number of participants with total mean scores in each group.

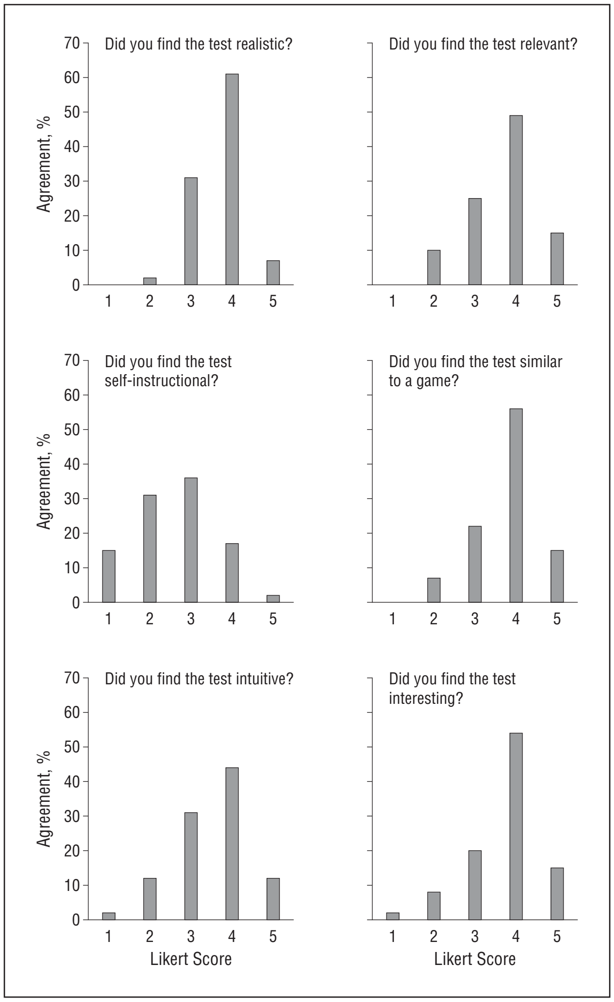


Figure 3. Answers to the 6 questions about the quality of the Script Concordance Test for the ear, nose, and throat discipline, expressed as percentages of agreement on the 5-point Likert scale (1, not at all; 5, enormously).

the expert panel (all $P<.001$). The Cronbach α coefficient for the test was 0.95.

QUALITATIVE ASSESSMENT

All participants completed the test in less than 1 hour. **Figure 3** summarizes the 5-point Likert-scale answers to the questionnaire about the test. With scores of 4 and 5 on the 5-point scale, more than two-thirds of the 65 participants (44 [68%]) found the test realistic and similar to clinical situations they encounter in otorhinolaryngology practice, 46 (71%) found the test similar to a game, 45 (69%) found it interesting, 42 (65%) believed it was relevant, and 36 (55%) indicated that it was intuitive. The self-instructional aspect of the test was rated low (23 participants [35%], a rating of 3). There was no significant difference in the responses to the qualitative questions between the 3 levels of training. Comments were provided by 24 of the 65 participants (37%). In general, students and residents were stimulated by the incorporation of text, images, and video. However, participants were disappointed that they could not review their answers after choosing a response.

COMMENT

The SCT format assesses the interpretation of clinical data, a crucial element of clinical reasoning and clinical judgment.[7] Scores for a series of questions about cases reflect the degree of concordance between participants' answers and those of an expert reference panel. High scores signify high concordance with the expert panel and, thus, the quality of clinical reasoning.[12]

Our testing of the SCT instrument specifically developed for assessment of ENT training showed excellent reliability (Cronbach α, 0.95). A test is considered sufficiently reliable with a Cronbach α coefficient of internal consistency of 0.80.[12] Our hypothesis, that mean total test scores on the SCT increase with clinical experience, was verified by our assessment of the ENT-SCT. Experienced clinicians had significantly higher mean SD scores than residents (76.81 [3.31 vs 69.05 [4.35]), who had significantly higher scores than students (58.29 [5.86]). These results indicate that the ENT-SCT can be used to document the progression of knowledge and experience with training in ENT.

Medical trainees, as well as experienced clinicians, gave high ratings for the test format and content. Interestingly, there was no significant difference in the responses to the qualitative questions between trainees and panel members. This suggests that testing clinical reasoning in a realistic manner is of interest. More than two-thirds of the participants considered the ENT-SCT to be relevant for clinical reasoning assessment; they found that it was interesting and contained cognitive tasks that were similar to those encountered in clinical practice. Furthermore, the test was considered to be intuitive by more than half the participants. These results must be considered in light of the fact that the test was taken without any special preparation, which differs from other assessment tools that, for optimal results, require specific knowledge of the domain. Nevertheless, the test was not considered a training tool because the tool did not assess performance for each question.

Online test administration offers many improvements over traditional testing. The scoring is automated and objective, participants can receive their scores immediately with explanations for the experts' answers on the Likert scale, and hyperlinks can direct participants to additional information in the form of text, video conferences, and other multimedia learning resources. The test format therefore becomes a method of initial or continuing medical education. These applications are being used in Spanish-speaking countries (eg, Spain, Argentina, and Chile; http://www.script.edu.es),where tests are coordinately constructed and administered in different countries; cases with SCT questions are provided online daily, and credits for training are given after completion.

Residents are practitioners who wish to acquire qualification in a specific domain of expertise. Therefore, comparing their performance in interpreting clinical data with that of a panel of experts in the discipline is legitimate for certification purposes. National and international collaboration can be envisioned. A common bank of cases and questions could be developed. Tests of cases and questions from such a bank could be administered at regular intervals to document the residents' progression during training and to detect areas of weakness that require remediation.

This study has several limitations. Although the test was developed with a deliberate effort to address major clinical topics in ENT, with only 20 cases, 94 questions, and 1 hour of testing time, it does not cover the whole discipline. Our response rate was 61%, and 15 of the 80 respondents had too many missing responses for analysis of their data. We did not have the means to detect whether the characteristics of nonparticipants differed from those of participants and whether the results would have differed with the inclusion of nonparticipants. The test was presented on a voluntary basis, and its unusual format may have deterred some participants. More studies are needed before considering its use in formal settings. Participants were recruited from a limited geographic area. It is uncertain whether the ENT-SCT, with its panel composed of specialists from one country, could be taken by equivalently trained cohorts in other countries and achieve similar results. This concern is particularly salient for cases or questions regarding management of the same condition. There is some evidence[14] showing that results are robust across panels and different countries; however, the findings need to be confirmed by studies conducted in differing contexts. Finally, SCT is an assessment tool able to reliably detect differences in capacity of interpreting clinical data among participants. Although there is evidence[2,3,8] showing that SCT scores progress along the level of training, to our knowledge, there are no reports in the educational literature on methods available to improve the performance of individuals whose test scores are low. Nevertheless, there are data[15] showing that, in the context of continuing professional development, the SCT format can be an efficient method to improve management practice.

In conclusion, this study has validated an online SCT for ENT that uses authentic clinical scenarios to compare trainees' judgment skills with those of experts. Most participants considered the test to be realistic, interesting, relevant, and intuitive to assess clinical data interpretation. Although more study is necessary, the tool may have a strong potential for assessing clinical reasoning and learning in initial or continuing medical education.

Submitted for Publication: October 26, 2010; final revision received February 19, 2011; accepted April 21, 2011.

Correspondence: Romain E. Kania, MD, PhD, Department of Otorhinolaryngology, Head and Neck Surgery, Lariboisière Hospital, 2 rue Ambroise Paré, 75010 Paris, France (romain.kania@lrb.aphp.fr).

Author Contributions: Drs Kania, Verillaud, and Charlin had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. *Study concept and design:* Kania, Tran Ba Huy, and Charlin. *Acquisition of data:* Kania, Tran, and Kazitani. *Analysis and interpretation of data:* Kania, Verillaud, Herman, and Charlin. *Drafting of the manuscript:* Kania, Gagnon, and Charlin. *Critical revision of the manuscript for important intellectual content:* Kania, Verillaud, Tran, Kazitani, Tran Ba Huy, Herman, and Charlin. *Statistical analysis:* Kania, Gagnon. *Obtained funding:* Tran Ba Huy. *Administrative, technical, and material support:* Kania, Tran, Kazitani, and Charlin. *Study supervision:* Tran Ba Huy, Herman, and Charlin.

Financial Disclosure: None reported.

Additional Contributions: We thank the medical students, residents, and experts who participated in this study. Nicolas Leboulanger, MD, and the staff of SJORLP (Société des Jeunes ORL Parisiens) provided significant help in contacting residents and experts. Mr Thomas Ehrlich assisted with mailing the questionnaire to the participants.

REFERENCES

1. Charlin B, Boshuizen HP, Custers EJ, Feltovich PJ. Scripts and clinical reasoning. *Med Educ.* 2007;41(12):1178-1184.
2. Meterissian S, Zabolotny B, Gagnon R, Charlin B. Is the Script Concordance Test a valid instrument for assessment of intraoperative decision-making skills? *Am J Surg.* 2007;193(2):248-251.
3. Charlin B, van der Vleuten C. Standardized assessment of reasoning in contexts of uncertainty: the script concordance approach. *Eval Health Prof.* 2004;27(3):304-319.
4. Schmidt HG, Norman GR, Boshuizen HP. A cognitive perspective on medical expertise: theory and implication. *Acad Med.* 1990;65(10):611-621.
5. Charlin B, Roy L, Brailovsky C, Goulet F, van der Vleuten C. The Script Concordance Test. *Teach Learn Med.* 2000;12(4):189-195.
6. Hall KH. Reviewing intuitive decision-making and uncertainty: the implications for medical education. *Med Educ.* 2002;36(3):216-224.
7. Carrière B, Gagnon R, Charlin B, Downing S, Bordage G. Assessing clinical reasoning in pediatric emergency medicine: validity evidence for a Script Concordance Test. *Ann Emerg Med.* 2009;53(5):647-652.
8. Lambert C, Gagnon R, Nguyen D, Charlin B. The Script Concordance test in radiation oncology. *Radiat Oncol.* 2009;4:7. doi:10.1186/1748-717X-4-7.
9. Lubarsky S, Chalk C, Kazitani D, Gagnon R, Charlin B. The Script Concordance Test. *Can J Neurol Sci.* 2009;36(3):326-331.
10. van der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ Theory Pract.* 1996;1:41-67.
11. Sibert L, Darmoni SJ, Dahamna B, Hellot MF, Weber J, Charlin B. On line clinical reasoning assessment with Script Concordance test in urology: results of a French pilot study. *BMC Med Educ.* 2006;6:45. doi:10.1186/1472-6920-6-45.
12. Fournier JP, Demeester A, Charlin B. Script concordance tests: guidelines for construction. *BMC Med Inform Decis Mak.* 2008;8:18. doi:10.1186/1472-6947-8-18.
13. Gagnon R, Charlin B, Coletti M, Sauvé 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? *Med Educ.* 2005;39(3):284-291.
14. Sibert L, Charlin B, Corcos J, Gagnon R, Lechevallier J, Grise P. Assessment of clinical reasoning competence in urology with the Script Concordance test. *Eur Urol.* 2002;41(3):227-233.
15. Petrella RJ, Davis P. Improving management of musculoskeletal disorders in primary care: the Joint Adventures Program. *Clin Rheumatol.* 2007;26(7):1061-1066.