Assessing clinical reasoning using a script concordance test with electrocardiogram in an emergency medicine clerkship rotation

Caroline Boulouffe,[1] Bruno Doucet,[1] Xavier Muschart,[1] Bernard Charlin,[2] Dominique Vanpee[1]

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/emermed-2012-201737).

[1]Emergency Department, CHU Mont Godinne-Dinant, Université Catholique de Louvain, Yvoir, Belgium
[2]Faculty of Medicine, Center of Pedagogy Applied to Health Sciences, University of Montreal, Quebec, Canada

Correspondence to
Dr Caroline Boulouffe, Emergency Department, CHU Mont Godinne-Dinant, UCL, Avenue Dr Gaston Therasse, 1, Yvoir 5530, Belgium; caroline.boulouffe@uclouvain.be

Received 11 July 2012
Revised 15 November 2012
Accepted 17 January 2013
Published Online First 13 April 2013

ABSTRACT

Objectives Script concordance tests (SCTs) can be used to assess clinical reasoning, especially in situations of uncertainty, by comparing the responses of examinees with those of emergency physicians. The examinee's answers are scored based on the level of agreement with responses provided by a panel of experts. Emergency physicians are frequently uncertain in the interpretation of ECGs. Thus, the aim of this study was to validate an SCT combined with an ECG.

Methods An SCT-ECG was developed. The test was administered to medical students, residents and emergency physicians. Scoring was based on data from a panel of 12 emergency physicians. The statistical analyses assessed the internal reliability of the SCT (Cronbach's α) and its ability to discriminate between the different groups (ANOVA followed by Tukey's post hoc test).

Results The SCT-ECG was administered to 21 medical students, 19 residents and 12 emergency physicians. The internal reliability was satisfactory (Cronbach's α=0.80). Statistically significant differences were found between the groups ($F_{0.271}$=21.07; p<0.0001). Moreover, significant differences (post hoc test) were detected between students and residents (p<0.001), students and experts (p<0.001), and residents and experts (p=0.017).

Conclusions This SCT-ECG is a valid tool to assess clinical reasoning in a context of uncertainty due to its high internal reliability and its ability to discriminate between different levels of expertise.

INTRODUCTION

Disparity between the clinical reasoning abilities of novice versus expert clinicians, and the high level of uncertainty experienced during practice of emergency medicine have been the subjects of research for many years.[1] A recent study showed that the training of future emergency physicians remains challenging.[2]

Script concordance tests (SCTs) are tools for evaluating clinical reasoning in the context of uncertainty. They are well validated and have been widely described in the literature for use in education, as well as in emergency medicine.[2–5] SCTs assess mental processes occuring during clinical reasoning by comparing the responses of examinees with those of expert physicians.[3] The examinee's answers are scored based on the level of agreement with the panel of experts,[3] and these scores are used to measure the patient care skills of healthcare providers in uncertain situations.[6] SCTs are based on script theories from cognitive psychology and are useful for assessing reasoning skills in medicine. Moreover, SCTs can be used in ambiguous or uncertain situations, and are therefore valuable for measuring real-life scenarios that are not adequately evaluated with current tests.[7] Thus, these tests uniquely provide a method for assessing the clinical reasoning ability of students during situations that mimic reality and involve a high degree of uncertainty.[2 3 6 7]

The interpretation of an ECG, which is a commonly performed test, includes a level of uncertainty.[8] Therefore, the aim of this study was to construct and validate a new SCT based on ECG data.

MATERIALS AND METHODS

Construction of the SCT

The SCT was constructed according to published guidelines and was composed of clinical vignettes based on actual cases, each followed by a series of three to four questions.[3] The vignettes did not contain all the information needed by the student/resident to propose an unequivocal diagnosis, treatment or complementary examination. These types of vignettes mimic real-life emergency situations, when the physician lacks information. The representative clinical scenarios used in this study were selected by a group of three emergency physicians and based on various cases encountered in our urgent care department and prehospital setting.

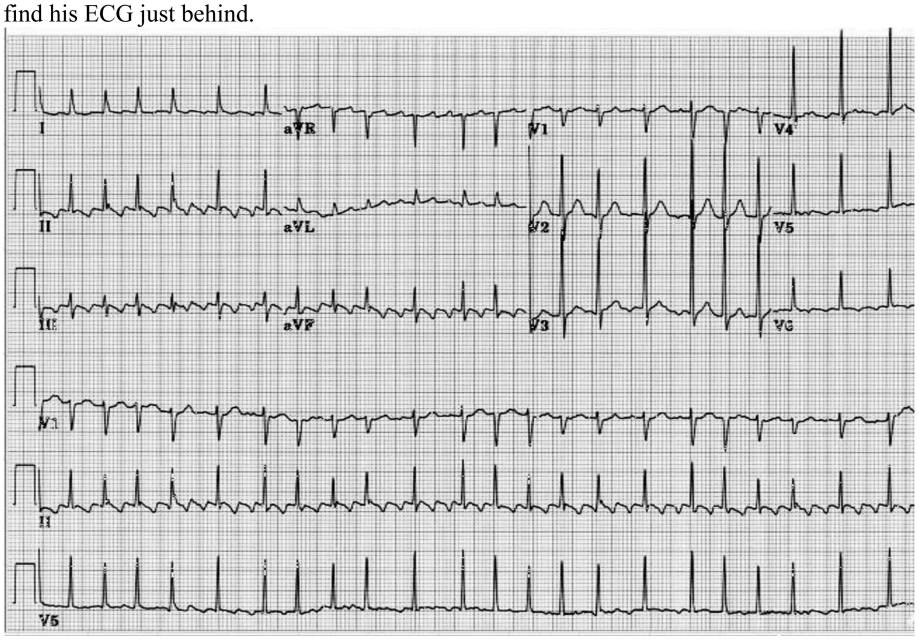
In this study, the authors provided an ECG compatible with the clinical situation. An example of an SCT vignette is shown in figure 1. This clinical situation is highly representative of our daily prehospital practice. The prehospital team is a part of our emergency department. This team is composed of an emergency physician, a nurse, a priority vehicle and all the material to resuscitate, treat and take care of a medical or traumatic patient.

The SCT structure differed slightly depending on the objective of the assessment (diagnosis, treatment or complementary examinations), and for each vignette, the questions that followed were ordered by format (eg, some items on diagnosis, followed by some items on treatment or complementary examinations).[3] Each test question consisted of three parts.[3] The first part concerned a diagnosis, treatment option or a further examination relevant to the situation and the ECG described in the vignette (ie, 'Your initial treatment is…'). The second part presented new information, such as a sign or condition, which might influence the student's/resident's diagnosis, treatment or

To cite: Boulouffe C, Doucet B, Muschart X, *et al*. *Emerg Med J* 2014;**31**:313–316.

Figure 1 Example of a clinical vignette.

You work in a pre-hospital medical setting. You are sent to take care of a 45-year-old patient who felt faint with palpitations. He suffers from Wolff Parkinson White disease. You can find his ECG just behind.



Your initial treatment is…	And then, you learn that…	In relation to this new circumstance, your initial treatment is…				
Adenosine ® 6 mg iV bolus	The patient is treated with Carbamazepin ® for epilepsy	-2	-1	0	+1	+2
Amiodarone ® 300 mg iV infusion in 30 minutes	The patient took an Sotalol ® 80 mg tablet 10 minutes ago	-2	-1	0	+1	+2
Cardioversion 50 J after sedation	The patient has just eaten. BP 10/6 Saturation 96%	-2	-1	0	+1	+2

-2 totally contraindicated; -1 less appropriate or less useful
0 The new circumstance doesn't influence your initial treatment
+1 More appropriate or even helpful; +2 absolutely necessary
Comment: Each test question consisted of three parts/columns. The first column concerned a diagnosis, treatment option or a further examination relevant to the situation described in the vignette and the ECG ("Your initial treatment is…"). The second column presented new information that might have influenced on the initial proposition (diagnose, treatment, or complementary examination…). The third part consisted of a five-point Likert scale on which the responder rated his reaction for the initial proposition regard this new element.

complementary examination (ie, 'And then you learn that…'). The third part consisted of a five-point Likert-type scale on which the student/resident rated his/her reaction to the information presented for that item (ie, 'In relation to this new circumstance, your initial treatment is…'). It was also clearly specified in the instructions that each vignette included three to four questions that were independent of each other. The goal of the questions was not to determine the additive effect of a series of clinical information elements, but to determine the effect of an isolated element of clinical information on the student's/resident's diagnosis, treatment or complementary examination.[3] Thus, a 27-case SCT-ECG composed of 64 questions was developed. In all, 24 questions concerned diagnosis, 21 questions were related to treatment options and 19 were related to further examinations. There were several levels of difficulty for the questions, which were mixed throughout the test as proposed by Fournier *et al*.[3] In addition, the questions were reviewed for content validity by two emergency physicians and a cardiologist.

Scoring method

The SCT was scored using an aggregate method described by Norman and Norcini *et al*, which takes into account the variability of responses of experienced clinicians to particular clinical situations.[9 10] Thus, for the construction of the SCT answer key, we sought the input of a panel of 12 experienced emergency physicians.[3 11] The scoring process was based on the principle that any answer given by this group of physicians reflected the opinion of an expert. Expert and expertise's definitions are still debating.[12] So as proposed in a recent publication, we decided, for this test, that an expert is a graduate emergency physician with at least a daily expertise in emergency care for 3 years.[12] Therefore, when unanimous agreement could not be reached among the experienced practitioners, it did not mean that an answer should be discarded.[13] In other words, all answers given by an expert were considered to have an intrinsic value, even if other experienced physicians did not agree with them.[14] An example of the scoring method is illustrated in figure 2, and expert responses are illustrated in figure 1 (second

Likert-scale	-2	-1	0	+1	+2	Comments
Response's number among the panel	4	8	0	0	0	Identification of the most frequent answer (in this example: -1)
Scoring creation	4/8	8/8	0/8	0/8	0/8	Division by the number of members who have given the most frequent answer (in this example: 8 experts)
Item's credit	0.5	1	0	0	0	Points obtained by the student for this item. If the responder answered -2, he get 0.5 credit If the responder answered -1, he get 1 credit

Figure 2 Scoring methods: aggregate scores. Experts' responses for the second question in figure 1.

question). The credit assigned to a particular answer was dependent on the number of experienced practitioners who gave that same response. Based on this, the maximum score for a given item was 1 point, and answers that did not match any of the responses of experienced practitioners received no credit. In this way, some answers were scored a fraction of a point based on partial agreement with experts. Scores obtained on each question were summed in order to assign a total test score, which was then divided by the number of questions and multiplied by 100 to get a percentage score.[3]

For each item, the credit given to each possible answer depended on the number of experienced practitioners who gave that response. The maximum credit for a given item was 1 point. Other responses were assigned a fraction of a point, and the responses that were not chosen by any of the experienced practitioners received no credit. Scores obtained on each question were added to obtain a total score for the test. This number was then divided by the number of questions and multiplied by 100 to get a percentage score.[10]

A disadvantage of the aggregate method is that examinees often have difficulty interpreting their scores in isolation. The scoring schemes of tests with single-right-answer formats, such as multiple choice questionnaires, provide examinees with an intuitive appreciation of their achievement; for example, a score of 67 clearly indicates that an examinee has given correct responses to 67% of the test questions. With the aggregate scoring method, SCT scores reflect concordance with those obtained by members of a reference panel. Since scoring is highly dependent of the panel used, for these scores to be meaningful, it is therefore necessary to report the value of the test panel's mean and SD.

Material and respondents

The SCT-ECG was administered online to students and residents during their emergency medicine clerkship rotations. These students were in the 4th–7th-year of their medical formation, and the residents were certified physicians who had made their specialisation in internal medicine or emergency medicine. In contrast, the expert physicians selected for the SCT were full-time emergency physicians with more than 3 years of experience following certification in the field of emergency medicine.

The experienced practitioners answered the SCT-ECG under the same conditions as students/residents. They were given 2 h to complete the test, using only their medical knowledge. The SCT-ECG was created within Adobe Acrobat X Pro and was sent by email to all respondents. Students and residents took the test within a hospital computer room during two organised sessions. Experts received the test via mail and were asked to

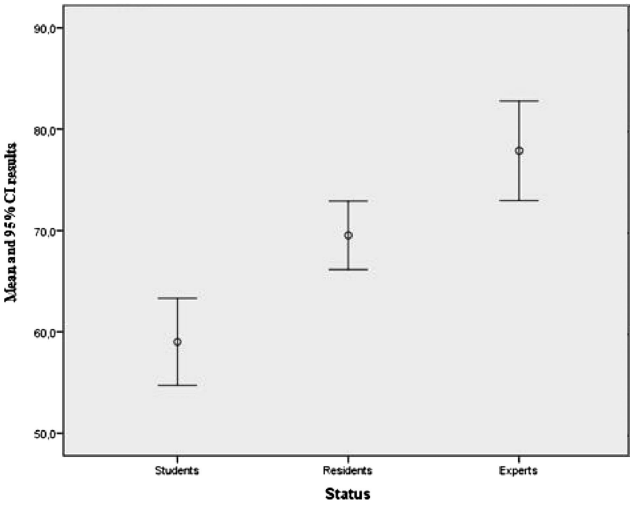


Figure 3 Table of the mean and 95% CI of respondents.

respond under the same condition as the students and residents without any theoretical support. The scoring procedure was based on the data obtained from a panel of 12 expert emergency physicians. The Excell corrector was used to determine the scores of all respondents. This program is available as a free download online (http://www.cpass.umontreal.ca/sct.html).

Statistical analysis

Statistical analyses focused on determining the internal reliability (Alpha Cronbach, Excell corrector) of the SCT-ECG and its ability to discriminate between the different groups (ANalysis Of VAriance (ANOVA), followed by Tukey's post hoc test, SPSS V.19, IBM).

RESULTS

The SCT-ECG was administered to 21 medical students, 19 residents and 12 experienced emergency physicians. The internal reliability of the test was satisfactory (Cronbach's $\alpha=0.80$). The mean results were: 59.01% (95% CI 54.71 to 63.13) for medical students, 69.53% (95% CI 66.15 to 72.91) for residents and 77.88% (95% CI 72.96 to 82.79) for emergency physicians.

Figure 3 illustrates the descriptive statistics of the interval confidence extracted from SPSS.

The ANOVA test revealed statistically significant differences among the three groups (emergency physicians, students and residents; $F_{0.271}=21.07$; $p<0.0001$). Post hoc analysis indicated significant differences between students and residents ($p<0.001$), students and emergency physicians ($p<0.001$), and residents and experts ($p=0.017$).

DISCUSSION

This SCT-ECG is a valid tool as it presents a high internal reliability and is able to discriminate among the different groups based on their level of expertise. In medicine, and especially in emergency medicine, SCTs provide unique data in comparison with current examination tools such as multiple choice questions (MCQs) or short open-answer questions (SOAPs).[15] This is because SCTs directly assess the degree of agreement between examinees and experts, while MCQs have often been considered to test facts, and SOAPs test logical thinking, judgment and application.[16 17] Moreover, SCTs, as opposed to MCQs, were shown to discriminate between novices and experts in terms of

the level of expertise in the field of emergency medicine.[15] Furthermore, SCTs can assess clinical reasoning in the context of uncertainty,[2 5] which is another advantage of SCTs as compared with other tools (MCQs or SOAPs) that more effectively assess factual knowledge.[16]

Finally, in contrast to other assessment tools that use aggregate scoring methods,[10] a unique characteristic of SCTs is the use of expert responses for scoring purposes.[3 11] Each expert's answer is considered when calculating the student's score. This is particularly applicable to emergency scenarios, which are ambiguous situations with a poor consensus.[3 6] Also, SCTs can differentiate among different levels of expertise.[4 11]

As reported by Fournier *et al* (for MCQs and SCTs) and Palmer *et al* (for SOAPs, MCQs and objective structured clinical examination), a better reliability can be obtained though combined use of various assessment tools.[15 18] Thus, this SCT-ECG could be used along with other tools, such as SOAPs and MCQs, for evaluating students/residents during their emergency clerkship rotation.

Unlike what has been described in other studies, we found that the results of the three groups (students, residents and experts) fell within the same value range.[2 19] All assessment tools require a long period of time for development, and the examinations and scoring processes can also be time-consuming.[18] Thus, other classic assessment tools, such as SOAPs, are laborious, and their scores are not always reliable.[18] One of the main limitations of SCTs is the time required for their construction. However, an SCT can be designed using computer software such as Excell Corrector, which allows the results to be more easily calculated. Here, another limitation involved the difficulty of recruiting experts since emergency medicine is a young specialty in our country. Finally, this study was limited in that it was performed in a single centre. Further investigation should be conducted in order to compare this ECG-SCT tool with other formats such as SOAPs, MCQs or illness scripts.

CONCLUSIONS

This SCT-ECG tool assesses clinical reasoning in the context of uncertainty. The use of SCTs has increased in several medical specialties; however, they have been relatively poorly used in emergency medicine, where uncertainty is an important component. In fact, the use of aggregate scoring of an expert panel closely reflects the reality of emergency medical practice. Thus, this method for evaluating clinical reasoning is valuable for the formation and assessment of future emergency physicians. We propose that the SCT-ECG is a valid tool which can complement existing tests currently administered to students and residents throughout Europe.

Acknowledgement The authors wish to thank the voluntary respondants and experts.

Funding The authors were supported by a fellowship in Medical Pedagogy from the Mont-Godinne Foundation.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 Sklar DP, Hauswald M, Johnson DR. Medical problem solving and uncertainty in the emergency department. *Ann Emerg Med* 1991;20:987–91.
2 Humbert AJ, Besinger B, Miech EJ. Assessing clinical reasoning skills in scenarios of uncertainty: convergent validity for a Script Concordance Test in an emergency medicine clerkship and residency. *Acad Emerg Med* 2011;18:627–34.
3 Fournier J-P, Demeester A, Charlin B. Script concordance tests: guidelines for construction. *BMC Med Inform Decis Mak* 2008;8:18.
4 Lubarsky S, Charlin B, Cook DA, *et al*. Script concordance testing: a review of published validity evidence. *Med Educ* 2011;45:329–38.
5 Carrière B, Gagnon R, Charlin B, *et al*. Assessing clinical reasoning in pediatric medicine: validity evidence for a script concordance test. *Ann Emerg Medicine* 2009;53:647–52.
6 Charlin B, Roy L, Brailovsky C, *et al*. The script concordance test, a tool to assess the reflective clinician. *Teach Learn Med* 2000;12:189–95.
7 Charlin B, van der Vleuten C. Standardized assessment of reasoning in context of uncertainty: the script concordance approach. *Eval Health Prof* 2004;27:304–19.
8 Lux RL. Uncertainty of the electrocardiogram: old and new ideas for assessment and interpretation. *J Electrocardiol* 2000;33(Suppl):203–8.
9 Norman GR. Objective measurement of clinical performance. *Med Educ* 1985;19:43–7.
10 Norcini J, Shea J, Day S. The use of the aggregate scoring for a recertification examination. *Eval Health Professions* 1990;13:241–51.
11 Gagnon R, Charlin B, Coletti M, *et al*. Assessment in the context of uncertainty: how many members are needed on the panel of reference of a script concordance test? *Med Educ* 2005;39:284–91.
12 Lubarsky S, Dory V, Duggan P, *et al*. Script concordance testing: from theory to practice: AMEE Guide No. 75. Medical Teacher, Early Online1–10, 2013.
13 Charlin B, Tardif J, Boshuizen HPA. Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. *Acad Med* 2000;75:182–90.
14 Sibert L, Darmoni SJ, Dahamna B, *et al*. On line clinical reasoning assessment with Script Concordance test in urology; results of a French pilot study. *BMC Med Educ* 2006;6:45.
15 Fournier J-P, Thiercelin D, Pulcini C, *et al*. Clinical reasoning assessment in emergency medicine: script concordance tests are more efficient to detect clinical experience than rich-context multiple choice questions. *Pédagogie Médicale* 2006;7:20–30.
16 Stratford P, Pierce-Fenn H. Modified essay question. *Phys Ther* 1985;65:1075–9.
17 Wass V, Van der Vleuten C, Shatzer J, *et al*. Assessment of clinical competence. *Lancet* 2001;357:945–9.
18 Palmer E, Duggan P, Devitt P, *et al*. The modified essay question: its exit from the exit examination? *Medical Teacher* 2011;32:e300–7.
19 Kania R, 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.

Assessing clinical reasoning using a script concordance test with electrocardiogram in an emergency medicine clerkship rotation

Caroline Boulouffe, Bruno Doucet, Xavier Muschart, Bernard Charlin and Dominique Vanpee

Emerg Med J 2014 31: 313-316 originally published online March 28, 2013
doi: 10.1136/emermed-2012-201737