

CONSTRUCTION AND EVALUATION OF A SCRIPT CONCORDANCE TEST FOR INTERNAL MEDICINE RESIDENTS

Moeen-uz-Zafar^{1*}, Homaidan T. Alhomaidan²

1- Assistant professor, Department of Medicine, 2 -Assistant professor, Department of Family and Community Medicine, College of Medicine, Qassim University, Saudi Arabia

*Email id of corresponding author- m.zafar@qumed.edu.sa

Received: 28/09/2017

Revised : 01/11/2017

Accepted: 05/11/2017

ABSTRACT

Background: Recently there has been a trend of using a new method of evaluation for the medical students and junior doctors. This is called ‘Script concordance test’ (SCT). It helps to check the concepts and decision making power of a physician. **Materials and Methods:** A cross-sectional study, conducted from December 2014 till April 2015. We constructed 5 scenarios in the field of Internal Medicine according to current guidelines for script concordance test. The questions were from the specialties of nephrology; pulmonology; gastroenterology; neurology and cardiology. Each scenario was followed by three questions, hence there were 15 questions. A total of 26 Internal Medicine residents and registrars were tested with these questions. A Performa was provided to the residents at the end of the test to evaluate the testing strategy and give their opinion. **Results:** The number of examinees was 26 with 16 residents and 10 registrars. All of them answered the 15 questions. The mean score for specialists and consultants was 77.7 ± 14.2 and for the examinees it was 50.4 ± 15.1 ($p < 0.05$). The Cronbach’s alpha for the test was 0.62. Item analysis showed 2 fair and 13 good items. A large proportion i.e. 76.9% (20/26) of the examinees did not know about the existence of the SCT. Those who were satisfied with the format of the test were 57.7% (15/26), whereas 53.8% (14/26) did not understand the instructional value of the SCT. **Conclusions:** Script concordance test is a useful tool for the evaluation of internal medicine residents though not easy to construct.

KEYWORDS: Internal medicine, residents, script concordance test

INTRODUCTION

Evaluation of the knowledge and skills of a medical student or a junior doctor is imperative in medical education. It not only assesses the competency but enhances it too (1-4). There are many possible ways to evaluate the competency of a medical student or a young graduate, like multiple choice questions, short and long essay type questions, objective structured clinical examination etc. They assess different levels of

Bloom’s taxonomy and all have been found useful (5-7). Clinicians are faced with complex situations in their day to day practice; they develop networks of knowledge to solve these problems. For example when a patient comes with chest pain a physician mobilizes his network of knowledge for this problem like the possible underlying disorders, specific questions to be asked in history, examination technique to

be applied, relevant investigations and management plans. These networks modify with the results obtained during each step e.g. a young patient presents with unilateral headache of two days duration associated with vomiting, the concerned physician mobilizes his network of knowledge about migraine but during examination the patient is found to have fever and slight neck stiffness, this new information changes the paradigm and the physician mobilizes another 'script' of information in his mind. This process of clinical reasoning makes the clinician a better judge of the situation (8-10).

This clinical reasoning process can be assessed by a relatively new form of evaluation called script concordance test. Here, the knowledge networks of a junior are matched to that of experts. It not only assesses the ability to decide correctly in un-certain situations but also develops the skill to shift knowledge scripts when some new piece of information 'pops up' (11-14).

SCT has been found valid and reliable tool in the assessment of examinee's skill to confirm or reject a clinical hypothesis in the setting of a clinical encounter (15-17). An SCT is drafted from a real and ambiguous clinical situation. A short clinical scenario is formulated and a series of choices regarding diagnosis, further investigation or management are asked. Then some new information is provided that may challenge the initial choice, the responses are recorded on a Likert scale (ranging from -2 to +2) (12, 13).

The exam can be run on an electronic platform or as a paper based test (18). The scoring is done in a very interesting way. A panel of experts, ideally 10 or more, are given the same scenarios.

Let us say that there is a panel of 10 experts and 8 of them choose one answer then the value of this choice would be 8/10 i.e. 0.8; if 4 choose another answer then the value of that answer would be 4/10 i.e. 0.4; if 2 experts choose another one then the value of that would be 2/10 i.e. 0.2; other choices would carry a weight age of 0. The examinees would be marked according to that scale (12, 13).

To summarize, the key features of a SCT are, firstly, examinees have to choose between possible (but having variable importance) options in an ambiguous clinical situation and that reflects ability to use complex clinical reasoning skills. Secondly, and most importantly, there is no single correct answer but the variability of expert opinions is reflected in the weight age of each option.

OBJECTIVES

Since this form of testing has not been experimented much in the subject of internal medicine at resident level, we conducted a study to construct and run a script concordance test for the internal medicine residents. Moreover we planned the evaluation of the script concordance testing strategy by the examinee residents.

MATERIALS AND METHODS

We conducted a cross-sectional study, from December 2014 till April 2015 at Qassim University College of medicine (QUCOM) and King Fahad Specialist Hospital (KFSH), Buraidah, Saudi Arabia. Ethical approval was obtained from the QUCOM research ethics committee.

The panel of experts was composed of 5 consultants from the hospital and 5 faculty members from the department of medicine, QUCOM. The experts included 5 internists; 2 nephrologists; 1 neurologist; 1 gastroenterologist and 1 cardiologist.

We constructed 5 clinical vignettes from difficult cases coming in the practice of the experts. They belonged to the specialty of cardiology; pulmonology; gastroenterology; nephrology and neurology. There were 3 questions per vignette, hence a total of 15 questions. There were 9 questions addressing investigations and 6 about management. (Appendix A)

Construct and face validity of the questions was determined by another committee of experts in medical education and internal medicine.

A questionnaire was designed to take the opinion of junior doctors about the testing strategy of SCT. (Appendix B)

A total of 26 Internal Medicine residents and registrars working at KFSH were tested with these questions.

We used online Google forms to conduct the test. The SCT calculator available at the University of Montreal website was used to compute the results. SPSS21 was used to analyze the data of Performa.

RESULTS

The number of examinees was 26 with 16 residents and 10 registrars. The mean score for the panel of experts was 77.7 ± 14.2 and for the examinees it was 50.4 ± 15.1 ($p < 0.05$). The Cronbach's alpha for the test was 0.62. The items

analysis showed 13 good items and 2 fair items (with a bad correlation of 0.05 and good correlation of 0.2).

Regarding the knowledge about SCT, 84.5% (14/16) residents and 60% (6/10) registrars did not know about its existence ($p > 0.05$). When asked about its role in their future practice, 70% (7/10) registrars considered SCT to be not useful for their future practice, whereas 50% (8/16) residents thought so ($p < 0.05$). For high stake examinations, 69% (11/16) residents and 70% (7/10) registrars did not recommend it to be a part of SMLE ($p > 0.05$) this means that they were unanimous in this point. When asked about experience of this exam, (90%) 9/10 registrars whereas 56.25% (9/16) residents were uncomfortable with this type of exam ($p < 0.05$). In the opinion of 81% (13/16) residents and only 20% (2/10) registrars, they were satisfied to have participated in the SCT ($p < 0.05$) again stressing that at earlier stage of training the trainees in internal medicine may be satisfied with SCT.

DISCUSSION

We constructed the SCT as per the guidelines available (11-13). It is recommended that the scenarios have some element of ambiguity (12) to challenge the clinical reasoning of candidates, we asked the hospital consultants to provide details of real cases, where they faced some uncertainty in decision making as regards diagnosis, further investigations or management. A test blue print was created beforehand, incorporating some most common medical problems like chest pain, cough, weakness, diarrhoea, jaundice, hypertension, loss of vision etc. After collecting the cases, the authors wrote the scenarios and the related questions. The scenarios which were lacking in ambiguity and straight forward answers were discarded. This

process was found to be difficult and time consuming. It needed a lot of support from the hospital staff.

A brochure about the SCT details, objectives and some related examples was circulated among 15 hospital consultants and 10 faculty members of QUCOM. Only 5 consultants from the hospital and 5 faculty members from the department of medicine agreed to be a part of this experiment. The experts included 5 internists; 2 nephrologists; 1 neurologist; 1 gastroenterologist and 1 cardiologist. In our opinion, this is also a cumbersome part in the formulation of SCT. It is mentioned that the panel of experts to be at least 15-20 in number (12). In our experience, this is not easy, as the staff and faculty are overwhelmed by other duties and it takes effort to introduce a new form of evaluation to them and then ask them to participate in its conduction. A composite panel has been suggested to give high psychometric quality of the SCT and we were lucky to have a panel composed of multiple specialties (19).

SCT has been run in different specialties like radiology, radiation oncology, paediatric emergency medicine, neurology, urology, geriatric medicine and rheumatology with good validity and reliability (20-26). We used it as a tool of evaluation for internal medicine residents. We circulated the brochure for SCT information and examples to 30 internal medicine residents and registrars. They were asked to provide their e-mail addresses. On specified date and time the test was sent to their e-mails as 'Google form' and responses were collected for an hour. It was a smooth and easy process. Researchers have found it easy to run a paper based or electronic platform for SCT. Only 4 junior doctors did not participate as they were occupied by other work. Our participation rate was satisfactory, whereas

people have reported lower participation rates in SCT (26). A similar process was used for the expert panel.

All the participants answered all the questions; hence our response rate was very high although literature has reported response rate of 94% to as low as 9 % (27, 28).

The data were entered in the SCT calculator available at the University of Montreal website (www.cpass.umontreal.ca/tcs.html) to compute the results and the analysis of the psychometric properties. It has been recommended to set the passing score at 2 to 4 SD below the composite marks of the experts (12), but since we had a small sample with high SD hence we set the passing score at 2 SD below the experts. With this limit ($77.7-28.4=49.3$), the number of candidates passing the exam was 14/26 i.e. 54%. The average score of registrars (n=10) was 55.4% and that of residents was 47.2% indicating that the progress in training leads to better performance in the SCT proving its discriminative value. This has been supported by the study of Brailovsky et al. (29) and Humbert et al. (30).

The psychometric properties of the test were found to be satisfactory in our research with fair reliability (Cronbach's Alpha of 0.62) and good construct validity indicated by higher mean score of the panel of experts as to that of candidates. Several studies have indicated such results like Lambert et al found a Cronbach's alpha of 0.90 among radiation oncology candidates and Mathieu et al found it to be 0.82 in a rheumatology examination (21, 26). The quality of our test was satisfactory, as indicated by 13 good items and 2 fair items (with a bad correlation of 0.05 and good correlation of 0.2) as computed by the SCT calculator.

As regards the threats to the validity indicated in the study of Lineberry et al. (31), we failed to analyze using the test-retest method and depended on the Cronbach's alpha. As suggested in the mentioned study, candidates tend to opt for the central choice and we found a similar trend in our study (Figure 1), we plan to overcome this by reducing the likert scale to 3 choices. Another limitation of our study was the small number of scenarios and questions, but since it was probably the first experiment in Arabic peninsula to have a look at the process of SCT construction, running and acceptability; this can be ignored to some extent.

Regarding the evaluation of SCT by the juniors, 77% of them did not know about the existence of this form of evaluation and 57.6% were uncomfortable with this strategy. This is in concordance with other studies and this can be overcome by establishing the rules for SCT and giving clear instructions and information to the candidates. An interesting observation was that the doctors in internal medicine at higher level of training (registrars) were relatively unsatisfied with this type of assessment and failed to find it useful for their future practice whereas those at earlier stages of training (residents) found it useful to develop their skills and were satisfied to have participated in SCT ($p < 0.05$), this supports the idea of introducing the SCT at earlier stages of training in internal medicine. The idea of introducing SCT in high stake summative assessment like SMLE was rejected by almost 70% of candidates (residents: registrars) ($p > 0.05$) we recommend to introduce it only as a formative or learning process in internal medicine training. Our findings support the study of Mathieu et al. (26) and Nseir et al. (32).

CONCLUSION

To the best of our knowledge this is one of the pioneering studies about script concordance test in KSA and the whole Arab peninsula. We find SCT a useful tool for the evaluation of internal medicine residents though not easy to construct. The most difficult is the development of vignettes and questions. Recruiting a panel of experts may be a time consuming and costly exercise. The examinees are generally satisfied with this form of test but hesitate to recommend it for higher level exams. We need to educate the candidates in this form of testing and probably apply it as a learning tool at earlier stages of training.

REFERENCES

1. Schuwirth LW, Van der Vleuten CP. Different written assessment methods: what can be said about their strengths and weaknesses? *Med Educ* 2004; 38: 974-9.
2. Amin Z, Chong YS, Khoo HE. *Practical guide to medical student assessment*. New Jersey: World Scientific Publishing Co; 2006.
3. Van der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. *Med Educ* 2005; 39: 309-17.
4. Van Der Vleuten CP. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ Theory Pract* 1996; 1: 41-67.
5. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med* 1990; 65: S63-7.
6. Bloom BS. *Taxonomy of educational objectives. Handbook I: cognitive domain*. New York: David McKay; 1956.

7. Bordage G, Page G. An alternate approach to PMPs, the key feature concept. In: Hart I, Harden R, editors. Further developments in assessing clinical competence. Montreal: Can-Heal Publications; 1987.pp. 57-75.
8. Charlin B, Van der Vleuten C. Standardized assessment of reasoning in contexts of uncertainty: the script concordance approach. *Eval Health Prof* 2004; 27: 304-19.
9. Farmer EA, Page G. A practical guide to assessing clinical decision-making skills using the key features approach. *Med Educ* 2005; 39: 1188-94.
10. Hall KH. Reviewing intuitive decision-making and uncertainty: the implications for medical education. *Med Educ* 2002; 36: 216-24.
11. Dory V, Gagnon R, Vanpee D, Charlin B. How to construct and implement script concordance tests: insights from a systematic review. *Med Educ* 2012; 46: 552-63.
12. Lubarsky S, Dory V, Duggan P, Gagnon R, Charlin B. Script concordance testing: from theory to practice: AMEE guide no. 75. *Med Teach* 2013; 35: 184-93.
13. Fournier JP, Demeester A, Charlin B. Script concordance tests: guidelines for construction. *BMC Med Inform Decis Mak* 2008; 8: 18.
14. Gagnon R, Charlin B, Lambert C, Carriere B, Van der Vleuten C. Script concordance testing: more cases or more questions? *Adv Health Sci Educ Theory Pract* 2009; 14: 367-75.
15. Lubarsky S, Charlin B, Cook DA, Chalk C, Van der Vleuten CP. Script concordance testing: a review of published validity evidence. *Med Educ* 2011; 45: 329-38.
16. 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: 248-51.
17. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med* 2006; 119: 166.e7-16.
18. Sibert L, Darmoni SJ, Dahamna B, Weber J, Charlin B. Online clinical reasoning assessment with the script concordance test: a feasibility study. *BMC Med Inform Decis Mak* 2005; 5: 18.
19. Pleguezuelos EM, Hornos E, Dory V, Gagnon R, Malagrino P, Brailovsky CA, et al. Script concordance testing in continuing professional development: local or international reference panels? *Adv Health Sci Educ Theory Pract* 2013; 18: 779-85.
20. Charlin B, Brailovsky CA, Brazeau-Lamontagne L, Samson L, Leduc C, Van der Vleuten C. Script questionnaires: their use for assessment of diagnostic knowledge in radiology. *Med Teach* 1998; 20: 567-71.
21. Lambert C, Gagnon R, Nguyen D, Charlin B. The script concordance test in radiation oncology: validation study of a new tool to assess clinical reasoning. *Radiat Oncol* 2009; 4: 7.
22. Carriere 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: 647-52.
23. Lubarsky S, Chalk C, Kazitani D, Gagnon R, Charlin B. The script concordance test: a new tool assessing clinical judgement in neurology. *Can J Neurol Sci* 2009; 36: 326-31.
24. Sibert L, Charlin B, Corcos J, Gagnon R, Lechevallier J, Grise P. Assessment of clinical reasoning competence in urology with the script concordance test: an exploratory study across two sites from different countries. *Eur Urol* 2002; 41: 227-33.

25. Piovezan RD, Custodio O, Cendoroglo MS, Batista NA, Lubarsky S, Charlin B. Assessment of undergraduate clinical reasoning in geriatric medicine: application of a script concordance test. *J Am Geriatr Soc* 2012; 60: 1946-50.
26. Mathieu S, Couderc M, Glace B, Tournadre A, Malochet-Guinamand S, Pereira B, et al. Construction and utilization of a script concordance test as an assessment tool for DCEM3 (5th year) medical students in rheumatology. *BMC Med Educ* 2013; 13: 166.
27. Braithwaite D, Emery J, De Lusignan S, Sutton S. Using the internet to conduct surveys of health professionals: a valid alternative? *Fam Pract* 2003; 20: 545-51.
28. Eysenbach G, Wyatt J. Using the internet for surveys and health research. *J Med Internet Res* 2002; 4: E13.
29. 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. *Med Educ* 2001; 35: 430-6.
30. 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. *Med Teach* 2011; 33: 472-7.
31. Lineberry M, Kreiter CD, Bordage G. Threats to validity in the use and interpretation of script concordance test scores. *Med Educ* 2013; 47: 1175-83.
32. Nseir S, Elkalioubie A, Deruelle P, Lacroix D, Gosset D. Accuracy of script concordance tests in fourth-year medical students. *Int J Med Educ* 2017; 8: 63-9.

APPENDICES

Appendix A

Example of a vignette and questions:

A 23 year old man has recently been diagnosed with hypertension.

If you were thinking of	And you find	The hypothesis becomes
Q.1 Starting ramipril	Serum K ⁺ is 5 mmol/l	-2; -1; 0; +1; +2
Q.2 Starting propranolol	Elevated serum triglycerides	-2; -1; 0; +1; +2
Q.3 Starting Verapamil	Slight pitting edema of both lower limbs	-2; -1; 0; +1; +2

(-2 strongly contraindicated; -1 contraindicated; 0 neither more nor less indicated; +1 indicated; +2 strongly indicated)

Appendix B

Examinee's assessment questionnaire about the script concordance test

1. Did you know about the existence of the script concordance tests?

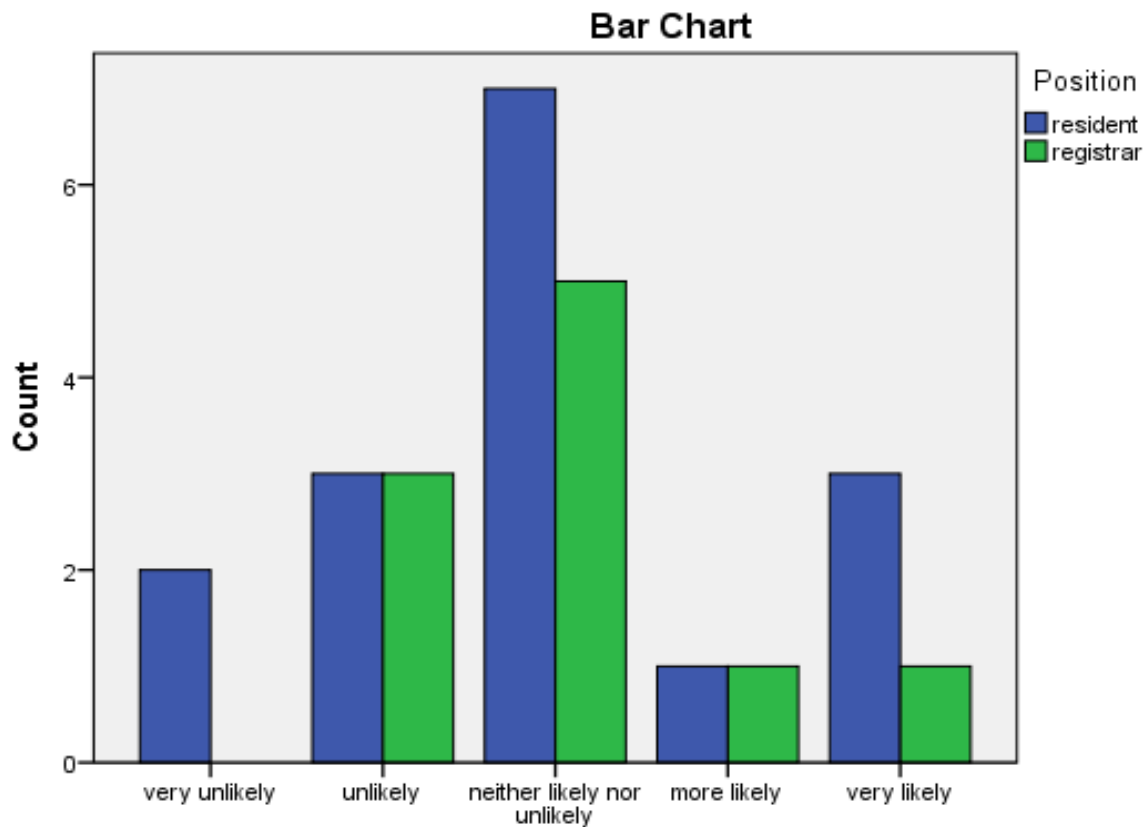
No Yes

2. Are you satisfied to have participated in this test?

No, not at all
 No, not really
 Yes, somewhat
 Yes, completely

3. Did you understand the instructional value of taking such a test?
 - No, not at all
 - No, not really
 - Yes, somewhat
 - Yes, completely
4. Did you feel at ease when completing this test?
 - No, not at all
 - No, not really
 - Yes, somewhat
 - Yes, completely**
5. Were you uncomfortable with the format of the questions?
 - No, not at all
 - No, not really
 - Yes, somewhat
 - Yes, completely**
6. Do you think this type of test is useful for your future medical practice?
 - No, not at all
 - No, not really
 - Yes, somewhat
 - Yes, completely**
7. Do you think this type of test should be part of the SMLE?
 - No, not at all
 - No, not really
 - Yes, somewhat
 - Yes, completely**
8. If you were invited to participate in another script concordance test, would you accept?
 - No, not at all
 - No, not really
 - Yes, somewhat
 - Yes, completely**
9. Would you recommend this test to your colleagues who have not yet taken it?
 - No
 - Yes

FIGURES 1



if you were thinking of ordering a LP with spinal fluid analysis and you find asymmetric weakness more pronounced in proximal muscles of lower limbs and distal muscles of upper limbs with absent position sense in the big toe of left foot