

Psychometric characteristics of script concordance test (SCT) and its correlation with routine multiple choice question (MCQ) in internal medicine department.

Mitra Amini¹, Amene Shahabi^{2*}, Mohsen Moghadami³, Mesbah Shams⁴, Amir Anooshirvani⁵, Hossein Rostamipour⁶, Javad Kojuri¹, Marzie Dehbozorgian¹, Parisa Nabeiei¹, Mohamad Jafari¹, Shirin Ghanavati⁷, Bernard Charlin⁸

¹Clinical Education Research Center, Shiraz University of Medical Sciences, Shiraz, Iran

²Shiraz Medical School, Shiraz University of Medical Sciences, Shiraz, Iran

³Non-Communicable Disease Research Center, Shiraz University of Medical Sciences, Shiraz, Iran

⁴Endocrinology and Metabolism Research Center, Shiraz University of Medical Sciences, Shiraz, Iran

⁵Internal Medicine Department, Shiraz University of Medical Sciences, Shiraz, Iran

⁶Jahrom University of Medical Sciences, Jahrom, Iran

⁷Iran University of Medical Sciences, Center for Educational Research in medical sciences, Tehran, Iran

⁸Medical School, University of Montreal, Montreal, Canada

Abstract

Background: Clinical reasoning is defined as the applying knowledge and expertise for solving clinical problems. This ability has the major role in physician abilities in diagnosis and management of diseases. One of the major methods in clinical reasoning assessment is SCT. In this study we used script concordance test in internal medicine department. The major aim of this study is to measure psychometric qualities of SCT and determining correlation between this test and routine MCQ in internal medicine department of Shiraz medical school.

Methods: In this study 100 interns participated. Beside the routine multiple question exams taken from the students at the end of the department semester, 20 SCT questions were given to the students and they were asked to answer these questions carefully. The subjects of the questions included basic and important areas of internal medicine. The reliability, item difficulty, item discrimination, item total correlation and correlation between SCT and MCQ test was measured.

Results: The mean score of student in SCT was 11.2/20, in MCQ 14.32/20, item difficulty was 0.3-0.8, item discrimination (0.02-0.27), item total correlation was 0.04-0.6, correlation between SCT and MCQ was 0.3 and reliability of SCT was 0.71 and reliability for MCQ was 0.82.

Discussion: The result of this study showed that correlation between SCT and MCQ was intermediate. This finding emphasizes that clinical reasoning tests measures examinees' abilities in clinical decision making and problem solving while MCQs measure the examinees' knowledge. The result of reliability of SCT, item discrimination, item difficulty and item total correlation showed that this test may be used as a substituted of MCQ in clinical wards as a better measure for clinical reasoning skill.

Keywords: Script concordance test, Routine multiple choice question, Clinical reasoning.

Accepted on September 24, 2017

Introduction

The progress from novice to expert is an important process in medicine. Clinical experts must monitor this progress and determine whether students have accomplished adequate levels of ability for independent practice. Clinical reasoning is one of

the most essential skills for medical doctors to solve clinical problems [1,2].

Due to multiple choice questions as a method to measure knowledge, there is still a lack of best assessing tool for clinical reasoning. Today, assessment of clinical reasoning in utmost clinical training programs includes global ratings by

faculties who observed medical students in the clinical wards over the length of a clinical rotation [3]. Global ratings, however, tend to determine limited reliability and objectivity [4].

The Script Concordance Test (SCT) is a tool for assessing clinical reasoning in medical students [5]. The SCT is based on the principles of script theory, which developed from the cognitive psychology. Scripts are networks of knowledge in physicians' minds. Clinicians use scripts in decision for diagnosis and treatment options [6]. Script theory suggests that the definition of an expert is not only the quantity of accumulated knowledge possessed, but also how knowledge is structured in the expert's mind [7]. In clinical medicine, based on the script theory each physician uses networks of knowledge, called "illness scripts" for clinical decision making. First scripts are derived when training at medical schools is started and this would be developed in postgraduate due to increased clinical experiences [8]. When a physician is encountered to a new patient, received data (such as history, physical examination, laboratory data, etc.) activate appropriate previous networks of knowledge ("illness scripts") which direct the selection and interpretation of collected data [9]. The development of "illness scripts" allows clinical experts to make precise clinical decisions rapidly, effectively, and often with minimal conscious work, even in the context of incomplete information [6].

The SCT measure the development of "illness scripts" in medical students during their progress from novice to expert learners by comparing their performance on this test to the performance of a panel of expert clinicians. SCT has been reported in previous studies to be a reliable assessing tool in clinical reasoning field [10-15].

Table 1. Format of an SCT.

If you were considering the hypothesis	And below finding is available	This finding confirms this hypothesis
lung cancer	Normal lung CT scan	+2 +1 0 -1 -2
Chronic bronchitis	Normal lung radiography	+2 +1 0 -1 -2
Lung tuberculosis	Three times negative AFB smears	+2 +1 0 -1 -2

+2: This finding confirms this hypothesis strongly. +1: This finding confirms this hypothesis somewhat 0: This finding does not have any effect on this hypothesis. -1: This finding weakens this hypothesis somewhat. -2: This finding weakens this hypothesis strongly.

Scoring

In scoring of the SCT examinees' answers to every question are matched with the responses of a panel of experts to those questions. The SCT measures how the clinical judgments of examinees are similar with those of medical experts. Each expert completes the SCTs and the aggregated responses of the experts to each SCT forms the SCT answer grid. For achieving a good and acceptable reliability in SCTs a reference expert group of 15-20 members is required [17]. 16 physicians participated in our reference panel of experts. A five-point LIKERT scale was used to distinguish the distance from the correct answer to incorrect. The answer that was chosen by

The aim of this study was to design and validate a new SCT tests for evaluation of last year medical students and compare its efficacy with a multiple choice question exam.

Methods

Study subjects

100 last year medical students of Shiraz medical school were randomly selected. All of participants participated in exam.

Study design

Beside the routine multiple question exams taken from the students at the end of the department semester, 30 SCT questions were given to the students and they were asked to answer these questions carefully. The issues of the questions consisted of fundamentals and important aspects of internal medicine.

Table 1 gives examples of the format of an SCT. The format is based on the "Hypothetico-Deductive" (HD) reasoning model. Generated hypotheses have directed HD reasoning. Generally, while solving a clinical problem, the medical physician connects theoretical knowledge of the disease based on specific symptoms and signs of the patient [16]. In SCT the case description shows early patient cues, and the three columns below the case description, resemble to the stages of hypothesis generation.

A sample of SCT question:

Question: A 45 y old with the history of smoking 20 cigarettes a day with a 100 cc hemoptysis has referred to emergency department.

most of the experts was considered the correct answer, and the weight for other answers was determined by considering their credit and their distance from the correct answer. In this scoring system the credit for the best answer was 100%, and credit for other answers was measured based on the percentage of clinical reference panel that chose that answer. We used the method $1/(1+x)$, where x is defined by way of the distance from the correct answer (values of x alternated from a minimum of 1 to a maximum of 4). This innovative scoring system was derived from our previous research [18].

Psychometric characteristics of script concordance test (SCT) and its correlation with routine multiple choice question (MCQ) in internal medicine department

Statistical analysis

Reliability of examinees' scores was examined by Cronbach's alpha coefficient, Test optimization was done by deleting questions with item total correlations lower than 0.05 as explained by Gagnon in previous studies [19]. A Spearman nonparametric correlation was calculated to estimate the strength of the relationship between scores on SCT and MCQ. Psychometric analysis of SCT was done based on the classic measurement theory by dividing population into high level and low level students based on their scores and using Whitney and Sabers method. All p values were considered significant at ≤ 0.05 .

Results

One hundred respondents participated in this study including 55 female and 45 male students and they answered 20 questions of SCT exam within 30 min. The whole score for the exam was 20. The average time for completion of the tests was 24.5 min (range 17.3 min-30 min).

The minimum and maximum score for SCT were 6.53 and 15.9 respectively and mean score was 11.21 ± 6.79 .

The Table 2 shows discrimination index calculated with the method of Witney and Spears [20]. All of the questions achieved positive discrimination index. It should be noted that the questions that receive negative coefficient should be deleted or changed properly. Zero coefficients show that the question could not separate the high level and low level students.

Difficulty index also was calculated for questions by the method of Witney and Spears. This coefficient was also acceptable for all questions (0.3-0.7). Whatever the difficulty index is near to 1, the question is easier and far from 1 it will be considered harder. Maximum Difficulty Index was 0.8 for question 8 and the minimum was 0.41 for question 4 (Table 2).

Correlation coefficient and statistical differences for SCT questions are mentioned in Table 3.

Table 2. Discrimination and difficulty index for questions of SCT.

	Question 1	Question 2	Question 3	Question 4	Question 5	Question 6	Question 7	Question 8	Question 9
Discrimination Index	0.1	0.4	0.2	0.52	0.44	0.06	0.51	0.8	0.72
Difficulty index	0.67	0.51	0.73	0.61	0.65	0.66	0.59	0.8	0.72
	Question 10	Question 11	Question 12	Question 13	Question 14	Question 15	Question 16	Question 17	Question 18
Discrimination index	0.15	0.12	0.08	0.21	0.08	0.08	0.18	0.1	0.27
Difficulty index	0.59	0.65	0.65	0.59	0.67	0.71	0.68	0.63	0.63
	Question 19	Question 20							
Discrimination index	0.21	0.19							
Difficulty index	0.52	0.64							

Table 3. Item total correlation and statistical difference (P-value) for questions of SCT.

	Question 1	Question 2	Question 3	Question 4	Question 5	Question 6	Question 7	Question 8	Question 9
Correlation coefficient	0.04	0.04	0.06	0.16	0.18	0.03	0.19	0.02	0.09
Statistical difference	0	0	0	0	0	0.52	0	0.64	0
	Question 10	Question 11	Question 12	Question 13	Question 14	Question 15	Question 16	Question 17	Question 18
Correlation coefficient	0.44	0.37	0.29	0.53	0.32	0.71	0.68	0.63	0.6
Statistical difference	0	0	0	0	0	0.01	0	0	0
	Question 19	Question 20							
Correlation coefficient	0.49	0.41							
Statistical difference	0	0							

The correlation between SCTs' scores and MCQs' scores was intermediate (Pearson coefficient: 0.3)

Discussion

In this study, we have evaluated the psychometric characteristics of SCT in last year medical students in internal medicine department in Iran. This test has been used in different studies for assessment of clinical reasoning in medical students all over the world [19,21].

The reliability of this test was 0.71 that it is similar to our previous studies and other studies [11,18,22,23]. This coefficient shows the amount of the attention to the exam and depends on the correlation of the mentioned attention. In other words, correlation between the scores of participants is considered as an index for internal consistency. We chose a good sample for our expert panel we feel that our selected members reflect acceptable community standards of expertise in internal medicine as mentioned in previous studies [24,25].

Discrimination index was positive for most of the questions in SCT exam and its difficulty index was within 0.3-0.8. Correlation coefficient for each question in regard to whole exam was positive (0.04-0.6) and it means that designed questions can separate weak students from strong ones.

In this study an intermediate correlation was achieved between SCT exam and MCQ exam scores.

Some explanations can be stated for these findings: First, clinical reasoning measures the students' ability in data gathering, making hypothesis, evaluating of hypothesis and problem solving while in MCQ exams, only knowledge of the students are measured. Second; in our country the exams are taken using MCQ forms of the exams, so, because the SCT is new, the participants are not completely aware of that. The third reason can be this matter that in medical universities more attentions are on member of the lessons and other aspects of clinical reasoning are not considered and this matter can be the cause for moderate correlation between SCT and multiple choice questions scores. A review of validity of all published articles about SCT by Lubarsky showed that there is poor correlation between assessing factual knowledge and clinical reasoning [11].

The most important strengths of our study are an adequate number of our expert panel for scoring SCTs, and calculating difficulty index and discrimination index for each test. Limitations of the present study were pen and pencil format of the tests, the fact that we did not compare diverse scoring methods, inability to measure the impact of SCTs method on students' learning and the fact that SCT determine matching between the end responses of students with that of experts and does not measure the process of clinical problem solving of students.

Our present study shows that the SCT is a reliable and valid tool for assessing clinical reasoning in internal medicine for last year medical students. It also appears to be practical, authentic, and versatile. The SCT can be used as a valuable

tool in comparison to other traditional standardized methods for medical trainees' assessment.

Conclusion

The results of the present study indicate that SCT was a valid and reliable approach which can be a good replacement of multiple choice exams in major wards such as internal ward. With growing body of research in SCTs, this test can be applied routinely in the future. It is recommended to do more researches if SCT is going to be used in summative evaluations.

Acknowledgements

This manuscript was extracted from the thesis of the second author Amene Shahabi (proposal No. 6354). All of the budget for the present study covered by Shiraz University of Medical Sciences. We also acknowledge the sincere contribution of administrators, teachers, and specially students who participated in this study.

References

1. Russell T. Can reflective practice be taught? *Reflect Pract* 2005; 6: 199-204.
2. Montgomery K. *How doctors think: clinical judgment and the practice of medicine*. Oxford Univ Press 2005.
3. Huddle TS, Heudebert GR. Viewpoint: taking apart the art: the risk of anatomizing clinical competence. *Acad Med* 2007; 82: 536-541.
4. Lurie SJ, Mooney CJ, Lyness JM. Measurement of the general competencies of the accreditation council for graduate medical education: a systematic review. *Acad Med* 2009; 84: 301-309.
5. Charlin B, van der Vleuten C. Standardized assessment of reasoning in contexts of uncertainty: the script concordance approach. *Eval Health Profess* 2004; 27: 304-319.
6. Kalyuga S, Rikers R, Paas F. Educational implications of expertise reversal effects in learning and performance of complex cognitive and sensorimotor skills. *Edu Psychol Rev* 2012; 24: 313-337.
7. Elstein AS, Schwarz A. Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. *BMJ* 2002; 324: 729-732.
8. Suebnukarn S, Haddawy P. Modeling individual and collaborative problem-solving in medical problem-based learning. *User Model User Adapt Interact* 2006; 16: 211-248.
9. Charlin B, Boshuizen HP, Custers EJ, Feltovich PJ. Scripts and clinical reasoning. *Med Educ* 2007; 41: 1178-1184.
10. Brazeau-Lamontagne L, Charlin B, Gagnon R, Samson L, van der Vleuten C. Measurement of perception and interpretation skills during radiology training: utility of the

Psychometric characteristics of script concordance test (SCT) and its correlation with routine multiple choice question (MCQ) in internal medicine department

- script concordance approach. *Med Teach* 2004; 26: 326-332.
11. 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: 329-338.
 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: 184-193.
 13. Aldekhayel SA, ALselaim NA, Magzoub ME, AL-Qattan MM, Al-Namlah AM, Tamim H. Constructing a question bank based on script concordance approach as a novel assessment methodology in surgical education. *BMC Med Educ* 2012; 12: 100.
 14. Kow N, Walters MD, Karram MM, Sarsotti CJ, Jelovsek JE. Assessing intraoperative judgment using script concordance testing through the gynecology continuum of practice. *Med Teach* 2014; 1-6.
 15. Amini M, Kojuri J, Karimian Z, Lotfi F, Moghadami M, Dehghani M. Talents for future: Report of the second national medical science Olympiad in Islamic republic of Iran. *Iran Red Cresc Med J* 2011; 13: 377.
 16. Elstein AS, Shulman LS, Sprafka SA, Allal L. *Medical problem solving: an analysis of clinical reasoning*: Harvard Univ Press Cambridge MA 1978.
 17. 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: 284-291.
 18. Amini M, Moghadami M, Kojuri J, Abbasi H, Abadi AAD, Molaee NA. An innovative method to assess clinical reasoning skills: Clinical reasoning tests in the second national medical science Olympiad in Iran. *BMC Res Notes* 2011; 4: 418.
 19. 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: 647-652.
 20. Whitney D, Sabers D. *Improving essay examinations III: Use of item analysis*. Iowa City: University of Iowa, University Evaluation and Examination Service (Technical Bulletin No 11) 1970.
 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. *Rad Oncol* 2009; 4: 7.
 22. Iravani K, Amini M, Doostkam A, Dehbozorgian M. The validity and reliability of script concordance test in otolaryngology residency training. *Journal of Advances in Med Educ Professional* 2016; 4: 93-96.
 23. Linn AM, Tonkin A, Duggan P. Standard setting of script concordance tests using an adapted Nedelsky approach. *Med Teach* 2013; 314-319.
 24. Simmons B. Clinical reasoning: concept analysis. *J Adv Nurs* 2010; 66: 1151-1158.
 25. Rezaee R, Nabeiei P, Sagheb MM. Evaluation of the consultation program in Shiraz University of Medical Sciences. *J Adv Med Educ Professional* 2014; 2: 27-32.

***Correspondence to**

Amene Shahabi

Shiraz medical School

Shiraz University of Medical Sciences

Iran