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

Mitra Amini¹, Amene Shahabi²*, Mohsen Moghadami³, Mesbah Shams⁴, Amir Anooshirvani⁵, Hossein RostamiPour⁶, Javad Kojuri¹, Marzie Dehbozorgian¹, Parisa Nabeiei¹, Mohamad Jafari¹, Shirin Ghanavati⁷, Bernard Charlin⁸

¹Quality Improvement in Clinical 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 SC 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.

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
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.

**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 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].

---

Table 1. Format of an SCT.

<table>
<thead>
<tr>
<th>If you were considering the hypothesis</th>
<th>And below finding is available</th>
<th>This finding confirms this hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>lung cancer</td>
<td>Normal lung CT scan</td>
<td>+2 +1 0 -1 -2</td>
</tr>
<tr>
<td>Chronic bronchitis</td>
<td>Normal lung radiography</td>
<td>+2 +1 0 -1 -2</td>
</tr>
<tr>
<td>Lung tuberculosis</td>
<td>Three times negative AFB smears</td>
<td>+2 +1 0 -1 -2</td>
</tr>
</tbody>
</table>

+2: This finding confirms this hypothesis strongly. +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.
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>
<thead>
<tr>
<th>Question 1</th>
<th>Question 2</th>
<th>Question 3</th>
<th>Question 4</th>
<th>Question 5</th>
<th>Question 6</th>
<th>Question 7</th>
<th>Question 8</th>
<th>Question 9</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Discrimination Index</strong></td>
<td>0.1</td>
<td>0.4</td>
<td>0.2</td>
<td>0.52</td>
<td>0.44</td>
<td>0.06</td>
<td>0.51</td>
<td>0.8</td>
</tr>
<tr>
<td><strong>Difficulty index</strong></td>
<td>0.67</td>
<td>0.51</td>
<td>0.73</td>
<td>0.61</td>
<td>0.65</td>
<td>0.66</td>
<td>0.59</td>
<td>0.8</td>
</tr>
<tr>
<td><strong>Question 10</strong></td>
<td><strong>Discrimination Index</strong></td>
<td>0.15</td>
<td>0.12</td>
<td>0.08</td>
<td>0.21</td>
<td>0.08</td>
<td>0.08</td>
<td>0.18</td>
</tr>
<tr>
<td><strong>Difficulty index</strong></td>
<td>0.59</td>
<td>0.65</td>
<td>0.65</td>
<td>0.59</td>
<td>0.67</td>
<td>0.71</td>
<td>0.68</td>
<td>0.63</td>
</tr>
<tr>
<td><strong>Question 19</strong></td>
<td><strong>Discrimination index</strong></td>
<td>0.21</td>
<td>0.19</td>
<td>0.52</td>
<td>0.64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Question 20</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Question 1</th>
<th>Question 2</th>
<th>Question 3</th>
<th>Question 4</th>
<th>Question 5</th>
<th>Question 6</th>
<th>Question 7</th>
<th>Question 8</th>
<th>Question 9</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Correlation coefficient</strong></td>
<td>0.04</td>
<td>0.04</td>
<td>0.06</td>
<td>0.16</td>
<td>0.18</td>
<td>0.03</td>
<td>0.19</td>
<td>0.02</td>
</tr>
<tr>
<td><strong>Statistical difference</strong></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.52</td>
<td>0</td>
<td>0.64</td>
</tr>
<tr>
<td><strong>Question 10</strong></td>
<td><strong>Correlation coefficient</strong></td>
<td>0.44</td>
<td>0.37</td>
<td>0.29</td>
<td>0.53</td>
<td>0.32</td>
<td>0.71</td>
<td>0.68</td>
</tr>
<tr>
<td><strong>Statistical difference</strong></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Question 19</strong></td>
<td><strong>Correlation coefficient</strong></td>
<td>0.49</td>
<td>0.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Statistical difference</strong></td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The correlation between SCTs’ scores and MCQs’ sores 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 for the first time 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 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 it 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 or 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 choices 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 use 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

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


*Correspondence to

Amene Shahabi
Shiraz medical School
Shiraz University of Medical Sciences
Iran

Biomed Res 2017 Volume 28 Issue 19