

Studying Clinical Reasoning of Internal Medicine Residents in Medical City Campus Using Script Concordance Test

Batool A.G. Yassin¹⁾, Ali Almothaffar²⁾, Manal Aladhadh²⁾, Mutaz F. Hussein²⁾,
Zuhair Abdul Ghafour²⁾, Faiq I. Gorial²⁾

ABSTRACT

Objective: Clinical reasoning is one of the key cognitive skills used by doctors to solve medical situations. The aim was to assess if there is a difference in clinical reasoning skills between junior and senior internal medicine residents using script concordance test.

Design: A cross sectional study.

Subjects and Methods: The test was developed in Medical City Campus April 2019 to April 2020. Twenty cases covering common medical problems were created and reviewed by five faculty members. Ten experienced physicians made the reference panel for the purpose of scoring, the mean scores of reference panel, senior and junior residents were compared using ANOVA.

Results: Fifty-nine residents agreed to participate, 39 juniors and 20 seniors. The mean score was 79.09 ± 5.8 for the reference panel, 66.4 ± 6.8 for junior and 66.7 ± 6.3 for senior residents. The reliability using Cronbach alpha was 0.67.

ANOVA test showed statistically significant differences between mean scores of the panel and both first grade and final board trainee whereas the differences between mean scores of the junior and the senior residents was statistically not significant.

Conclusion: The mean score of junior residents did not differ statistically from that of senior residents in SCT. Further study is needed.

KEY WORDS

clinical reasoning, script concordance test, assessment

INTRODUCTION

Clinical reasoning is one of the most important cognitive skills used by doctors to solve every day vague medical situations specially in emergency situations. Clinical experience enhances these skills which expand over time and eventually adapt to specific cognitive tasks that clinicians routinely perform¹⁻⁴⁾.

Many assessment methods have been introduced to assess this important clinical skill. The Script Concordance Test (SCT) is an innovative assessment instrument originally developed by Charlin et al, which is based on script theory from cognitive psychology. It uses written case descriptions (SCRIPTS) and requires from the examinees to make choices and decisions about the patient problems. The process of decision making reflects real-world decision-making processes in which clinicians retrieve their 'illness scripts' or network of knowledge (about similar patient problems and presentations stored in their memory) when faced with uncertainty in a clinical presentation⁵⁾. Answers are provided on a Likert scale intended to capture the variability of responses of experts to clinical problems for which single correct answers may not be evident⁶⁾. Scoring is based on comparing the responses of individual examinees with those of a panel of expert clinicians⁶⁾. The reliability and construct validity of the SCT are well documented when used to differentiate advanced medical students, residents, and expert attending physicians⁷⁾. In addition, the SCT has been used to assess the clinical reasoning skills of medical students during their clerkship years, as well

as those of residents and fellows (3). The SCT is limited in its ability to assess how students address psychological, interpersonal, and emotional issues when dealing with clinical situations that involve ambiguity and uncertainty. Yet it can be used in formative assessment of undergraduate students and for postgraduate students as a tool to measure the progress of the students in developing their clinical reasoning skills. Rather than using the SCT outcomes as a criterion for advancement, or a relative ranking instrument, the intended purpose is to provide students with personal skills development informed by individualized feedback about clinical reasoning ability³⁾.

SCTs are challenging to develop, although they can be administered with the same ease as MCQs. Research concerning SCTs suggests that scores offer a valid reflection of diagnostic reasoning, with test performance correlating with clinical experience and in training examination scores^{6,8,9)}.

In Iraq, the general medicine specialty degree is awarded to residents after completing 4-year training board program and passing the final examination. The assessment of clinical reasoning of these trainees is achieved through single best answer-based case scenario questions and the long case, the short case and the oral examinations with all their limitations.

The aim of this study is to assess the clinical reasoning skills of the residents of internal medicine board program in Medical City Campus / Baghdad using script concordance test (SCT). The research question was whether the mean score in SCT will be different between senior and junior residents reflecting clinical reasoning skills.

Received on July 31, 2020 and accepted on October 6, 2020

1) Family & Community Medicine Department, College of Medicine, University of Baghdad
Baghdad, Iraq

2) Department of Medicine, College of Medicine, University of Baghdad
Baghdad, Iraq

Correspondence to: Ali Almothaffar
(e-mail: amjmam@yahoo.com)

Table 1: An example of an SCT vignette and related items

Q: A 33-year-old woman was admitted to the hospital suffering from thyrotoxicosis. (Scenario)

If you were thinking to start... (Treatment option)	And then you find that... (New Information)	Then your plan of action becomes... (Likert Scale)					
1. Methimazole	She had been treated previously with propylthiouracil, but she developed agranulocytosis	A	B	C	D	E	-2: Much less appropriate
2. Propranolol	she has pronounced exophthalmos	-2	-1	0	+1	+2	-1: Less appropriate
3. Radioactive iodine	She had hysterectomy for uterine fibroids 2 y ago	A	B	C	D	E	0: Neither more nor less appropriate
		-2	-1	0	+1	+2	+1: More appropriate
		A	B	C	D	E	+2: Much more appropriate
		-2	-1	0	+1	+2	

Table 2: SCT mean scores (60 items) of the ten expert panel members and residents

Participants	No.	Range	Mean Scores ± SD	Coefficient of Variation
The Expert Panel Members	10	71.5 ± 88.2	79.09 ± 5.8	7.3
First year(junior) board residents	39	46.2 ± 78.3	66.4 ± 6.8	10.2
Final year(senior) board residents	20	56.6 ± 78.5	66.7 ± 6.3	9.4

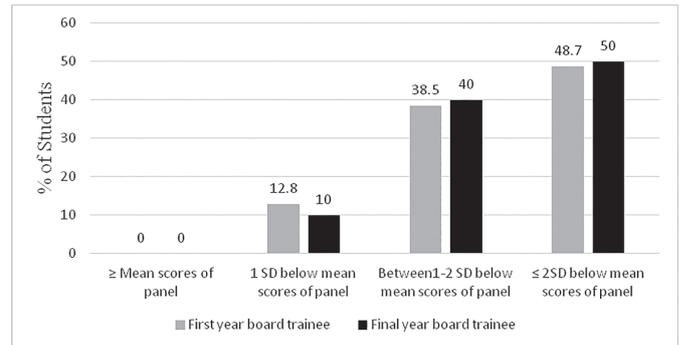


Figure 1: Standardized scores for students

SUBJECTS AND METHODS

Test Principles

The test is case based. Cases, described as short scenarios, always incorporate uncertainty. Scenarios are followed by a series of questions, presented in three parts. The first part ("if you were thinking of") contains a relevant diagnostic or management choice or opinion. The second part ("and then you were to find") presents a new clinical information that may affect this choice, such as a physical sign, an imaging study or a laboratory test result. The third part ("this option would become") is a five-point Likert scale that captures examinees' decisions that results from the mental process that incorporates the scenario, the choice and the effect of the new information. The task for examinees is to decide what effect the new finding has on the status of the option, in direction (positive, negative, or neutral) and intensity. This effect is captured with a Likert scale because script theory assumes that clinical reasoning is composed of a series of qualitative judgments⁹.

SCT Development

The SCT was developed in Medical City Campus and College of Medicine/ University of Baghdad from April 2019 to April 2020. Twenty cases covering common medical problems in internal medicine were created first separately by the members of the research committee and then they were all reviewed by the whole 5-member committee and necessary changes had been made in a collaborative way so that the cases will be compatible with the purpose of the test. The committee members all have more than 10 years' experience in undergraduate and postgraduate student teaching and assessment in the department of internal medicine.

The SCT was constructed according to the above-mentioned principles and guidelines⁹.

It was composed of vignettes based on genuine clinical medical cases including emergency cases. The case descriptions contained situations of uncertainty that cannot be answered by single-correct-answer solutions. The scoring system aimed to compare responses of the residents with those of practicing physicians who are > 10 years in the practice; this latter group formed the reference panel used to set the test's scoring grid. An example of a SCT vignette and related items is shown in Table 1. Each test item consists of a clinical vignette followed by three columns. The first column provides a hypothesis relevant to the situation described in the scenario. In the second column, a clinical information (e.g., a symptom, a sign, or a test result) is provided. The

third column contains a 5-point Likert-type scale for the examinee to use to indicate the perceived effect of this new information (column 2) on the proposed hypothesis (column 1)⁹.

Reference Panel

Ten experienced board-certified physicians were recruited from the division of internal medicine in Medical City Campus who had been practicing internal medicine and participating in residents training for > 10 years. Each panel member independently completed a paper-based version of the SCT.

Residents

The internal medicine board residents who are actively involved in Medical City Campus health services including emergency department services were eligible to participate in the study. All of them were invited to participate in the SCT and after explaining the aims of the study only those who accepted to participate were included in the study. They were given 60 minutes to complete the test.

Scoring System

There are no single-best answers to SCT items; several responses to each item may be considered acceptable. The examinee's response to each item is compared with the responses of the test's expert panel members. Credit is assigned to each response based on how many of the experts on the panel choose that response. A maximum score of 1 is given for the response chosen by most of the experts (the modal response). Other responses are given partial credit, depending on the fraction of experts choosing them. Responses that experts did not select receive 0.

Data Analysis

Microsoft Excel, 2010 and Statistical Package for Social Sciences (SPSS) version 20 were used to analyze the data. Descriptive statistics, including means, standard deviations (SDs), minimum and maximum values of participants' scores and coefficient of variation were obtained. Internal reliability was estimated using Cronbach alpha. The Z- distribution of the panel based on the mean and SD was used as a standard scale and the score of each participant was converted to that scale. Analysis of variance and Post Hoc Test (Tukey HSD) were used to test the differences between mean scores. P value of less than 0.05 was considered

Table 3: Analysis of Variance (ANOVA) for the participants' scores

Groups	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1356.310	2	678.155		
Within Groups	2810.302	66	42.580	15.926	.000
Total	4166.613	68			

statistically significant.

Ethical Approval

The study was approved by the scientific committee of the department of internal medicine/College of Medicine/University of Baghdad.

RESULTS

Ten expert panel participated in this study, their scores ranged from 71.5 - 88.2 with a mean of 79.09 ± 5.8 Standard Deviation (SD) and 7.3% Coefficient of Variation. (Table 2).

There were 80 residents eligible to participate, 59 (73.8%) of them agreed to participate. A total of 39 first year board residents and 20 final year board residents took the test and completed the 60-items (of the 20 cases) within 60 minutes. The score of the first-year board residents ranged from 46.2 - 78.3 with a mean of 66.4 ± 6.8 SD and 10.2% Coefficient of Variation.

The score of the final year board residents ranged from 56.6 - 78.5 with a mean of 66.7 ± 6.3 SD and 9.4% Coefficient of Variation. (Table 2)

The reliability using Cronbach alpha was calculated and was found to be 0.67.

On calculating the standardized scores for students it was found that none of the residents, from both grades had scores above the mean score of the panel, five residents (12.8%) from first grade and two (10.0%) from the final had scores below the mean score of the panel by one SD, 15 (38.5%) of the first and 8(40.0%) of the final had scores between one and two SD below the mean score of the panel and 19 (48.7%) of the first and 10 (50.0%) of the final had scores below 2 SD of the mean score of the panel (Figure 1).

Analysis of Variance (ANOVA) and post hoc tests showed statistically significant differences between mean scores of the panel and both first year residents and final year residents ($F = 15.926$, $P < 0.0001$) whereas the differences between mean scores of the first year residents and the final year residents was statistically not significant ($P > 0.05$) (Tables 3 and 4).

DISCUSSION

The results of this study are worth noticing although the number of participants was limited. The mean score of junior (first year) residents in the SCT did not differ much from that of senior (final year) residents which is unexpected if we take in consideration the difference in experi-

ence between the 2 groups.

In this study, the clinical reasoning of internal medicine board residents in Medical City Campus, the largest training center in Iraq, had been assessed using the recommended steps in generation, implementation and scoring of SCT.

The question format was new for the expert panel and the residents and so explanation with an example was made for them before taking the test. We chose this number and format because it seems that a test set comprising 20 cases and 60 questions, for one hour of testing time is convenient and it is advisable to ask several questions for each case, as long as those questions address critical elements⁹. It has been shown that, for high stake examinations, 15 panel members are required but for lower stake examinations, for instance formative assessment, smaller panels can be used, however panels with less than 10 persons are associated with more error in reliability estimates⁶.

The test's internal consistency and discriminative capacity to distinguish residents from experienced physicians supported construct validity⁹. The Cronbach reliability factor was 0.67 which was acceptable for the purpose of research. We tried to improve it after item deletion, but it did not make much improvement. The coefficient of variations for the 3 groups were good. The mean score of junior residents did not differ statistically from that of senior residents which in a way similar to the study of Nouh 2012, where the R5 residents showed a dip in their mean score¹⁰, while Ducos and colleagues found statistically significant difference between junior and senior anesthesia residents using SCT¹¹. The results of the current study need further study to verify their consistency and the cause behind them before making further conclusion.

The mean scores of both groups differ from that of the expert panel significantly and this can be explained by the effect of more than 10-year experience of the panel.

The generation of the test, administration and scoring was a cumbersome process as stated by others and this is an obstacle for its wide use in summative assessment in our country¹⁰.

In their review, Daniel and colleagues addressed the issue of assessment of this skill through different test formats that had been tried including multiple choice questions, essay questions, OSCE and oral exams, with all their limitations. The different methods of assessment of clinical reasoning were compared and the score of each test on the different components of clinical reasoning was estimated. The SCT scored better than MCQ, key feature exam and extended match questions but less than modified essay and short answer questions, all belonging to non-workplace-based assessments. If we want to assess many students, the essay questions are difficult to score¹². Nevertheless, MCQ examinations are used widely due to factors such as grading consistency, reliability, and simplicity in administration and grading. The cost-effectiveness of the MCQ examination is also attractive, especially when large numbers of examinations must be scored^{13,14}.

The assessment of clinical reasoning is a complex process and using multiple assessment tools is warranted since no one level or assessment tool should take precedence and clinical teachers should be prepared and trained to assess from knowledge through to performance using multiple methods to gain a more accurate picture of their learners' skills⁹. Thammasitboon and colleagues proposed Assessment of Reasoning Tool (ART) to assess this core skill to address the complex mental process used by doctors during everyday practice¹⁵.

LIMITATIONS

Limitations included limited number of participants and the modest

Table 4: Post Hoc Test (Tukey HSD) for ANOVA; Dependent Variable: Scores of the test

(I) participants' Class	(J) participants' Class	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
First Year board residents	Final Year board Trainee	-.37	1.8	.977	-4.7	3.9
	Expert Panel	-12.7*	2.3	.000*	-18.3	-7.2
Final Year board residents	First Year Board Trainee	.37	1.8	.977	-3.9	4.7
	Expert Panel	-12.3*	2.5	.000*	-18.4	-6.3
Expert Panel	First Year Board Trainee	12.7*	2.3	.000*	7.2	18.3
	Final Year board Trainee	12.3*	2.5	.000*	6.3	18.4

* The mean difference is significant at the 0.05 level.

reliability of the exam set. We suggest another study that includes larger number of residents from all training centers with a new set of question that contain more items to improve reliability before we can reach to a conclusion regarding its relation to the training program.

CONCLUSION

In conclusion, the mean score of junior residents, as an indicator of clinical reasoning ability skill, did not differ statistically from that of senior residents using SCT.

REFERENCES

- Schmidt H, Norman G, Boshuizen H. A cognitive perspective on medical expertise: Theory and implications. *Acad Med* 1990; 65: 611-621
- Lateef F. Clinical Reasoning: The Core of Medical Education and Practice. *Int J Intern Emerg Med*. 2018; 1(2): 1015
- Humbert, Aloysius & Johnson, Mary & Miech, Edward & Friedberg, Fred & Grackin, Janice & Seidman, Peggy. Assessment of clinical reasoning: A Script Concordance test designed for pre-clinical medical students. *Medical teacher*. 2011; 33: 472-7.
- Bowen J: Educational Strategies to Promote Clinical Diagnostic Reasoning. *N Engl J Med* 2006; 355(21): 2217-25.
- Piovezan RD, Custódio O, Cendoroglo MS, Batista NA, Lubarsky S & Charlin B. Assessment of Undergraduate Clinical Reasoning in Geriatric Medicine: Application of a Script Concordance Test. *Journal of the American Geriatrics Society*. 2012; 60(10): 1946-50
- Fournier, J.P., Demeester, A. & Charlin, B. Script Concordance Tests: Guidelines for Construction. *BMC Med Inform Decis Mak* 2008; 8, 18. <https://doi.org/10.1186/1472-6947-8-18>
- Charlin B, Van Der Vleuten C. Standardized assessment of reasoning in contexts of uncertainty: The script concordance approach. *Eval Health Prof* 2004; 27: 304-319.
- Ilagan JS, Humbert AJ, Kuhn G, Hansen ML, Norman GR, Eva KW, Charlin B, Sherbino J. Assessing diagnostic reasoning: a consensus statement summarizing theory, practice, and future needs. *Acad Emerg Med*. 2012 Dec; 19(12): 1454-61. doi: 1111/acem.12034.
- Thampy H, Willert E, Ramani S. Assessing Clinical Reasoning: Targeting the Higher Levels of the Pyramid. *J Gen Intern Med*. 2019 Aug; 34(8): 1631-1636. doi: 10.1007/s11606-019-04953-4.
- Nouh T, Boutros C and Ganong R, Script concordance test in surgical education, *The American Journal of Surgery* .2012; 203, 530-534
- Ducos G, Lejus C, Sztark F, Nathan N, Fourcade O, Tack I, Asehnoune K, Kurrek M, Charlin B, Minville V, The Script Concordance Test in anesthesiology: Validation of a new tool for assessing clinical reasoning. *Anaesthesia Critical Care & Pain Medicine*. 2015; 34 (1): 11-15.
- Daniel MI, Rencic J, Durning SJ, Holmboe E, Santen SA, Lang V, Ratcliffe T, Gordon D, Heist B, Lubarsky S, Estrada CA, Ballard T, Artino AR Jr, Sergio Da Silva A, Cleary T, Stojan J, Gruppen LD. Clinical Reasoning Assessment Methods: A Scoping Review and Practical Guidance. *Acad Med*. 2019 Jun; 94(6): 902-912. doi: 10.1097/ACM.0000000000002618.
- Veloski JJ, Rabinowitz HK, Robeson MR, Young PR. Patients do not present with five choices: An alternative to multiple-choice tests in assessing physicians' competence. *Acad Med*. 1999; 74: 539-546.
- White H. Problem-based testing. *Biochem Mol Biol Educ*. 2002; 30: 56.
- Thammasitboon S, Rencic JJ, Trowbridge RL, Olson APJ, Sur M, Dhaliwal G. The Assessment of Reasoning Tool (ART): structuring the conversation between teachers and learners. *Diagnosis (Berl)*. 2018 Nov 27; 5(4): 197-203. doi: 10.1515/dx-2018-0052.