Canadian Medical Education Journal

Major Contribution/Research Article

The Reliability and Validity of a Paediatric Script Concordance Test with Medical Students, Paediatric Residents and Experienced Paediatricians

Jean-Francois Lemay,[1] Tyrone Donnon[1] and Bernard Charlin[2]

University of Calgary[1] and University of Montreal[2]

Published: 30 June, 2010

CMEJ 2010, 1(2):e89-e95 Available at http://www.cmej.ca

Abstract

Background: The Script Concordance (SC) approach was used as an alternative test format to measure the presence of knowledge organization reflective in one's clinical reasoning skills (i.e., diagnostic, investigation and treatment knowledge).

Method: We investigated the reliability and validity of a 40-item paediatric version of the SC test with three groups representing 53 medical students (novices), 42 paediatric residents (intermediates) and 11 paediatricians (experts).

Results: A comparison between scoring techniques based on experts' ratings of the items showed internal reliability coefficients from 0.74 for the one-best answer up to 0.78 for alternative scoring techniques. An ANOVA showed an increase in test performance from medical students through to expert paediatricians ($F = 84.05$, $p < .001$), but did not differentiate between the postgraduate year 1 to 3 paediatric residents. A large effect size (Cohen's *d*) difference of 1.06 was found between medical students and residents total SC test scores.

Conclusions: These results support other findings indicating the SC test format can be used to differentiate between the clinical reasoning skills of novices, intermediates and experts in paediatrics. An alternative scoring method that includes one best answer and partial marks was also supported for grading SC test items.

Correspondence: Dr. Jean-Francois Lemay, Alberta Children's Hospital, 2888 Shaganappi Trail NW, Calgary, AB Canada, T3B 6A8. Phone: (403)955-2911; Fax: (403)955-7665; Email: Jean-Francois Lemay – jf.lemay@albertahealthservices.ca

Introduction

In recent years, emphasis has been placed on using principles of cognitive psychology in the theoretical understanding and assessment of clinical expertise.[1-3] In particular, cognitive structures referred to as "illness scripts" have focused less on the organization of relevant clincial information about diseases, their consequences and the context under which they present and more on the in-depth knowledge or superior reasoning skills of pathophysiological states.[4] Based on the diagnosis of similar and previously encountered clinical presentations, research has shown that experienced practitioners possess more elaborate knowledge networks better suited to the daily tasks of clinical problem solving.[5-6] Using both hypothetico-deductive theory and the illness script concept of knowledge organization, Charlin et al.[7] introduced the Script Concordance (SC) approach to assess the skills that produce clinical competency.

The evaluation of clincial competency is only partly based on the person's ability to apply well-recognized solutions to well-defined clinical problems. In practice, clinical reasoning is more complex in that appropriate clinical decisions must be made based on patients' presentations that are often poorly defined. From the perspective of cognitive problem-solving, experienced physicians are able to narrow their diagnostic hypotheses and develop a more accurate clincial plan of action than novices. Although dependent on the mastery of the content associated with a particular domain, clinical experience supports the development of mental models, prototypes, or schemes that faciliate the diagnostic process.[2-3] The assumption is that better clinical reasoning skills are a result of the elaborate, semantic networks made between the clinical presentations and diagnostic categories.[8-9]

The SC test was designed to measure the presence of links between examinees' knowledge and the organization of this knowledge as it is adapted for efficiency in the clinical tasks associated with the most appropriate diagnostic, investigative strategies and treatment options available.[10-11] From the perspective of decision making in medicine, the clinician's post-test probability of correctly identifying a disease is a function of the previous knowledge of the disease's prevalence and the strength of the evidence provided in the clinical presentation.[12] Recognising the cognitive errors associated with a heuristic approach and biases in the estimation of probabilities, the basis of the clinical reasoning process is that experts are consistently more accurate in their abilities to diagnosis, investigate and manage clinical presentations.[13]

In most research concerning written assessment of competence, experienced clinicians score little better and sometimes even worse than end-of-training residents, even though one expects that greater experience should be reflected in higher scores.[14] In contrast, the objective of the SC test format is to be a more sensitive measure reflecting clinical experience. In SC tests, examinees must solve relatively ill-defined problems representative of the real world clinical profession and also must answer questions that experts consider of crucial importance in the process of solving that problem.[15] In particular, the probability of a diagnosis or usefulness of an investigation or treatment makes the one-best answer somewhat ambiguous in many clinical presentations. Therefore, rewarding partial marks to examinees in the SC approach reflects the development of the expert clinicians' scoring key – indicative of the variability expected by practicing clinicians' appraisal of the diagnosis, investigation or knowledge about treatment. The SC test's format has been found to be well accepted, relevant, and interesting to complete by students, residents, and physicians.[16]

Ideally, one would want to create an examination that reflects all the competencies necessary for clinical expertise including clinical reasoning. The SC test attempts to measure the links between examinees' ability to adapt to the introduction of new diagnosis, investigation and treatment information that imitate problem solving effectiveness found in real clinical settings.

The goal of this present study was to investigate the reliability and validity of the SC test in the assessing of differences in clinical knowledge between novices (medical students), intermediates (paediatric residents) and experts (experienced paediatricians). In particular, the two main objectives identified were: 1) to measure the internal reliability and effect size differences of examinees on a paediatric version of the SC test, 2) to test a standard (one best answer) scoring method for items with two other alternative scoring techniques, and 3) to determine the validity of the SC test and scoring

method for assessing clinical reasoning skills (i.e., diagnostic, investigation and treatment knowledge) of a sample of medical students, paediatric residents (postgradute years 1 to 3), and paediatricians.

Method

A paediatric version of the SC test was developed based on the core competencies criteria outlined for the paediatric specialization by the Royal College of Physicians and Surgeons of Canada.[17] A table of specifications based on these paediatric objectives ensured that relevant content was used in the identification of each test item. The relevance, appropriateness, and quality of the items developed for the diagnostic, investigative and treatment knowledge sections of the SC test was completed through written and verbal consultations with both paediatric residents and paediatricians. The resulting 40 item SC test was constructed to measure residents' abilities to diagnose common paediatric presentations, to identify appropriate laboratory tests for confirmation of diagnoses, and to choose appropriate treatment or management options. This study received approval from the Conjoint Health Research Ethics Board of the University of Calgary.

The SC test provides examinees with written clinical presentations that require them to interpret relevant data in the process of making appropriate diagnostic, investigative or treatment decisions. A panel of experienced paediatric physicians, chosen among a list of certified general and specialist paediatricians with an expressed interest in medical education reviewed the SC test for both face and content validity. Each test item of the paediatric test was designed to have four components: 1) the patient's presentation is written in the form of a clinical vignette, 2) a diagnostic hypothesis, an investigation action, or a treatment option is provided that is relevant to the particular situation, 3) new information is introduced in the form of a condition that might have an effect on the diagnostic hypothesis, investigative action, or treatment option, and 4) a 5-point Likert-type scale is used to record the examinees' response (Figure 1).

To compare a range of potential responses to the paediatric version of the SC test, we identified three functionally different groups of participants with a range of clinical experience in paediatrics (i.e., medical students with no or limited clincial experience, residents in their first three years of a paediatric residiency program, and the panel of experienced paediatricians). One of the main protocols of the SC test is to use the responses obtained by a group of experienced clinicians as the standard for the scoring key and, hence, from which other examinees' (i.e., medical students and residents) clinical knowledge is assessed. The variability of the expert responses reflect the complexity associated with making a clear-cut distinction between clinical diagnoses.[18] In the one best answer and partial marks alternative method for scoring our paediatric version of the SC test, for example, if on an item six out of 11 panel experts have selected the -1 response option this choice receives 1 point (see Figure 2). If three of the expert physicians chose the -2 response, the examinee would receive 0.33 (i.e., 3/11), and if two experts chose the 0 response, this option would receive 0.18 (2/11). This scoring technique combines the standard one best answer method (allowing a maximum of 1 mark per question) while rewarding examinees with partial marks for options that may be more or less probable based on the ambiguity of the diagnosis, investigation and treatment information provided in the item. As such, the SC test was completed initially by 11 experienced paediatricians (M = 8.5 years as a practicing physician, SD = 4.6). Once the scoring key was derived for each of the 40 items, a cohort of 53 medical students doing a second year paediatric course at the University of Calgary and 42 paediatric residents from the University of Oregon (Portland, Oregon, USA) and the University of Washington (Seattle, Washington, USA) completed the paediatric version of the SC test. The number of residents that participated in the study were distributed evenly across the first three years (n = 14, for each year) of the postgraduate paediatric residency program.

Results

The internal reliability analysis for the paediatric version of the 40-item SC test derived an Cronbach's alpha value of 0.74 for the one-best answer, 0.77 for the one best answer and partial marks, and 0.78 for the partial marks only scoring methods (see Table 1). Although the reliability between the two alternative scoring methods was shown to be equivalent to the standard one best answer, the one best answer and partical marks scoring technique allows for the total test to be graded on a

Figure 1. Examples Of paediatric diagnostic, investigation and treatment knowledge test items based on the Script Concordance (SC) test method.

Diagnostic Knowledge: A 7 year-old boy presents with cough and clear-yellowish nasal discharge that has lasted for 2 weeks without any improvement. Parents report that the child has had a low-grade fever and frontal headaches for the last 5 days. Physical examination is unremarkable.

If you were thinking of…	And then you find…	This hypothesis becomes…				
Sinusitis	Sinus X-ray: Normal	-2	-1	0	1	2

Investigation Knowledge: A 13 year-old female adolescent presents to the Emergency Department with an episode of loss of consciousness during a basketball practice. Physical examination is unremarkable including the measurements of her vital signs.

If you were considering ordering…	And then you find…	This investigation becomes…				
Electrocardiogram	Maternal aunt died of sudden cardiac arrest at the age of 21	-2	-1	0	1	2

Treatment Knowledge: An 8 year-old boy returns from a Scouts' camping trip with fever, sore throat, and a headache. The boy does not appear toxic in your walk-in clinic. The rapid strep test (throat swab) is positive.

If you were thinking of prescribing…	And then you find…	That prescription becomes…				
Penicillin (PO)	Medical file shows the child had a few red plaques with Cefaclor (cephalosporins 2nd generation PO)	-2	-1	0	1	2

Scoring Key			
Diagnostic Knowledge		**Investigation and Treatment Knowledge**	
- 2 =	Ruled out or almost ruled out	- 2 =	Contra-indicated totally or almost totally
- 1 =	Less probable	- 1 =	Not useful or even detrimental
0 =	Neither less or more probable	0 =	Nor less nor more useful
+1 =	More probable	+1 =	Useful
+2 =	Certain or almost certain	+2 =	Necessary or absolutely necessary

total score of 40 marks. Using the partial marks only scoring technique resulted in a total maximum score of 29.3 which was found to be more difficult to score and less inituitive for the examinees or examiners. The internal reliability of the diagnostic, investigation and treatment knowledge subtests using the one best answer and partial marks scoring method were $\alpha = 0.51$, $\alpha = 0.55$, and $\alpha = 0.50$ respectively.

The descriptive statistics and an ANOVA on the paediatric version of the SC test and its' three subtests (i.e., the diagnostic, investigation and treatment knowledge sections) are shown in Table 2 for the medical students, paediatric residents and paediatricians. The medical students had the lowest mean SC test results out of a total possible score of 40 (M = 18.3; SD = 2.8), followed by the residents (M = 21.6; SD = 3.4) and then the expert clinicians (M = 32.0; SD = 4.3). An ANOVA on the total test scores showed a significant difference between the three groups and on each of the diagnostic, investigation, and treatment knowledge subtests ($p < .001$). A post hoc analysis using Tukey's HSD test, however, showed that for the investigation knowledge subtest there is no statistically significant difference between the medical students and paediatric residents. Large effect size differences were found between medical students' and residents' scores on the total SC test (d = 1.06) and diagnostic (d = 0.79),

Table 1. Example of scoring on one item and the internal reliability of one best answer and two alternative scoring techniques on the SC test.

Response Scale for SC Test Item	-2	-1	0	+1	+2	Max Total Score	Cronbach's alpha
Scored by Expert Paediatricians (N=11)	3/11	6/11	2/11	0/11	0/11		
One Best Answer	0	1	0	0	0	40	0.74
One Best Answer plus Partial Marks	0.27	1.0	0.18	0.0	0.0	40.0	0.77
Partial Marks Only	0.27	0.55	0.18	0.0	0.0	29.3	0.78

Table 2. ANOVA results for diagnostic, investigation, treatment and total SC test scores by medical students, paediatric residents and practicing paediatricians.

							*Subset for alpha= .05	
		n	***M***	***SD***	***F***	***p***	**1**	**2**	**3**
Diagnostic	Medical Students	53	7.7	1.72	41.39	.001	7.7		
Knowledge	Paediatric Resident	42	9.0	1.57				9.0	
	Paediatricians	11	12.7	1.90					12.7
Investigation	Medical Students	53	6.2	1.37	52.76	.001	6.2		
Knowledge	Paediatric Resident	42	7.1	1.70			7.1		
	Paediatricians	11	11.4	1.93				11.4	
Treatment	Medical Students	53	4.4	1.26	27.27	.001	4.4		
Knowledge	Paediatric Resident	42	5.5	1.66				5.5	
	Paediatricians	11	7.8	1.31					7.8
SC Test	Medical Students	53	18.3	2.78	84.05	.000	18.3		
Total Score	Paediatric Resident	42	21.6	3.41				21.6	
	Paediatricians	11	32.0	4.29					32.0

*$p < .05$; Tukey HSD uses harmonic mean sample size = 22.5 as the group sizes are unequal. Note: Maximum possible SC test score = 40)

investigation (d = 0.58), and treatment (d = 0.75) knowledge subtests. The interpretation of the magnitude of the effect size is based on Cohen's[19] suggestions of a d of 0.30 as "small", d of 0.50 as "medium", and d of 0.70 as "large".

In a separate subgroup analysis of the paediatric residents, there was no significant differences found between postgraduate year 1 to 3 residents' total test score performance (F = 0.05, p = .95).

Discussion

In the present study, there were three main findings. First, the paediatric version of the SC test shows relatively high internal reliability and validity as a testing format to differentiate the clinical knowledge assessment skills between medical students, paediatric residents and experienced paediatricians. Second, the internal reliability coefficients of the three scoring methods tested were equivalent, but the one best answer and partial marks scoring techniques have better reported face validity. Third, a comparison of the postgraduate year 1 to 3 residents' performance on the paediatric version of the SC test did not show significant differences for the total SC test and knowledge, investigation, and treatment subtest scores. The strength of the SC test format was supported by results that indicate significant differences in knowledge organization between novice medical students, intermediate paediatric residents and expert paediatricians. They reflect an increasing level of clinical experience and expertise in the organization of clinical knowledge specific to paediatric cases by the respective groups. Although we were able to show high internal

reliability with the paediatric version of the SC test, the potential to differentiation between the diagnostic, investigation and treatment knowledge skills using the SC test format was inconclusive with the residents only. In support of the combined benefits of using an alternative scoring technique (to allow for the partial marks in making a more or less probable diagnosis) and a standard scoring method (i.e., one-best answer = 1 mark) that allows for a maximum possible score equvialent to the number of items on the test, the alternative one best answer and partial marks scoring technique was found to be the more accepted method of presenting the results to the examinees. Medical diagnosis is a categorization task that allows physicians to make predictions about features of clinical situations and to determine appropriate courses of action. The script concept provides a theoretical framework to explain how medical diagnostic knowledge can be structured for diagnostic problem solving. In medical education assessment, there are not many testing methods available that allow comparison of clinical knowledge among subjects as different as medical students, residents and faculty members. Bordage[5] has shown that some clinicians organize their knowledge well and that others do not. Therefore, we believe that the use of the SC approach as an assessment method or potential approach to teaching medical students and residents to organize their knowledge for more efficient use in their clinical work warrants further investigation and research.

Acknowledgement

The authors gratefully acknowledge the financial support of the Royal College of Physicians and Surgeons of Canada.

References

1. Elstein AS, Shulman LS, Sprafka SA. Medical problem solving: an analysis of clinical reasoning. Cambridge, MA: Harvard University Press; 1978.
2. Bordage G, Zacks R. The structure of medical knowledge in the memories of medical students and general practitioners: categories and prototypes. Med Educ. 1984;18:406-11.
3. Patel VL, Groen GJ. Knowledge based solution strategies in medical reasoning. Cogn Sc. 1986;10:91-116.
4. Schmidt HG, Norman GR, Boshuizen HPA. A cognitive perspective on medical expertise: theory and implications. Acad Med. 1990;65 (10):611-21.
5. Bordage G. Elaborated knowledge: a key to successful diagnostic thinking. Acad Med. 1994;69 (11):883-5.
6. Norman GR, Brooks LR. The non-analytical basis of clinical reasoning. Adv Health Sc Educ. 1997;2:173-84.
7. Charlin B, Brailovsky C, Leduc C, Blouin D. The diagnosis script questionnaire: a new tool to assess a specific dimension of clinical competence. Adv Health Sc Educ. 1998;3:51-8.
8. Bordage G, Lemieux M. Semantic structures and diagnostic thinking of experts and novices. Acad Med. 1991;66 (9):S70-2.
9. Schmidt HG, Boshuizen HPA. On acquiring expertise in medicine. Educ Psych Rev. 1993;5(3):205-21.
10. Charlin B, Roy L, Brailovsky C, Goulet F, van der Vleuten C. The script concordance test: a tool to assess the reflective clinician. Teach Learn Med. 2000;12 (4),189-95.
11. Charlin B, van der Vleuten C. Standardized assessment of reasoning in contexts of uncertainty: the script concordance approach. Eval & Health Prof. 2004:27(3):304-19.
12. Sackett DL, Richardson WS, Rosenberg W, Haynes RB. Evidence-based medicine: how to practice and teach EBM. New York: Churchill Livingstone, 1997.
13. Elstein AS. Heuristics and biases: selected errors in clinical reasoning. Acad Med. 1999;74 (7):791-4.
14. Van der Vleuten CPA. The assessment of professional competence: developments, research and practical applications. Adv Health Sc Educ. 1996;1:41-67.
15. 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.
16. Charlin B, Tardif J, Boshuizen HPA. Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. Acad Med. 2000;75 (2):182-90.
17. The Royal College of Physicians and Surgeons of Canada. Skills for the new millennium: report of the societal needs working group. CanMEDS 2000 Project, 1996.
18. Charlin B, Gagnon R, Pelletier J, Coletti M, Abi-Rizk G, Nasr C, Sauve E, Van Der Vleuten C. Assessment of clinical reasoning in the context of uncertainty: the effect of

variability with the reference panel. Med Educ. 2006;40:848-54.

19. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Erlbaum, 1988.