Surgical Innovation
Volume 13 Number 2
June 2006 115-119

10.1177/1553350606291042
http://sri.sagepub.com
hosted at
http://online.sagepub.com

A Novel Method of Assessing Clinical Reasoning in Surgical Residents

Sarkis H. Meterissian, MD, MSc, FRCS, FACS

At present, surgical educators can readily assess knowledge with multiple-choice examinations, and inanimate models can be used to assess technical skills. Clinical judgment and reasoning are indispensable skills used by expert surgeons to solve ill-defined problems encountered in the emergency department, clinic, and operating room. The Script Concordance Test, a new tool of clinical reasoning assessment, can test the elaborated networks of knowledge that experienced surgeons develop over the years. It allows for multiple different approaches to the same problem and could be developed as both a formative and summative assessment tool in general surgery residency programs. This article explores the theoretical and practical aspects of the Script Concordance Test.

Keywords: Clinical reasoning; script-concordance; surgical residency

Theories of Clinical Reasoning

Clinical reasoning is a crucial component of physician competence.[1] Three models have been proposed over the years to understand clinical problem solving. The first originated from studies at McMaster University and the University of Michigan and is known as the "hypothetico-deductive" method. It states that physicians, within minutes of a clinical encounter, generate several diagnostic hypotheses and then gather data to rule in or out each hypothesis on their list. Studies showed that this method was too general, unrelated to expertise, and in most cases, diagnostic accuracy was more related to expert knowledge than a particular method.[2,3] In the 1980s, memory recall was thought to form the basis of clinical reasoning. Findings in other fields such as chess and classical music could not, however, be replicated in clinical medicine.[4-6] More recently, the concept of mental representations has been advanced to explain clinical reasoning.

Schmidt et al[7] proposed that with increasing experience, clinicians move through 3 kinds of mental representations: basic mechanisms of disease, to illness scripts, to a bank of cases derived from experience. Clinical medicine is fraught with ill-defined problems that clinicians solve in a variety of ways, depending on their experience.[8] Barrows and Feltovich[9] have proposed that expert clinicians have a bank of "illness scripts" consisting of a story-like narration of a typical case of the condition that can be used when confronted with a new case. In addition, it has been hypothesized that expert clinicians have mental probability matrices that they use in conjunction with illness scripts to arrive at the right diagnosis.[10]

Decision making is a process that includes problem recognition and values analysis (choosing from a set of alternatives depending on their probability).[11] Experienced practitioners possess elaborated networks of knowledge fitted to the regular tasks they perform, known as "scripts."[12] The script-concordance approach, which is based on cognitive psychology script theory,[12] may provide a way to build a theory-based tool to assess decision-making skills such as those in the intraoperative setting.

The Script Concordance Test (SCT) is a new tool of clinical reasoning assessment developed by Charlin et al[13] based on these theoretical and empirical findings about clinical reasoning. Unlike multiple-choice questions, the SCT assesses the reasoning skills of examinees and how they actively process information to confirm or eliminate hypotheses with a series of qualitative judgments. These judgements are used daily by experienced practitioners and are refined with increasing experience. The SCT should therefore not

From the Division of General Surgery and Center for Medical Education, McGill University, Montreal, Quebec, Canada.

Address correspondence to: Sarkis H. Meterissian, MD, 687 Pine Ave West S10.22, Montreal, Quebec H3A 1A1, Canada; e-mail: sarkis.meterissian@muhc.mcgill.ca.

only measure skills at solving ill-defined problems but should also reflect refinement with increasing residency level.

The script-concordance approach has 3 principles[14]: (1) examinees are faced with a challenging authentic clinical situation in which several options are relevant, (2) responses follow a Likert scale that reflects script clinical reasoning theory, and (3) scoring is based on the aggregate scoring method to take into account the variability of the clinical reasoning process among experts. This article reviews the principles of constructing and scoring an SCT and summarizes studies that have used the SCT to assess clinical reasoning.

Previous Studies Using the Script-Concordance Approach

Studies to date have demonstrated both construct and predictive validity for the SCT. Early studies showed higher mean scores achieved by physicians compared with residents and students, supporting the construct validity of the SCT.[13] The predictive validity of the SCT was determined in a family medicine residency, where a study was done to determine whether scores obtained on the SCT taken at the end of clerkship predicted performances on short answer management problems and simulated office orals that test clinical reasoning at the end of residency.[15] In a group of 24 students, the SCT was highly correlated with the short answer management problems test and simulated office orals but not with the objective structured clinical examination, which assesses hands-on skills rather than clinical reasoning.

Stability of the SCT across 2 different linguistic and learning environments has also been assessed.[16] The SCT was administered to urology residents from a French and a Canadian university. Two groups of urologists were used as the expert panel in the aggregate scoring method. This study also showed construct validity, with scores increasing with clinical experience in urology (Table 1). What was even more interesting was that the candidates obtained higher scores when correction was done using the answer key provided by the experts from the same country. These results give further credence to the validity of the aggregate scoring method.

The versatility of the SCT has been demonstrated in radiology, where it reliably measured skills that are traditionally difficult to measure, including perception and interpretation skills in film reading.[17] The SCT is thus a novel approach for the assessment of clinical reasoning and, as such, may be useful to assess surgical residents' clinical decision-making skills.

Table 1. Comparison of Mean Scores by Groups According to the Composition of Reference Panels

Composition of the Reference Panel	Residents[a] (n = 25)	Students[a] (n = 23)	P[b]
French urologists (n = 10)	53.16 ± 4.68	46.67 ± 5.60	.0001
Canadian urologists	55.15 ± 4.21	48.33 ± 5.64	.0001

a. Values are mean ± SD.
b. $P < .05$ was considered significant (univariate analysis of variance).

How to Construct a Scrips-Concordance Test

Number of Items

As determined from the work of Charlin et al, 50 to 60 items are needed in an SCT to achieve a reliability coefficient (Cronbach α) of 0.80. Exams should have about 100 questions, because a number will usually be eliminated using an analysis of item–total item correlation to select the best items. The format of a typical item is shown in Figure 1. Each item of the test must be built so that reflection in action is necessary to answer it. When preparing the clinical vignette, an attempt should be made to keep the clinical scenario authentic, but each item should require reasoning skills and some experience.

Scoring Process

According to Charlin, 10 to15 experts are required to develop the scoring system. The experts can either be subspecialists or board-certified general surgeons without subspecialization. The experts are required to complete the test individually. Scores for each item are computed from the frequencies given to each point of the Likert-type scale (aggregate scoring method),[18,19] as shown in Figure 2. Credits for each answer are transformed proportionally to get a maximum score of 1 for modal experts' choice(s) on each item; other experts' choices are given partial credit. Answers not chosen by experts receive 0. For example, if on an item, 8 of 15 experts have chosen response +1, this choice receives 1 point (8/8). If 6 experts have chosen response +2, this choice receives 0.75 points (6/8). If 1 expert has chosen response 0 this choice receives 0.125 points (1/8). Because none of the experts chose responses −1 or −2, these choices receive 0 points. The total score for the test is the sum of credits on all items. This scoring system takes into account the

A 28 year-old pregnant woman presents to the ER with severe biliary colic. This is her third attack and an ultrasound reveals gallstones. In your management:

If you were planning…	and found…	the planned management is:				
An open cholecystectomy	That the patient was in the first trimester	-2	-1	0	+1	+2
A laparoscopic cholecystectomy	That the patient is in the second trimester and has gestational diabetes	-2	-1	0	+1	+2

-2	Absolutely contraindicated
-1	Should be reconsidered
0	Not affected
+1	Supported
+2	Absolutely correct

Figure 1. An example of a Script-Concordance Test question.

range of potential scores and the expected distribution should be broad (if only 1 answer is chosen by all the experts, the SCT question becomes a multiple-choice question) but clustered around a mean (too broad a distribution invalidates the question).

Applications in Surgery

Most current methods of clinical competence assessment use either performance-based methods[20] (eg, objective structured clinical examinations) or tests assessing the "technical rationality" part of clinical reasoning (eg, multiple-choice questions). These assessment tools fail to capture the uncertainty of clinical problems such as those encountered in the operating room, emergency department, or clinic. Problem-solving in the operating room for example requires a mixture of knowledge and experience. Problems are encountered which can force the surgeon to deviate from his or her preoperative surgical plan and such decisions under pressure can significantly affect the patient's outcome.

Board examinations for certifying general surgeons presently incorporate a multiple-choice question test and an oral examination. The oral examination's purpose is to determine how the candidate can "think on his or her feet" and solve problems in a short period of time. Oral exams can assess professional knowledge[21] needed to solve ill-defined problems, but they have limitations such as difficulty of standardization, reliability of scoring, or ease of administration to large groups of examinees. At a time when cognitive psychology has become the major conceptual framework in medical education,[22] it is necessary to add to these methods a way to assess clinical reasoning which measures its process rather than just its outcome.

The SCT could be used as a formative assessment tool and for summative evaluation. If a standard examination could be developed and validated in a number of North American centers, then standard scores by resident postgraduate year could be identified. Program directors could use these scores to identify residents having trouble with clinical

	-2	-1	0	+1	+2
Number of experts choosing answer	0	0	5	4	1
Score	0	0	5/10	4/10	1/10
Transformed score	0	0	5/5	4/5	1/5
Credit per item	0	0	1	0.8	0.2

Figure 2. The scoring grid.

decision making (>2 standard deviations below the mean) and provide remediation.

The exam could also be used in the same way as the annual in-training exam administered to surgical residents in North American training programs, in that residents would have to pass it in order to proceed to the next year. The in-training exam measures knowledge whereas the SCT would determine whether the resident has reached the same level of clinical reasoning as his/her peers.

Conclusion

At the present time, tools have been developed to assess surgical skills and knowledge but there is as yet no accurate, reliable, and valid way of measuring how our residents think. The SCT is a novel way of assessing clinical reasoning and should be developed and validated in North American surgical training programs as a formative and summative evaluation instrument.

References

1. Norman G. Research in clinical reasoning: past history and current trends. *Med Educ*. 2005;39:418-427.
2. Elstein AS, Shulman LS, Sprofka SA. Medical problem-solving: an analysis of clinical reasoning. Cambridge, MA: Harvard University Press, 1978.
3. Neufeld VR, Norman GR, Feightner JW, Barrows HS. Clinical problem-solving by medical students: a longitudinal and cross-sectional analysis. *Med Educ*. 1981;5: 315-322.
4. Muzzin LJ, Norman GR, Freightner JW, Tugwell P, Guyatt G. Expertise in recall of clinical protocols in two specialty areas. Proceedings of the 22nd conference on Research in Medical Education, Washington, DC. *Proc Annu Conf Res Med Educ*. 1983;122-126.
5. Patel VL, Groen GJ. Knowledge based solution strategies in medical reasoning *Cogn Sci*. 1986;10:91-115.
6. Boshuizen HPA, Schmidt HG. Biomedical knowledge and clinical expertise. *Cogn Sci*. 1992;16:153-184.
7. Schmidt HG, Norman GR, Boshuizen HPA. A cognitive perspective on medical expertise: theory implications. *Acad Med*. 1990;65:611-621.
8. Grant J, Marsden P. Primary knowledge, medical education and consultant expertise. *Med Educ*. 1988;22:173-179.
9. Barrows HS, Feltovich PJ. The clinical reasoning process. *Med Educ*. 1987;21:86-91.
10. Papa FJ, Shores JH, Meyer S. Effects of pattern matching, pattern discrimination and experience in the development of diagnostic expertise. *Acad Med*. 1990;5: 521-2.
11. Mullen JD, Roth BM. *Decision Making: Its Logic and Practice*. Savage, MD: Rowman & Littlefield, 1991.
12. Charlin B, Tardif J, Boshuizen HPA. Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. *Acad Med*. 75:182-190, 2000.
13. Charlin B, Roy L, Brailovsky CA, Goulet F, van der Vleuten C. The Script Concordance test: a tool to assess the reflective clinician. *Teach Learn Med*. 2000;2:189-195.
14. Charlin B, van der Vleuten CPM. Standardized assessment of reasoning in contexts of uncertainty: the script-concordance approach. *Eval Health Prof*. 2004;27: 304-319.
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: 430-436.
16. Sibert L, Charlin B, Corcos J, Gagnon R, Grise P, van der Vleuten C. Stability of clinical reasoning assessment results with the script concordance test across two different linguistic, cultural and learning environments. *Med Teach*. 24:537-542, 2002.

17. Brazeau-Lamontagne L, Charlin B, Gagnon R, Samson L, van der Vleuten C. Measurement of perception interpretation skills during radiology training: utility of the script concordance approach. *Med Teach*. 2002;26:326-332.
18. Norman GR. Objective measurement of clinical performance. *Med Educ*. 1985;19:43-47.
19. Norcini JJ, Shea JA, Day SC. The use of the aggregate scoring method for a recertification examination. *Eval Health Prof*. 1990;13:241-251.
20. Dauphinee W.D. Assessing clinical performance: where do we stand and what might we expect? *JAMA*. 1995; 275:741-743.
21. Schon DA. *The Reflective Practitioner: How Professionals Think in Action*. New York, NY: Basic Books; 1983.
22. Irby D. Editorial. *Acad Med*. 1997;72:116.