A Novel Method of Assessing Clinical Reasoning in Surgical Residents

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

Clinical medicine is fraught with ill-defined problems that clinicians solve in a variety of ways, depending on their experience. Barrows and Feltovich 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.

Decision making is a process that includes problem recognition and values analysis (choosing from a set of alternatives depending on their probability). Experienced practitioners possess elaborated networks of knowledge fitted to the regular tasks they perform, known as “scripts.” Schmidt et al. 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. Barrows and Feltovich 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.

Decision making is a process that includes problem recognition and values analysis (choosing from a set of alternatives depending on their probability). Experienced practitioners possess elaborated networks of knowledge fitted to the regular tasks they perform, known as “scripts.” The script-concordance approach, which is based on cognitive psychology script theory, 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. 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...
only measure skills at solving ill-defined problems but should also reflect refinement with increasing residency level.

The script-concordance approach has 3 principles:\(^1^4\): (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.\(^1^3\) 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.\(^1^5\) 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.\(^1^6\) 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.\(^1^7\) 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.

**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 \(\alpha\)) 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 to 15 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),\(^1^8,1^9\) 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

---

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

<table>
<thead>
<tr>
<th>Composition of the Reference Panel</th>
<th>Residents(^a) ((n = 25))</th>
<th>Students(^a) ((n = 23))</th>
<th>(P^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>French urologists ((n = 10))</td>
<td>53.16 ± 4.68</td>
<td>46.67 ± 5.60</td>
<td>.0001</td>
</tr>
<tr>
<td>Canadian urologists</td>
<td>55.15 ± 4.21</td>
<td>48.33 ± 5.64</td>
<td>.0001</td>
</tr>
</tbody>
</table>

\(a\). Values are mean ± SD.

\(b\). \(P < .05\) was considered significant (univariate analysis of variance).
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:

<table>
<thead>
<tr>
<th>If you were planning...</th>
<th>and found...</th>
<th>the planned management is:</th>
</tr>
</thead>
<tbody>
<tr>
<td>An open cholecystectomy</td>
<td>That the patient was in the first trimester</td>
<td>-2 -1 0 +1 +2</td>
</tr>
<tr>
<td>A laparoscopic cholecystectomy</td>
<td>That the patient is in the second trimester and has gestational diabetes</td>
<td>-2 -1 0 +1 +2</td>
</tr>
</tbody>
</table>

-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 (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 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, 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
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**


