European Urology

European Urology 82 (2002) 1–7

Assessment of Clinical Reasoning Competence in Urology with the Script Concordance Test: an Exploratory Study Across Two Sites from Different Countries

Louis Sibert[a,*], Bernard Charlin[b], Jacques Corcos[c], Robert Gagnon[b], Joel Lechevallier[d], Philippe Grise[a]

[a]*Department of Urology, Rouen University Hospital-Charles-Nicolle, Pavillon Derocque 1, rue de Germont, 76031 Rouen Cedex, France*
[b]*Unit of Research and Development in Medical Education, Faculty of Medicine, University of Montreal, Montreal, Canada*
[c]*Department of Urology, Sir Mortimer B. Davis-Jewish General Hospital, McGill University, Montreal, Canada*
[d]*Department of Medical Education, Faculty of Medicine, Rouen, France*

Accepted 3 December 2001

Abstract

Objectives: The script concordance (SC) test is designed to measure the organisation of knowledge that allows interpretation of data in clinical reasoning. This study explores the use of this new written examination tool in urology.

Materials and Methods: An 80 items SC test was administered to participants from a French and a Canadian university. Three levels of experience were tested: urologists ($n = 22$), residents in urology ($n = 25$) and students ($n = 23$). Scores between groups were compared by analysis of variance. Reliability analysis was studied with Cronbach alpha coefficient.

Results: Mean global scores were 51.45 ± 5.29 for students, 58.19 ± 3.81 for residents and 62.27 ± 5.46 for urologists. The difference between the three groups was statistically significant ($P < 0.00001$). Interaction between levels of expertise and sites was apparently not significant ($P = 0.326$). Cronbach alpha was 0.79 for the test.

Conclusions: This study shows that the SC test is able to discriminate among participants according to their levels of clinical experience in urology. The results are similar in two different learning environments. The SC test appears as a simple and direct approach to testing organisation and use of knowledge in urology.

Keywords: Clinical reasoning; Knowledge structures; Assessment; Urology

1. Introduction

The primary goal of urology residency programmes is to produce competent physicians. Although a sound knowledge base and technical skills are vital for an urologist, clinical competence encompasses numerous other components, including interviewing and interpersonal skills, physical diagnosis skills, problem-solving abilities and clinical reasoning [1]. Most of the written methods for clinical reasoning assessment measure factual knowledge more than the organisation of knowledge that allows clinicians to recognise and handle situations effectively [2]. Other evaluation tools of clinical competence, such as oral examination, often lack either reliability or validity, even when they are structured by the use of rating forms.

It is well established that medical reasoning is a hypothetico-deductive process [3,4] characterised by early generation of hypotheses, oriented data collection and decision making judgement, using collected data to confirm or reject hypotheses. Several authors have hypothesised that, in clinical medicine, skilled and experienced practitioners differ from those less experienced and skilled because they possess elaborated networks of knowledge fitted to the tasks they regularly do [2,5–8]. These networks, named scripts [9–11], are

* Corresponding author. Tel. +33-2-32-88-82-05; Fax: +33-2-32-88-82-05.
E-mail address: louis.sibert@chu-rouen.fr (L. Sibert).

0302-2838/02/$ – see front matter © 2002 Published by Elsevier Science B.V.
PII: S0302-2838(02)00053-2

organised to fulfil goals within tasks concerning diagnosis, strategies of investigation or treatment options. They begin to appear when students are faced with their first clinical cases and are later developed and refined during their entire clinical career [10,11]. This theory considers that in diagnostic situations clinicians bring in their working memory knowledge related to each relevant hypotheses. Knowledge activated from memory is then used in a deductive process to actively seek information that will allow confirmation or rejection of entertained hypotheses [11].

From this cognitive theory of clinical expertise development, a new written assessment tool, the script concordance (SC) test was designed to measure the richness of these networks [12]. The test approach consists of presenting examinees with a series of patients' problems and then asking examinees to make diagnostic, investigative or therapeutic decisions when specific elements of information are provided. It places examinees in written but authentic clinical situations where they have to interpret data to make decisions. Earlier studies have shown that SC tests have good psychometric qualities with an increase in the mean scores of participants with different levels of clinical expertise [12–15].

In this context, this paper reports an exploratory study of the first utilisation of the SC test in the field of urology. The purpose of this study was to explore the possible use of the script concept to assess clinical competence in urology. In order to answer the research question, the SC test was administered to three groups of candidates with different levels of experience in urology (senior urologists, residents, students). To establish the value of the test across different learning environments, participants were recruited from two universities from two different countries.

2. Materials and methods

2.1. Construction of the SC test

Two faculty members were asked to describe clinical situations representative of urology practice and based on major educational objectives of urology training programmes. They were asked to specify for each situation (a) the relevant hypotheses, investigation strategies or treatment options; (b) the questions they ask, physical examinations they perform, and tests they request to solve the problem and (c) what clinical information, positive or negative, they would look for in these inquiries. Test items are built using the material obtained at this stage. Content of the SC test in urology is shown in Table 1.

The clinical situations were presented in short vignettes, each of them followed by a series of related test items. The item format differs with the objective of assessment (diagnosis, investigation or treatment). Each item consists of three parts. The first part includes a diagnostic hypothesis, an investigative action or a treatment option. The second presents new information (e.g. a clinical data, imaging study or laboratory test result) that might have an effect on the diagnostic hypothesis, investigative action or treatment option. The third part is a five-point Likert-type scale (see illustrations of the three formats in Table 2). Each item was built so that a reflection was necessary to answer it. It was also clearly specified in the instructions for participant that within the vignettes, each item is independent of the others. Hypotheses or options change for each question. Hence, a questionnaire of 85 items was then constituted (see example of items form the diagnostic section in Table 3). During their completion of the test, urologists were asked to identify the items they found confusing or not relevant. Five items were then discarded, 80 items were retained for the calculation of scores and statistical analysis.

2.2. Participants

The test was submitted to three groups of participants from the urology departments of two universities (McGill University, Canada; Rouen University Hospital, France): board certified urologists ($n = 22$, 12 from McGill University and 10 from Rouen University Hospital), residents ($n = 25$, 14 from McGill University and 11 from Rouen University Hospital) and medical students ($n = 23$, 8 from McGill University and 15 from Rouen University Hospital). All subjects were asked to participate volunteered. Criteria for inclusion were, for urologists, to have completed their entire

Table 1
Blueprint of the SC test in urology

Clinical problems	Context	Age	Sex	NB	Assessed components
Scrotum enlargement	C	42	M	5	DG
Scrotal trauma	E	27	M	5	TT
Renal colic	E	63	F	10	DG, I
Pelvis trauma	E	25	M	10	DG, TT
Infertility	C	31	M	5	DG
Urinary incontinence	C	57	F	5	DG
Urinary lithiasis	E	35	M	10	DG, TT
BPH, prostate cancer	C	58	M	10	I, TT
Erectile dysfunction	C	66	M	5	DG
Obstructive renal failure	E	73	F	5	TT
Urinary retention	E	71	M	5	TT
Kidney tumour	C	73	F	5	DG

NB: number of questions; C: consultation; E: emergency; M: male; F: female; DG: diagnostic; I: investigation; TT: treatment; BPH: benign prostatic hypertrophy.

Table 2
Illustration of questions and answering grids format

For diagnostic knowledge assessment

If you were thinking of	And then you find	This hypothesis becomes
(A diagnosis hypothesis)	*(A new clinical information, an imaging study or a laboratory test result)*	-2 -1 0 +1 +2

-2 the hypothesis is almost eliminated
-1 the hypothesis becomes less probable
0 the information has no effect on the hypothesis
+1 the hypothesis is becoming more probable
+2 it can only be this hypothesis

For investigation knowledge assessment

If you were considering to ask	And then you find	This investigation becomes
(A diagnostic test)	*(A new clinical information, an imaging study or a laboratory test result)*	-2 -1 0 +1 +2

-2 contra-indicated totally or almost totally
-1 not useful or even detrimental
0 nor less nor more useful
+1 useful
+2 absolutely necessary

For treatment knowledge assessment

If you were considering to prescribe	And then you find	The relevance of this treatment becomes
(A therapeutic option)	*(A new clinical information, an imaging study or a laboratory test result)*	-2 -1 0 +1 +2

-2 contra-indicated totally or almost totally
-1 not useful or even detrimental
0 nor less nor more useful
+1 useful
+2 necessary or absolutely necessary

The item format varies with the object of assessment (e.g. diagnostic, investigation, treatment).

training programme, for residents, to be included in the urology program at McGill or Rouen University Hospital, and for students to have had a rotation in a urology department in the last 6 months. All the residents had completed their first 2 years of core surgery and had begun their last 3 years of clinical urology training. All urologists were faculty members. In this exploratory research, the group of urologists was considered as the reference group (but not as a gold standard). The test was simultaneously administered in English and French. Translation was done by English and French professional translators. Great care was taken in the translation process in order to avoid word substitution, word omission, word addition or different meaning for a word in the other language. Each version of the test was then independently reviewed and cross-checked by experienced urologists in their own language. This translation process was previously reported and had proven effectiveness for minimising translation errors and assure equivalency of the English and French versions of clinical competence assessment tool [16].

2.3. Scoring process

The scoring process of the test is based on the principle that any experienced clinician answer is a reflection of expertise, and those answers for which there is no agreement among experts should not be discarded [13,14]. The answers of the reference group (urologists) were used to build the answer key. For each item, answers were assigned a weight corresponding to the proportion of the urologists who selected it. Credits for each answer were then transformed proportionally (division of all scores by the modal value) to obtain a maximum score of 1 for modal experts' choice(s) on each item, other experts' choices receiving a partial credit. Answers not chosen by any experts received zero. For example, if

Table 3
Example of items from the diagnostic section of the SC test in urology

If you were thinking of	And then you find	This hypothesis becomes				
Urethral rupture	Urethral bleeding	−2	−1	0	+1	+2
Retroperitoneal bladder rupture	Bladder distension	−2	−1	0	+1	+2
Urethral rupture	Upward and bulging prostatic apex at the digital rectal examination	−2	−1	0	+1	+2
Intra-peritoneal bladder rupture	Spontaneous micturition after the accident	−2	−1	0	+1	+2
Urethral rupture	Perineal haematoma	−2	−1	0	+1	+2

Note: −2: the hypothesis is almost eliminated; −1: the hypothesis becomes less probable; 0: the information has no effect on the hypothesis; +1: the hypothesis is becoming more probable; +2: it can only be this hypothesis. Clinical vignette: a 25-year-old male patient is admitted to the emergency room after a fall from a motorcycle with a direct impact to the pubis. Vital signs are normal. The X-ray reveals a fracture of the pelvis with a disjunction of the pubic symphysis.

on an item, 15 experts (out of 22) have chosen response +1, this choice received one-point (15/22 × 22/15), if seven experts have chosen response +2, this choice received 0.26 (7/22 × 22/15). The total score for the test is represented by the sum of the scores obtained at each item.

2.4. Statistical analysis

Descriptive statistics of the participants' scores on the SC test were performed, followed by a one-way analysis of variance to test the differences between groups' means. The homogeneity of group variances was estimated with the Levene's test in order to be able to interpret the results of the analysis of variance. The post-hoc Scheffe analysis was then used to determine precisely which differences of scores between groups of participants were significant. A two-factor analysis of variance was used to assess the interaction between the level of expertise and site with different learning environment. To evaluate the presence of a significant statistical difference, a $P < 0.05$ value was considered as significant. Reliability of the examination was assessed through the Cronbach alpha internal consistency coefficient. Medical education literature suggests that a Cronbach alpha coefficient of 0.80 or more is necessary for high-stake tests [17].

3. Results

Mean scores of the groups of participants according to their learning environment are illustrated in Fig. 1.

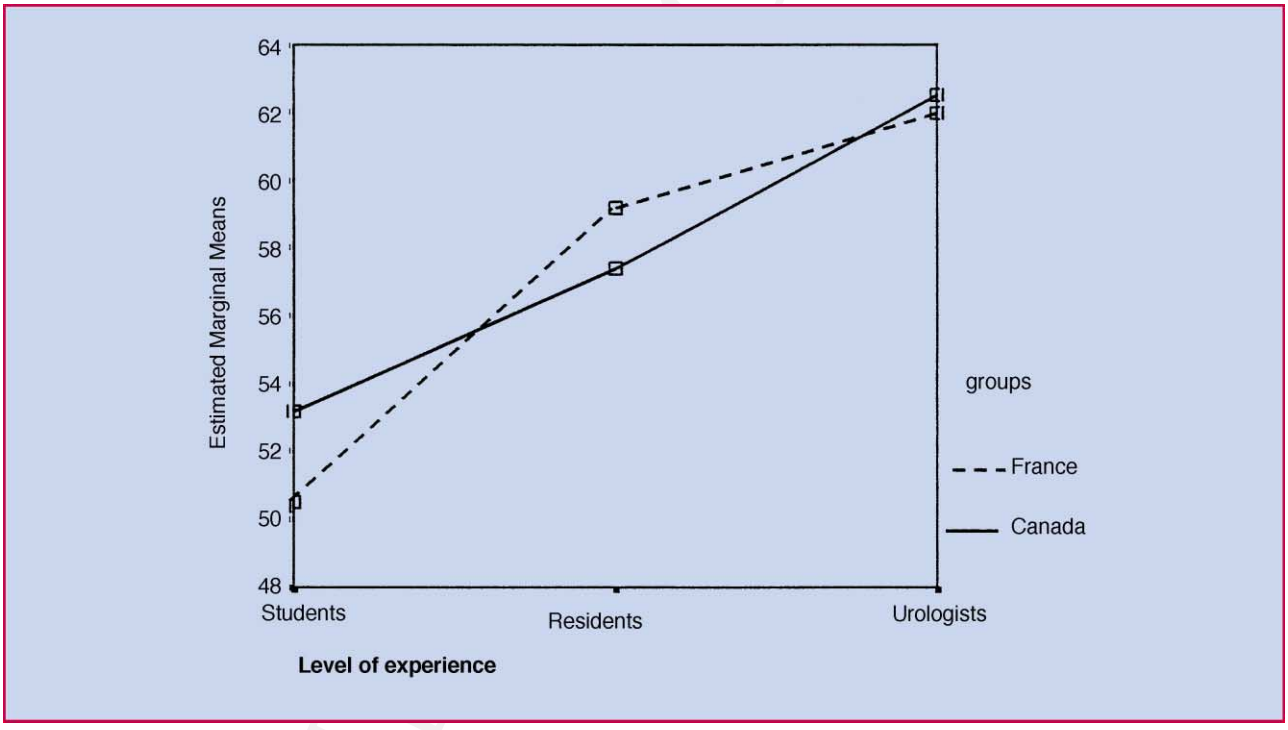

Fig. 1. Comparison of mean score progression with the level of expertise in the two learning environments. Values are means ± S.D. The group of urologists (both Canadian and French) was considered as the reference group to build the answer keys.

Table 4
Comparison of mean scores by groups according to level of expertise

Group	Size	Mean score	S.D.	Minimum	Maximum
Students	23	51.45	5.46	41.43	61.39
Residents	25	58.19	3.81	49.15	65.72
Urologists	22	62.27	5.29	42.21	67.08

Students vs. residents, students vs. urologists, residents vs. urologists: $P < 0.00001$ (analysis of variance and Scheffe post-hoc test); the group of urologists (both Canadian and French) was considered as the reference group to build the answer keys.

As concerns the Canadian participants, mean scores were 53.17 ± 4.18 for the students, 57.42 ± 4.78 for the residents and 62.52 ± 6.68 for the urologists. As regards the French participants, mean scores were 50.53 ± 5.96 for the students, 59.18 ± 1.78 for the residents and 61.98 ± 3.27 for the urologists. Global mean scores for the different groups of participants according their level of experience in urology are summarised in Table 4, the means scores were 51.45 ± 5.46 for the students, 58.19 ± 3.81 for the residents and 62.27 ± 5.29 for the urologists. One urologist had a relatively low score and one score achieved by a student was higher than the mean score of the residents. Therefore, the Levene's test of homogeneity of variance was used to verify whether the three groups' variances were equal. The results were not significant ($P = 0.203$), thus indicating that the three variances can be considered equal.

Differences between the mean scores for the three groups were statistically highly significant ($P < 0.00001$). The post-hoc Scheffe analysis indicated that significant differences were present between students and residents, between students and urologists, and between residents and urologists ($P < 0.05$). The interaction between mean scores and site was not significant ($P = 0.326$), indicating that there were no differences on groups' mean scores according the difference of sites. The Cronbach alpha reliability coefficient for the entire examination was 0.79.

4. Discussion

Contemporary assessment tools of clinical competence have repeatedly shown the puzzling fact that experienced clinicians score hardly better and sometimes worse than less experienced clinicians or students [17]. This counterintuitive finding indicates that most of these methods, especially written examinations, i.e. multiple-choice tests, measure clinical factual knowledge rather than clinical reasoning competence and are invalid indicators of the work clinicians actually do in the practice setting [18]. However, the development of expertise is largely a matter of knowledge reorganisation and cognitive processes to fit the demand of tasks within the domain of expertise [2]. SC test was developed in the perspective to explore the capacity of data interpretation when making clinical decisions, clearly a skill that belongs more to clinical competence than the simple recall of factual data.

Our results show an increase in the mean scores on the SC test of groups with different clinical expertise, with the students receiving lower scores than the residents and the residents receiving lower scores than the urologists. These observations were similar in the two sites, independently of the language of test and of the learning environment. This may mean that the SC test measures a dimension for which, as one should expect, experienced clinicians get better scores than less experienced subjects. This supports the construct validity of the instrument when applied in the field of urology. The Cronbach alpha reliability coefficient obtained in this test is 0.79, which is excellent in a context of formative assessment. Content validity of the SC test and relevance of competing hypotheses were validated by the panel of experts. The data of this test specifically addressed for clinical competence assessment in urology was in agreement with previous studies performed in other fields with a larger number of participants [12–15].

In SC test, examinees have to solve problems belonging to the real urology practice and must answer questions that expert urologists consider of crucial importance in the process of solving that problem. SC test has another major advantage for a testing method of being relatively easy to construct and use as well as being machine-scorable. It can be either paper or computer based. It is also important to mention that relatively modest resources are required to develop it. Another advantage of the scoring system is that the test can be used in situations where there is no consensus among experts, in literature or in practice.

It is well known that assessment has a strong impact on learning. Students and residents adapt what they learn to what they believe will be tested. SC test could reflect professional reality and is problem-solving oriented; hence it should influence the adaptation of students and residents' learning activities in that direction.

This exploratory study has some limitations. The first is its small cohort of participants. Hence, there is probably a lack of power in the statistical analysis to in fact establish the absence of significant interaction between levels of expertise and training. The trend of the two curves in Fig. 1 confirms that the score progression with the clinical expertise of group parti-

cipants is the same in the two sites as well as the absence of interaction for the students and the urologists. The graphic illustration nevertheless suggests a possible interaction for the residents. These data confirm the necessity of further investigation with larger groups of participants.

Another limitation is that one urologist had a relatively low score. This result might indicate that some urologists have a clinical reasoning process that differs from that of the majority. As regards the student with a higher score, this may be explained by the fact that he had previous on call duties as a resident in a surgical emergency unit for a few weeks. Therefore, his clinical exposition and training were most likely higher than the other students. The elimination of these disagreeing scores increased the means and decreased the standard deviations of the two groups of participants but had no influence on test reliability. It was decided to keep these participants in the analyses. In this study, experts benefited from a systematic advantage because by definition they received a credit on each item. The goal of this exploratory study was to assess if this tool is able to detect different levels of expertise in urology. Despite its limits, our scoring system was sufficient to detect this effect.

One other important limitation of this kind of assessment is well established. SC test is a written examination and could not adequately evaluate important skills, such as interviewing and interpersonal skills, physical examination skills and technical skills. Other evaluation tools, such as objective structured clinical examination, have proven useful for the assessment of clinical skills in urology, but they are not sufficiently valid to adequately assess clinical reasoning capacity [19]. Clinical competence is a multi-dimensional entity. No single assessment method can adequately measure it. SC test explores the capacity of data interpretation in the making of clinical decision and should be used in complement with other assessment tools.

5. Conclusions

The results show an increase in the mean scores on the SC test of groups with different clinical expertise in urology. This indicates that SC test measures a dimension for which experts obtained better scores than less experienced subjects. This is a strong argument in favour of the construct validity of the tool. The data also show that the test developed in urology has a high reliability coefficient, despite the fact that it was used in two different learning environments. Our data also favour the idea that during urological training, organised knowledge structures adapted to clinical tasks progressively build up. This type of examination appears promising and warrants further investigation to confirm its value as an assessment tool of clinical competence in urology.

Acknowledgements

The authors thank Richard Medeiros for his valuable advice in editing the manuscript.

References

[1] Norman GR. Defining competence: a methodological review. In: Neufeld VR, Norman GR, editors. Assessing clinical competence, vol. 7. New York: Springer, 1985. p. 15–35.

[2] Bordage G. Elaborated knowledge: a key to successful diagnostic thinking. Acad Med 1994;69:883–5.

[3] Elstein AS, Shulman LS, Sprafka SA. Medical problem-solving: an analysis of clinical reasoning. Cambridge (MA): Harvard University Press, 1978.

[4] Barrows HS, Norman GR, Neufeld VR, Feightner JW. The clinical reasoning of randomly selected physicians in medical general practice. Clin Invest Med 1982;5:49–55.

[5] Feltovich PJ. Expertise: reorganizing and refining knowledge for use. Prof Educ Res Notes 1983;4:5–9.

[6] Custers JFM, Regher G, Norman GR. Mental representations of medical diagnostic knowledge: a review. Acad Med 1996;71:S55–61.

[7] Regher D, Norman GR. Issues in cognitive psychology: implications for professional education. Acad Med 1996;71:988–1001.

[8] Zeitz CM. Some concrete advantages of abstraction: how experts' representations facilitate reasoning. In: Feltovich PJ, Ford KM, Hoffman RR, editors. Expertise in context: human and machine. Memlo Park (CA): AAAI press, 1997. p. 43–65.

[9] Feltovich PJ, Barrows HS. Issues of generality in medical problem-solving. In: Schmidt HG, De Volder ML, editors. Tutorials in problem-based learning: a new direction in teaching the health professions. Assen (Holland): Van Gorcum, 1984. p. 128–42.

[10] Schmidt HG, Norman GR, Boshuizen HPA. A cognitive perspective on medical expertise: theory and implications. Acad Med 1990;65:611–21.

[11] Charlin B, Tardif J, Boshuizen HPA. Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. Acad Med 2000;75:182–90.

[12] Charlin B, Brailovsky CA, Roy L, Van der Vleuten CP. The script concordance test: a tool to assess the reflective physician. Teach Learn Med 2000;12:189–95.

[13] Charlin B, Brailovsky CA, Leduc C, Blouin D. The diagnostic script questionnaire: a new tool to assess a specific dimension of clinical competence. Adv Health Sci Educ 1998;20:51–8.

[14] Charlin B, Brailovsky CA, Brazeau-Lamontagne L, Samson L, Leduc C. Script questionnaires: their use for assessment of diagnostic knowledge in radiology. Med Teach 1998;20:567–71.

[15] Brailovsky C, Charlin B, Beausoleil S, Coté 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] Marshall KG, Brailovsky CA, Grand'Maison P. French–English, English–French translation process of an objective structured clinical examination (OSCE) used for licensing family physicians in Quebec. Teach Learn Med 1995;7:115–20.

[17] Van der Vleuten CPM. The assessment of professional competence: development, research and practical implications. Adv Health Sci Educ 1996;1:41–67.

[18] Mc Gaghie WC. Evaluating competence for professional practice. In: Curry L, Wegin JF, editors. Educating professionals: responding to new expectations for competence and accountability. San Francisco: Jossey-Bath, 1993. p. 229–61.

[19] Sibert L, Grand'Maison P, Doucet J, Weber J, Grise P. Initial experience of an objective structured clinical examination in evaluating urology residents. Eur Urol 2000;37:621–7.