

Using the Script Concordance Test to Evaluate Clinical Reasoning Skills in Psychiatry

François Kazour¹ · Sami Richa¹ · Marouan Zoghbi¹ · Wissam El-Hage² · Fady G. Haddad¹

Received: 24 September 2015 / Accepted: 18 March 2016
© Academic Psychiatry 2016

Abstract

Objectives Although clinical reasoning is a major component of psychiatric training, most evaluating tools do not assess this skill properly. Clinicians mobilize networks of organized knowledge (scripts) to assess ambiguous or uncertain situations. The Script Concordance Test (SCT) was developed to assess clinical reasoning in an uncertainty context. The objective of this study was to test the usefulness of the SCT to assess the reasoning capacities of interns (7th year medical students) during the psychiatry training.

Methods The authors designed a SCT for psychiatry teaching, adapted to interns. The test contained 20 vignettes of five questions each. A reference panel of senior psychiatrists underwent the test, and we used their scoring as a reference for the student group. The SCT assessed the competence of students at the beginning and the end of their training in psychiatry.

Results A panel of 10 psychiatrists and 47 interns participated to this study. As expected, the reference panel performed significantly ($p < 0.001$) better (79.4 ± 5.1) than the students on the SCT. Interns improved significantly ($p < 0.001$) their scores between the beginning (58.5 ± 6.2) and the end (65.0 ± 5.3) of their psychiatry rotation. The students improved significantly ($p < 0.001$) their scores between the beginning and the end of the training (6.4 ± 4.8).

Conclusions This is the first study using the SCT in psychiatry. This study shows the feasibility of this procedure and its utility in the field of psychiatry for evaluating medical students in their clinical reasoning competence. It can provide a valid alternative to classical evaluation methods.

Keywords Script concordance test · Psychiatry · Evaluation · Reasoning · Intern

The capacity of medical training programs to evaluate clinical reasoning is limited and cannot be standardized and generalized to all students. The script concordance test (SCT) was developed by Charlin et al. to evaluate clinical reasoning in the context of uncertainty [1]. According to script theory, clinicians will use in uncertainty contexts networks of knowledge in order to reach solutions for clinical problems. Every new information given in clinical situations will contribute to either confirm or eliminate possible hypotheses.

The SCT consists of different vignettes followed by several items each. The vignette is a brief clinical case used to test diagnosis hypotheses, treatment, or management options. Each following item explores a possible hypothesis for the script.

The SCT was successfully studied in several medical fields including emergency medicine [2], dermatology [3], and radiation oncology [4]. This test has not been studied so far in psychiatry.

The primary objective of this study was to develop a SCT in the field of psychiatry and to introduce a new model of evaluation for mental health educators. The secondary objective of this study was to use this test for the evaluation of interns during the seventh year of medical studies at the beginning and at the end of their 1-month training in psychiatry.

✉ François Kazour
francoiskazour@hotmail.com

¹ Faculty of Medicine, Saint Joseph University, Beirut, Lebanon

² Université François-Rabelais de Tours, Inserm UMR U930, Tours, France

Methods

This study received the approval of the committee of pedagogy of the Faculty of Medicine—Saint Joseph University, Beirut, Lebanon.

The SCT is composed of short vignettes describing clinical situations in psychiatry. The clinical situations reflect real cases (incomplete data in a context of uncertainty) with uncertain diagnosis, or different management possibilities. Five items follow each vignette. Each item is associated to a

possible diagnosis or a treatment option, followed by new relevant information whose impact on the initial option is measured with a five-point Likert scale, from -2 (least probable) to $+2$ (most probable). The Likert scale measures if the new information makes the proposed diagnosis more or less probable, and the proposed treatment option more or less useful. Examples are given in Box 1.

Choices on the Likert scale are subject to a modal transformation [5]. To define the correct answers, the test was taken by a reference panel of senior experienced psychiatrists

Box 1: Examples of clinical vignettes included in the script concordance test: a. Diagnostic vignette; b. Treatment vignette

a. Vignette 1

Roger is 36-year-old man who presents with acute agitation that had been present for 48 h. He has an incoherent speech. He believes that his brothers want to hurt him because he's a powerful and wealthy man. He says he's the friend of the prime minister and has seen him on several occasions. His brother says he drinks 2 glasses of whiskey every night and takes cocaine with alcohol once every month.

-2 = rules out or almost ruled out
 -1 = less probable
 0 = neither less nor more probable
 $+1$ = more probable
 $+2$ = certain or almost certain

If you were thinking of:

- Q1. Acute psychotic episode
 Q2. Bipolar disorder
 Q3. Schizophrenia
 Q4. Substance induced psychosis
 Q5. Acute mania

And then you find:

- Auditory hallucinations
 He has racing thoughts, spends a lot of money since 1 week, but manages to sleep 5 to 6 h every night
 He saw "Jesus" yesterday who told him that "he was the chosen one"
 No drugs detected on urine tests
 Family history of depression

This hypothesis becomes:

- -2 -1 0 $+1$ $+2$
 -2 -1 0 $+1$ $+2$
 -2 -1 0 $+1$ $+2$
 -2 -1 0 $+1$ $+2$
 -2 -1 0 $+1$ $+2$

b. Vignette 2

A 32-year-old woman with a history of schizophrenia presents to consultation with her sister. She used to have persecutory delusions with auditory hallucinations. She was admitted 3 times to the psychiatric unit. The patient has been treated successfully for the past 3 years with Risperidone 4 mg/day. She has hypertension and smokes 20 cigarettes everyday. Her mother is treated for Alzheimer disease with Donepezil. She is presenting recurrent persecutory delusions for the last 2 weeks. She is isolated and refuses to talk to her family members.

-2 = useless
 -1 = less useful
 0 = neither less nor more useful
 $+1$ = useful
 $+2$ = very useful

If you were considering the utility of the following treatment...

- Q31. Risperidone 8 mg/day
 Q32. Clozapine
 Q33. Duloxetine
 Q34. Quetiapine
 Q35. Olanzapine

...and the following new information were to become available...

- She never stopped treatment but it used to sedate her
 She also was treated unsuccessfully with Ziprazidone
 She's hearing voices commanding her to kill herself
 She stopped her treatment because it made her gain weight
 She stopped her treatment because it induced tremors

...You would then consider this treatment...

- -2 -1 0 $+1$ $+2$
 -2 -1 0 $+1$ $+2$
 -2 -1 0 $+1$ $+2$
 -2 -1 0 $+1$ $+2$
 -2 -1 0 $+1$ $+2$

Table 1 Scores of Reference panel and students on the script concordance test of psychiatry

	Reference panel (<i>n</i> = 10)	Students pre-rotation (<i>n</i> = 47)	<i>p</i>
Mean	79.42	58.47	<0.001
Standard deviation	5.13	6.18	
Minimum	72.03	45.57	
Maximum	87.5	75.66	
Confidence interval (95 %)	77.8–81.04	57.3–59.47	

(between 5 and 18 years of experience). For any question, the most frequently chosen answer by the expert panel gave a credit of 1 point, while the never chosen answer gave 0 point. Other answers gave partial credit (the number of experts choosing this answer divided by the number of experts choosing the most frequent answer).

The test was designed to evaluate interns during the seventh year of medical studies. Clinical situations chosen for the SCT covered psychiatric disorders and treatment within interns' psychiatric program. The test was developed by a psychiatrist and reviewed by a senior psychiatrist and two senior residents to ensure that vignettes and items were clearly written and appropriate to the interns' level of competence. The test included 100 questions (20 vignettes followed each by 5 independent questions). The content of vignettes and questions covered the seventh year of medicine psychiatric program and included the following subjects: Schizophrenia and psychosis (3 vignettes); Depressive disorders (2 vignettes); Anxiety disorders (3 vignettes); Bipolar disorder (3 vignettes); Substance use disorders (3 vignettes); Somatoform disorders (2 vignettes); Child and adolescent disorders (2 vignettes); Dementia (2 vignettes). The test explored two axes of competence: Diagnostic clinical reasoning (15 vignettes) and therapeutic reasoning (5 vignettes). Each student and expert was given 90 min to take the test. The test was taken in a classroom without access to any reference. The language of the test was English.

Table 2 Scores of students before and after notation on the script concordance test of psychiatry

	Students pre-rotation (<i>n</i> = 47)	Students post-rotation (<i>n</i> = 47)	<i>p</i>
Mean	58.47	65.04	<0.001
Standard deviation	6.18	5.3	
Minimum	45.57	52.17	
Maximum	75.66	78.16	
Confidence interval (95 %)	57.3–59.5	64.3–65.8	

A reference panel included ten practicing psychiatrists affiliated to two universities (Saint Joseph University and The Lebanese University). They answered the SCT after the concept was explained to them.

The student group consisted of interns during their psychiatry training. Every month, four new interns have their psychiatry rotation at the Psychiatric Hospital of the Cross. All students doing their internship during a 12-month period were recruited. All students had previously pursued psychiatry courses (approximately 40 h of lectures and clinical case discussions) during the last 2 to 4 years. The same SCT was administered to interns on the first and last days of their 1-month rotation. All interns were informed about the study's objectives, the principle of the SCT, and consented to participate. The psychiatry internship is a training of 1 month, where interns are confronted to different clinical cases of psychiatry. Except the weekly staff meetings and article presentations, no formal lectures are given during this internship.

The reference panel scores were established using the following rules: most frequent answer gave a credit of 1 point; the never chosen answer gave 0 point; other answers gave partial credit. The students' scores were determined using the reference panel scores. For each question, an item analysis was performed regarding its ability to discriminate between students and its impact on the overall reliability. Thus, each question can be categorized as "good" (correlation coefficient >0.20), "fair" (correlation coefficient between 0.10 and 0.20) and "bad" (correlation coefficient <0.10).

All data was entered using Microsoft Excel. The internal coherence test coefficient was evaluated with the Cronbach's alpha test. This coefficient (values between 0 and 1) indicates greater homogeneity as values approach 1. Good reliability is indicated when the coefficient is above 0.80 [6]. Scores were determined within 3 groups: reference psychiatrists, interns before training, and interns after training. The statistical analysis included the scores of the three groups. Quantitative values were described by the mean, standard deviation and range. Confidence intervals of 95 % were calculated. Comparisons between the students (before and after training) and reference groups' scores were undertaken using the Mann-Whitney test. *p* values <0.05 were considered statistically significant.

Results

The test composed of 100 questions consisted of 20 vignettes (15 diagnostic and 5 therapeutic) with five questions each. The reference panel consisted of ten psychiatrists in the academic field, with 5 to 18 years of experience.

Experts gave extreme values (−2 (least probable) or +2 (most probable)) to 27 % of questions (27 questions) that have extreme modal answers.

Forty-eight students participated to this study. They were evaluated at the first and last days of their training, at 1-month interval. One student was excluded for having a previous clinical training in psychiatry. We analyzed the data of 47 students.

The reference group had a mean score (79.42) significantly higher than the mean score of the students (58.47) before their training ($p < 0.001$; Table 1). After training, the students improved 11 % of their scores on the SCT (65.04), significantly higher than the scores before the training ($p < 0.001$; Table 2). The mean difference between pre- and post-training scores was 6.57 ± 4.78 (0.57–18).

Cronbach's α coefficient of the SCT was 0.79. The SCT contained 19 bad items, 37 fair items and 44 good items. After removing 19 questions (bad items) with correlation coefficients < 0.10 , Cronbach's α coefficient increased to 0.87 and the number of bad items decreased to 3 (3.7 %).

Discussion

This study presents the SCT for mental health educators as an evaluation tool used for the first time in the field of psychiatry. It shows the feasibility of the SCT, and gives some evidence for its reliability in psychiatry. The cases used in our study were taken from real clinical situations and adapted to the level of knowledge of our students. Our test included 15 diagnostic and 5 therapeutic vignettes. This distribution was a reflection of the interns' theoretical teaching that focused more on diagnostic evaluation (75 %) rather than on therapeutic procedures (25 %) [7]. Psychiatrists forming the experts' panel were recruited from the psychiatry departments of two universities in Lebanon.

Despite the small sample size, Cronbach's α coefficient indicated the reliability of our test. As expected, expert psychiatrists had higher scores than the students on the SCT. This comparison shows the difference between the students' baseline score and the higher mean score of experts. This difference reflects the differences in knowledge and clinical expertise between the two groups. This difference shows that the SCT can distinguish participants according to their clinical experience [8]. The internal value of the test (Cronbach α) was good (0.87), reaching the value needed for high-stake evaluations [6]. Interns had significantly better scores after 1 month of clinical training in psychiatry. The SCT was able to measure this improvement in clinical reasoning gained by interns through their clinical training. Clinical reasoning includes a complex set of processes that may need several tools for their evaluations. The SCT allows us to evaluate the ability to interpret clinical data in a given context. In order to

complete the evaluation of clinical reasoning competence, SCT should be associated to other evaluation tools such as objective structured clinical examination (OSCE), case studies, or supervised clinical examinations.

SCT has been used in other medical fields with promising results [2–4, 9]. As far as we know, this is the first time it is used in psychiatry. Our test was designed following published guidelines [5].

We chose to have an expert panel of psychiatrists with clinical and university teaching experiences. All experts had hospital experience. This corresponds to the nature of the students' training, which is hospital inpatient training. The number of experts in the reference panel is 10, which is considered acceptable [5]. Increasing the number of panel members to 15 or more would have improved the test's reliability, but this would be counterbalanced by a decreased feasibility.

Some authors suggested that responses that intentionally avoided extreme values could be a threat to the validity of SCT, and recommended increasing the proportion of questions with extreme modal answers [10, 11]. Avoidance of extreme responses may increase students' scores and be a potential bias in SCT. In our test, 27 % of questions had extreme modal answers (–2 (least probable) or +2 (most probable)). If this proportion were to be increased to 50 %, we would be able to avoid this bias [11].

Many university programs rely on simulators or real patients to evaluate their students' clinical reasoning ability. These methods often require extensive resources and time. They cannot evaluate more than one student at a time, and may lack objectivity of reliability since they depend on the patient's or the simulator's performance during each evaluation. The SCT allows exploring several clinical situations over a short duration. It is objective, reliable, and can be repeated for an unlimited number of times. It mimics real clinical cases and helps in evaluating the clinical competence gained by students during their training. This test improves considerably the evaluation system in clinical context since it allows having the same evaluation for an unlimited number of students in a short period of time.

This study has several limitations that should be taken into consideration. It was conducted on one site, reflecting the evaluation of students in just one model of psychiatry rotation and cannot be generalized to all types of rotations. Our experts' group consisted only of 10 psychiatrists. It would have been better to have at least 15 experts, but since Lebanon does not count more than 60 psychiatrists [12], it was difficult to have a larger group. It would have been better to compare the students' group to another group of senior psychiatrists in order to confirm the validity of our test. Using the experts' group as a comparator is a limitation in this study. It would have

been better to have interns reviewing the test instead of residents, but since all interns were participants and the sample size was relatively small, we were not able to have other interns as reviewers. We did not account the test-retest effect, since one of the objectives of this study was to compare pre- and post-rotation evaluations by using the same SCT. The interns were informed that they would have to take the same test at the beginning and at the end of the rotation to evaluate the improvement of their clinical reasoning. Since it was a trial study, they were also informed that this test would not have a direct impact on their rotation's grades. Scores and questions were not released to the interns between the two evaluations. Students were expected to discuss with other students or residents about the vignettes they had. They were also expected to search during their rotation for possible correct answers since the vignettes covered major parts of their program. Informing the students about having the same test at the end may be a bias to the results of the study; however we estimate that it contributed to the learning process during this rotation. Finally, the sample size of the students' group is also an important limitation to take into consideration before generalizing results to all medical students.

Implications for Educators

- Classical evaluation methods cannot evaluate properly students' clinical reasoning in complex situations.
 - The script concordance test is objective, reliable, and easy to develop. It mimics real-life clinical situations and can be reproduced for an unlimited number of students.
 - The script concordance test should be associated with other testing methods like case studies, OSCE, supervised clinical examinations, or multiple-choice questions in order to have a global evaluation of clinical competence.
-

Acknowledgments The authors are grateful to the Psychiatric Hospital of the Cross for providing the training field for students who participated to this study.

Compliance with Ethical Standards

Disclosures On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. 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.
2. Fournier JP, Thiercelin D, Pulcini C, et al. Evaluation du raisonnement clinique en médecine d'urgence: les tests de concordance de script décèlent mieux l'expérience que les questions à choix multiples a contexte riche. *Pédagogie Médicale.* 2006;7:20–30.
3. Bursztejn AC, Cuny JF, Adam JL, Sido L, Schmutz JL, de Korwin JD, et al. Usefulness of the script concordance test in dermatology. *J Eur Acad Dermatol Venereol.* 2011;25(12):1471–5.
4. Lambert C, Gagnon R, Nguyen D, et al. The script concordance test in radiation oncology: validation study of a new tool to assess clinical reasoning. *Radiat Oncol.* 2009;4:7.
5. Fournier JP, Demeester A, Charlin B. Script concordance tests: guidelines for construction. *BMC Med Inform Decis Making.* 2008;8:18.
6. Charlin B, Gagnon R, Sibert L, et al. Le test de concordance de script, un instrument d'évaluation du raisonnement clinique. *Pédagogie Médicale.* 2002;3:135–44.
7. Richa S, Baladi A, Kesrouani A, Baddoura C. Teaching psychiatry for fifth-year medical students at Saint-Joseph University, Beirut: evaluation of the four-step method. *Acad Psychiatry.* 2009;33(4):344–5.
8. LaDuca A. Validation of professional licensure examinations: profession, theory, test design and construct validity. *Eval Health Prof.* 1994;17:178–97.
9. Haddad FG, Gaspard D, Nasr M, Abou Jaoude S, Nemr E. Etude exploratoire de l'utilisation du test de concordance de script pour l'évaluation des apprentissages de résidents de médecine interne au Liban. *Pédagogie Médicale.* 2008;3:135–40.
10. Lineberry M, Kreiter CD, Bordage G. Threats to validity in the use and interpretation of script concordance test scores. *Med Educ.* 2013;47(12):1175–83.
11. See KC, Tan KL, Lim TK. The script concordance test for clinical reasoning: re-examining its utility and potential weakness. *Med Educ.* 2014;48(11):1069–77.
12. Lebanese Psychiatric Society. Retrieved on December 12, 2015 from : <http://www.lpsonline.org/psychiatrists-psychiatres/>.