THE AGA KHAN UNIVERSITY

eCommons@AKU

Section of Urology

Department of Surgery

January 2019

Assessing clinical reasoning skills using Script Concordance Test (SCT) and extended matching questions (EMQs): A pilot for urology trainees

SYED MUHAMMAD NAZIM

Jamsheer J Jehangir Talati Professor
Aga Khan University, jamsheer.talati@aku.edu

SHEILA PINJANI
Aga Khan University

Syed Raziuddin Biyabani
Aga Khan University, raziuddin.biyabani@aku.edu

Hammad Ather
Aga Khan University, hammad.ather@aku.edu

See next page for additional authors

Follow this and additional works at: https://ecommons.aku.edu/pakistan_fhs_mc_surg_urol

Authors

SYED MUHAMMAD NAZIM, Jamsheer J Jehangir Talati Professor, SHEILA PINJANI, Syed Raziuddin Biyabani, Hammad Ather, and JOHN J. NORCINI

Assessing clinical reasoning skills using Script Concordance Test (SCT) and extended matching questions (EMQs): A pilot for urology trainees

SYED MUHAMMAD NAZIM[1*], JAMSHEER J TALATI[1], SHEILA PINJANI[2], SYED RAZIUDDIN BIYABANI[1], MUHAMMAD HAMMAD ATHER[1], JOHN J. NORCINI[3]

[1]Section of Urology, Department of Surgery, The Aga Khan University Hospital, Karachi, Pakistan; [2]Department for Educational development, The Aga Khan University Hospital, Karachi, Pakistan; [3]Foundation for Advancement of International Medical Education and Research (FAIMER), United States of America

Introduction: Clinical reasoning skill is the core of medical competence. Commonly used assessment methods for medical competence have limited ability to evaluate critical thinking and reasoning skills. Script Concordance Test (SCT) and Extended Matching Questions (EMQs) are the evolving tests which are considered to be valid and reliable tools for assessing clinical reasoning and judgment. We performed this pilot study to determine whether SCT and EMQs can differentiate clinical reasoning ability among urology residents, interns and medical students.

Methods: This was a cross-sectional study in which an examination with 48 SCT-based items on eleven clinical scenarios and four themed EMQs with 21 items were administered to a total of 27 learners at three differing levels of experience i.e. 9 urology residents, 6 interns and 12 fifth year medical students. A non-probability convenience sampling was done. The SCTs and EMQs were developed from clinical situations representative of urological practice by 5 content experts (urologists) and assessed by a medical education expert. Learners' responses were scored using the standard and the graduated key. A one way analysis of variance (ANOVA) was conducted to compare the mean scores across the level of experience. A p-value of <0.05 was considered statistically significant. Test reliability was estimated by Cronbach α. A focused group discussion with candidates was done to assess their perception of test.

Results: Both SCT and EMQs successfully differentiated residents from interns and students. Statistically significant difference in mean score was found for both SCT and EMQs among the 3 groups using both the standard and the graduated key. The mean scores were higher for all groups as measured by the graduated key compared to the standard key. The internal consistency (Cronbach's α) was 0.53 and 0.6 for EMQs and SCT, respectively. Majority of the participants were satisfied with regard to time, environment, instructions provided and the content covered and nearly all felt that the test helped them in thinking process particularly clinical reasoning.

Conclusion: Our data suggest that both SCT and EMQs are capable of discriminating between learners according to their clinical experience in urology. As there is a wide acceptability by all candidates, these tests could be used to assess and enhance clinical reasoning skills. More research is needed to prove validity of these tests.

Keywords: Clinical reasoning, Decision making, Education, Concordance, Urology

J Adv Med Educ Prof. January 2019; Vol 7 No 1

7

Introduction

Clinical competence (1) is a multi-dimensional, complex construct, representing the ability of a professional to use clinical judgment skills and reasoning in addition to knowledge to solve complex problems in a specific context (2). The currently used methods to assess competence in knowledge based application include multiple choice questions (MCQs), short answer questions (SAQs) and traditional unstructured viva. There is considerable debate about the validity, reliability and standardization (3, 4) of these tools which have limited ability to assess critical thinking and reasoning skills.

Problems encountered during the professional practice do not always have straightforward algorithmic solutions but require judgment and insight which can't be measured by conventional tools (5). The most appropriate tools suggested for assessment of judgment and clinical reasoning are the key feature problems (KFPs), extended matching items (EMIs) and script concordance test (SCT) (6).

Originally developed in the field of medicine, in SCT learners are presented with a clinical scenario followed by reveal of a new piece of information. This unique tool assesses the clinical reasoning and data interpretation skills for real clinical scenarios encountered in clinical practice under condition of uncertainty (7, 8). The SCT cases fall into various categories such as diagnosis, treatment and investigations related to a particular clinical condition (9, 10).

The extended matching questions, a form of multiple choice questions (MCQs) are found to be superior to traditional MCQs in assessing learners' clinical reasoning and problem solving abilities with higher reliability of scoring. The have less "recognition effect" and therefore, less chance for learners to guess the correct response (11, 12).

Both tests have been found to be valid and reliable tools which can discriminate levels of practice between experts, residents and medical students and can evaluate clinical reasoning skills (13).

We observed that trainees, despite a good knowledge base, are shaky deciding about management in uncertain situations and encountering them with real case scenarios via regular practice with a paper based assessment will enhance their decision making ability and confidence level. A more relevant written test for assessing clinical decision making i.e. EMQs and SCT in urology training program will also strengthen the existing assessment methods.

The purpose of this study was to assess the clinical reasoning skills of urology trainees using script concordance test (SCT) and extended matching questions (EMQs) and whether these tests could differentiate clinical reasoning ability among urology residents, interns and medical students.

Methods

This was a cross-sectional (pilot) study conducted by the section of Urology and department for educational development at the Aga Khan University after obtaining institutional review board approval. A non-probability convenience sampling was done. We selected 3 categories of participants according to their clinical experience in urology. The residents with at least one year of clinical experience, interns with an experience of 3 months and fifth year medical students who had completed 3 weeks of urology rotation were eligible for the study. All eligible urology residents and interns volunteered for the study. Twenty six 5th year medical students who completed the 3 week urology rotation in their surgery clerkship module were asked to participate and twelve of them volunteered for the test. The informed consent form was obtained from all the participants. It was a pencil and paper based test administered during a 90-minute time period in a proctored setting. The test was designed including a combination of eleven Script Concordance (SC) scenarios with 48 items & four themed Extended Matching Questions with 21 items.

Test construction

We constructed SCT vignettes (scenarios) comprising of a stem and between 4-5 items (questions) (14). The clinical vignette was made in order to keep the scenario authentic but requiring experience, reasoning skills and reflection in action. All the vignettes and questions were constructed from the clinical situation representative of urological practice by 5 content experts (urologists) and assessed by a medical education expert in the first round prior to the administration of the test (9). For each clinical situation, the questions were asked and actions taken to arrive at the diagnosis or decide on the adequate management of a patient. Each question had an answer key in the form of a 5-point Likert type scale (-2, -1, 0, +1, and +2), ranging from absolutely contraindicated (-2) to absolutely indicated (+2). Both simple (standard key) and aggregate (graduated key) scoring methods were used for scoring. Simple scoring using standard key was done by choosing the best response selected by experts and giving it a score

of 1 (consensus score) while rest of the responses were scored as zero. Aggregate scoring method using graduated key (15) was also used to include variability of answers by experts showing their reasoning process. Any answer given by an expert had an intrinsic value and was not discarded. A maximum score of 1 (modal answer) was given to responses chosen by most of the experts while other responses were given partial credit depending upon fraction of experts choosing them and dividing it by the modal value for the item. An example of a SCT case and the scoring grid using aggregate method are shown in Tables 1 and 2.

The EMQs were organized into 4 parts i.e. theme, an alphabetical option list, a lead in statement (Question) and a clinical scenario (Items) (12). An option could be correct for more than one question or may not be correct for any question. A total of 4 themed EMQs with between 9-12 option in each theme and a total of 21 items were developed by the same panel of experts. There was no negative marking for the EMQs.

Pre-test familiarization of candidates with the testing and scoring system

Since the learners were not fully aware of these methods of assessment, a voluntary practice session was held with them to make them familiar with the format.

Statistical analysis

The data was analyzed using statistical package for the social sciences software (SPSS), version 22. The scores were described by means, standard deviation and minimum and maximum scores. Validation for the test focused on the internal reliability (measured by Cronbach's alpha) and the ability of SCT to distinguish between different group of learners (i.e. Residents, interns and medical students). A One-way analysis of variance (ANOVA) was conducted to compare the difference in percent mean scores of all three groups of learners for EMQ and SCT. A p-value of <0.05 was kept as significant. A post HOC analysis of multiple comparisons between groups for SCT was also done. The test was immediately followed by a Focused Group Discussion (FGD) with the candidates.

Results

A total of 27 learners (9 urology residents, 6 interns rotating in urology and 12 fifth year medical students) participated in the test. All participants completed the questions within 90 minutes examination period.

Both SCT and EMQs successfully differentiated residents from interns and students in one-way analysis of variance (ANOVA). Significant difference in the mean score was found for script concordance test among residents, interns and students using aggregate scoring (graduated key) and simple scoring (using standard key) (one way ANOVA). The residents scored highest (66.96±4.61) followed by interns (54.48±12.68) and then students (54.40±6.05) on aggregate scoring. The mean scores were higher for all the groups as measured by graduated key compared to standard key (Table 3). On simple

---

Table 1: Case Description: Example of a case from diagnostic section of Script Concordance Test (SCT)

A 43 year old gentleman known case of diabetes mellitus (DM) and cardiomyopathy with ejection fraction (EF) 20% presented to emergency room with 2 weeks history of burning micturition, fever and 3 days history of right sided scrotal swelling. He is taking Tab.Warfarin and Digoxin. His pulse is 91/min, BP 85/40 mmHg, temperature 38.2° C and examination showed tender right hemi-scrotum.

If you were thinking of	Then you found on clinical presentation/ investigation	The hypothesis becomes
1. Testicular abscess	Bruising of scrotum	-2 , -1 , 0 , +1, +2
2. Scrotal hematoma	INR of 1.8	-2 , -1 , 0 , +1, +2
3. Epididymo-orchitis	Normal U/S scrotum	-2 , -1 , 0 , +1, +2
4. Stangulated hernia	Absent cough impulse	-2 , -1 , 0 , +1, +2
5. Testicular tumor	Normal tumor markers	-2 , -1 , 0 , +1, +2

Where: -2 Ruled out or almost ruled out; -1 Less probable; 0 Neither less or more probable; +1 More probable; +2 Certain or almost certain

---

Table 2: Scoring grid: Aggregate method via graduated key to calculate weighted scores

Credit per Question: No. of experts (out of 5) who choose the respective answers (-2 to +2)					
Item	**-2**	**-1**	**0**	**+1**	**+2**
1	0 (0)	1 (0.33)	3 (1)	1 (0.33)	0 (0)
2	0 (0)	0 (0)	1 (0.33)	3 (1)	1 (0.33)
3	1 (0.33)	2 (0.66)	2 (0.66)	0 (0)	0 (0)
4	0 (0)	1 (0.33)	3 (1)	1 (0.33)	0 (0)
5	1 (0.33)	2 (0.66)	2 (0.66)	0 (0)	0 (0)

1=Modal answer <1are credits

Table 3: Comparison of Script concordance test (SCT) and Extended matching questions (EMQ) scores between urology residents, interns and medical students rotating in urology using both standard and graduated keys

Variables	Participants level of experience (n=27)			p
	Residents (9)	Interns (6)	Students (12)	
Script concordance test (SCT)				
Standard key (Mean±SD)	49.7±2.53	39.84±10.11	40.78±6.11	0.020
Graduated Key (Mean±SD)	66.96±4.61	54.48±12.68	54.40±6.05	0.009
Extended matching questions (EMQs)				
Standard key (Mean±SD)	52.38±14.11	28.53±19.96	40.76±10.36	0.016
Graduated Key (Mean±SD)	57.65±13.41	32.26±22.9	46.21±9.47	0.015

scoring, residents scored 49.7±2.53, interns had a score of 39.84±10.11, and students scored 40.78±6.11; the difference was statistically significant (p=0.020).

Statistically significant difference was also found for EMQs among residents, interns and students in both aggregate scoring (graduated key) and simple scoring (standard key) (one way ANOVA). Interestingly, on aggregate scoring, the students scored higher (46.21±9.47) compared to the interns (32.26±22.9); however, residents mean scores (57.65±13.41) were highest (Table 3)

A post HOC analysis of multiple comparisons between groups for SCT using the graduated key showed that there was significant difference between residents vs. students (p<0.001). There was no significant difference between residents vs. interns (p=0.079) and interns vs. students (p>0.999) (Dunnett test).

Similarly, for EMQs, post HOC analysis of multiple comparisons using graduated key showed significant difference between the residents and interns (p=0.011); however, no significant difference was found between residents vs. students (p=0.286) and interns vs. students (p=0.141) (Tukey test). The internal consistency (Cronbach's α) was 0.53 and 0.6 for EMQs and SCT, respectively.

Students completed the test in the shortest time. It was shown that different scoring methods did not affect learners' scores significantly for EMQs but there was difference in scoring between the graduated and standard key for SCT.

Regarding focus group responses, the students thought that EMQs could be used as an assessment tool but the SCT although provoked thinking process, required increased clinical exposure so it might not be very well suitable for under-graduates. Residents found SCT to be easier and interesting compared to EMQs. The ease of answering SCT may be due to its similarity to day-to-day clinical practice; and the difficulty in EMQ might be because of wide choices and basic science content. Regarding the overall experience of the test, more than two-thirds of

the participants were satisfied with regard to time, environment, instructions provided and the content covered. Nearly all learners felt that the test helped them in thinking process particularly clinical reasoning.

Discussion

The purpose of this paper was to describe the development and implementation of a tool for assessment of clinical reasoning in the field of urology. This was first of its kind at our institution.

The current methods of professional competence assessment include performance based methods (e.g. Objective Structured Clinical Exam'OSCE') or those seeking the solutions to well-defined problems (e.g. MCQs) (1). These methods of assessment identify the examinee's ability to recall a depth and breadth of factual knowledge from memory rather than the organization of knowledge (16). Besides the psychometric properties of these tests, these tests have failed to assess an individual's ability to think critically, reason and proceed in an unknown encounter (7).

The concept of SCT and EMQs is to explore students' understanding and organization of knowledge base. EMQs and SCT assess the "knows how" level of Miller's pyramid, (17) which can potentially complement the other assessment tools situated at the lower level i.e. "knows" (e.g. MCQs) and higher level i.e. "shows how" (e.g. OSCE) and "does" (Multi-source feedback).

Extended matching items have a clinical scenario, or vignette, a lead in and a long list of options (up to 16) to choose from. This long list serves to reduce the recognition effect present in MCQs and used best when there are a large number of similar actions/ decisions to choose from (18).

The script concordance approach is closely linked to a model of clinical reasoning and diagnosis known as hypothetico-deductive (HD) method (19) and allows objective assessment of

the trainees' clinical competence (20) against that of expert clinicians in context of ill-defined, uncertain situations. Physicians facing the clinical problems mobilize a set/network of knowledge (scripts) in order to understand the situation and make a clinical decision (8, 11).

Scripts are dynamic structures which are modified by each new encounter. Algorithmic reasoning or pure recall of factual knowledge can't be used to answer a properly fashioned SCT (21). The aim of SCT is to explore the physicians' knowledge base in terms of both the content and structure of knowledge (13).

SCT has 3 key design features (10);

1. The examinees are faced with ill–defined but authentic clinical situations and must choose from several realistic options.

2. The response format should reflect the way of processing the information in complex problem-solving situations.

3. The scoring should take into account the variability of responses by experts to that clinical situation.

SCT is case-based, which are described as short scenarios incorporating a bit uncertainty. These are followed by a set of questions consisting of 3 parts. The first part ("if you were thinking of") contains a hypothesis in the form of a diagnostic possibility or a management option. The second part ("and then you were to find") presents a new clinical finding such as a physical examination sign, a pre-existing condition or a laboratory or imaging study result. The third part ("this option would become") contains a 5-point Likert response scale capturing the examinees' decisions (9).

The scoring of SCT involves comparing the answers provided by examinees with those of a reference panel of experts. Different scoring systems are being used.

We used both simple and aggregate methods of scoring but feel aggregate scoring to be better as it incorporates the variability expressed by the panel of experts when confronted with ill-defined clinical problems into the scoring process. The reason for higher scores for all 3 level of experience on aggregate scoring was due to incorporation of partial credit to the scores compared to simple scoring by consensus which lacked the credit. In SCT, there was increase in mean scores with level of experience which indicates that our test actually measured the desired dimension of knowledge organization rather than factual knowledge which builds with experience. This supports the construct validity of our test.

For the construction of SC test items, a collaboration of a small number of experts is necessary (two at the initial stage of item production) is necessary (9). However, Gangnon (22) has shown that around 15 panel members are required to obtain acceptable reliability estimates (Cronbach's alpha) in high stakes examinations. Choosing a large number of experts is faculty and labor intensive and since our study was a pilot one, only 5 clinical experts could be included in the panel however, the content validity of both SCT and EMQ with relevance of competing hypothesis was validated by the expert panel.

The SCT has been used in the field of surgery besides other disciplines (21) to assess residents' clinical reasoning skills. Its validity and reliability in differentiating novices from experienced clinicians has been well documented (23) across different linguistic, cultural and learning environments. In the test, the examinees are presented with a series of patients' problems and are then asked to make diagnostic, investigative or therapeutic decisions based on the specific elements of information provided (16). This test has been reported to be easy to construct, machine-scorable and can be used in undergraduate, postgraduate or continuing medical education (CME) (13).

A number of studies have shown that SC test has interesting psychometric properties with regard to reliability, face and construct validity (21). This test may be used for both formative and summative assessment in a variety of medical fields such as surgery, gynecology, family medicine, radiology and many others.

Studies have been done to determine the optimal number of the items or cases to reach a reliability coefficient of 0.8 (21, 24, 25). Fournier, *et al.* (9) has proposed that an SCT should have 20 cases with 60 items in order to achieve a reliability coefficient (Cronbach alpha) higher than 0.75. Similarly, association for medical education in Europe (AMEE) (10) guide no. 75 also recommends that in SCT, three questions should be nested per case. However, other studies have shown that adding questions (items) rather than cases were more feasible in terms of reducing the workload of the test designers and reading time of learners and were also effective in increasing the test reliability (26). Various studies have used up to 5 items per clinical scenarios (19, 23, 27). Our study included only 11 case scenarios with 48 items in SCT and 4 EMQs with 21 items, which might be the reason for low reliability coefficient.

SCT and EMQ testing the clinical reasoning competence have the ability to overcome the

intermediate effect (26) which is a limitation for some of the test formats based on written simulations of clinical problem solving. The discrimination validity of this test has been shown in studies as the scores of individuals increase with their level of experience. Our study showed similar results with residents having more experience scored higher compared to interns and students.

Our study had several limitations. First, the number of learners who took the test was relatively small and spread out over different years of training, with probable lack of power in statistical analysis describing interaction between levels of experience and scoring. The results may therefore not be generalized to the other settings. Second, the outcome measures were assessed based on differences by years of training which might exist for simple knowledge, clinical skills and other competences as well and may not be specific to clinical reasoning. As mentioned above, the number of experts in the panel and the number of clinical scenarios and items were also less in both SCT and EMQ affecting the test reliability. We also did not evaluate other methods which could affect the reasoning and decision making skills like interpersonal, physical examination and technical skills for which other tools such as objective structured clinical examination (OSCE) are done. We also did not make any comparison with the learners' end of clerkship/rotation or in-service examination scores.

The most important determinant of validity relates to consequences or educational impact. At present very little is known about this aspect of SCT. Development of clinical reasoning tests in residency programs might be useful in identifying the deficiencies in learning at early years of clinical practice. This could be used for constructive feedback and focus in teaching, resulting in improved confidence in decision making.

Conclusion

Our data suggest that both SCT and EMQs are capable of discriminating between learners according to their clinical experience in urology. As there is a wide acceptability by all candidates, these tests could be used to assess and enhance clinical reasoning skills. More research is needed to prove validity of these tests.

Conflict of Interest: None declared.

References

1. Sharma N. Redesigning competency based medical education in a world of many team players. Med Teach. 2017; 22:1.
2. Loughlin M, Bluhm R, Buetow S, Borgerson K, Fuller J. Reasoning, evidence, and clinical decision-making: The great debate moves forward. J Eval Clin Pract. 2017;23(5):905-14.
3. Veloski JJ, Rabinowitz HK, Robeson MR, Young PR. Patients don't present with five choices: an alternative to multiple-choice tests in assessing the physicians' competence. Acad Med.1999;74:539-46.
4. Schuwirth LW, Verheggen MM, Van der Vleuten C, Boshuizen HP, Dinant GJ. Do short cases elicit different thinking processes than factual knowledge questions do? Med Educ. 2001;35:348-56.
5. Bagg W, Clark K. Professionalism: medical students, future practice and all of us. Intern Med J. 2017;47(2):133-4.
6. Banning M. A review of clinical decision making: models and current research. J Clin Nurs. 2008;17(2):187-95.
7. Wan SH. Using the script concordance test to assess clinical reasoning skills in undergraduate and postgraduate medicine. Hong Kong Med J. 2015;21(5):455-61.
8. 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.
9. Fournier JP, Demeester A, Charlin B. Script concordance tests: guidelines for construction. BMC Med Inform Decis Mak. 2008;8:18.
10. Lubarsky S, Dory V, Duggan P, Gagnon R, Charlin B. Script concordance testing: from theory to practice: AMEE guide no. 75. Med Teach. 2013;35(3):184-93.
11. Hornos EH, Pleguezuelos EM, Brailovsky CA, Harillo LD, Dory V, Charlin B. The practicum script concordance test: an online continuing professional development format to foster reflection on clinical practice. J Contin Educ Health Prof. 2013;33(1):59-66.
12. Duthie S, Hodges P, Ramsay I, Reid W. EMQs: a new component of the MRCOG Part 2 exam. The Obstetrician & Gynaecologist. 2006; 8: 181–5.
13. Karila L, François H, Monnet X, Noel N, Roupret M, Gajdos V, et al. The Script Concordance Test: A multimodal teaching tool. Rev Med Interne. 2018;39(7):566-73.
14. Gagnon R, Charlin B, Lambert C, Carrière B, Van der Vleuten C. Script concordance testing: more cases or more questions? Adv Health Sci Educ Theory Pract. 2009;14:367–75.
15. Charlin B, Desaulniers M, Gagnon R, Blouin D, Van der Vleuten C. Comparison of an aggregate scoring method with a consensus scoring method in a measure of clinical reasoning capacity. Teach Learn Med. 2002;14(3):150-6.
16. Charlin B, Tardif J, Boshuizen HP. Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. Acad Med. 2000;75(2):182-90.
17. Miller G. The assessment of clinical skills/competence /performance. Acad Med.1990;65(9):63–7.
18. Vorstenbosch MA, Bouter ST, Van den Hurk MM, Kooloos JG, Bolhuis SM, Laan RF. Exploring the validity of assessment in anatomy: do images influence cognitive processes used in answering extended matching questions? Anat Sci Educ. 2014;7(2):107-16.
19. Charlin B, Van der Vleuten C. Standardised assessment of reasoning in contexts of uncertainty: the script concordance approach. Eval Health Prof. 2004;27 (3):304–19.
20. Tormey W. Education, learning and assessment:

current trends and best practice for medical educators. Ir J Med Sci. 2015;184(1):1-12.

21. Lubarsky S, Charlin B, Cook DA, Chalk C, Van der Vleuten CP. Script concordance testing: a review of published validity evidence. Med Educ. 2011;45(4):329-38.

22. Gagnon R, Charlin B, Coletti M, Sauvé E, Van der Vleuten C. Assessment in the context of uncertainty: how many members are needed on the panel of reference of a script concordance test? Med Educ. 2005; 39:284-91.

23. 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. 2002; 24:522–7.

24. Subra J, Chicoulaa B, Stillmunkès A, Mesthé P, Oustric S, Rougé Bugat ME. Reliability and validity of the script concordance test for postgraduate students of general practice. Eur J Gen Pract. 2017;23(1):208-13.

25. Gagnon R, Charlin B, Lambert C, Carriere B, Van der Vleuten C. Script concordance testing: more cases or more questions? Adv Health Sci Educ Theory Pract. 2009;14(3):367–75.

26. Dory V, Gagnon R, Vanpee D, Charlin B. How to construct andimplement script concordance tests: insights from a systematic review. Med Educ. 2012;46(6):552-63.

27. 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.