Original Article

A comprehensive test of clinical reasoning for medical students: An olympiad experience in Iran

Alireza Monajemi, Kamran Soltani Arabshahi[1], Akbar Soltani[2], Farshid Arbabi[3], Roghieh Akbari[4], Eugene Custers[5], Arash Hadadgar[6], Fatemeh Hadizadeh[6], Tahereh Changiz[7], Peyman Adibi[8]

Department of Philosophy of Science, Institute for Humanities and Cultural Studies, [1]Center for Educational Research in Medical Sciences, [2]Endocrinology and Metabolism Research Center (EMRC), Tehran University of Medical Sciences, Tehran, [3]Department of Oncology, Zahedan University of Medical Sciences, Zahedan, [4]Department of Internal Medicine, School of Medicine, Babol University of Medical Sciences, Babol, Iran, [5]Medical Education Development Center, University of Utrecht Medical Center, Utrecht, Netherlands, [6]Medical Education Development Center, [7]Medical Education Research Center, [8]Department of Internal Medicine, School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran.

ABSTRACT

Background: Although some tests for clinical reasoning assessment are now available, the theories of medical expertise have not played a major role in this filed. In this paper, illness script theory was chose as a theoretical framework and contemporary clinical reasoning tests were put together based on this theoretical model. **Materials and Methods:** This paper is a qualitative study performed with an action research approach. This style of research is performed in a context where authorities focus on promoting their organizations' performance and is carried out in the form of teamwork called participatory research. **Results:** Results are presented in four parts as basic concepts, clinical reasoning assessment, test framework, and scoring. **Conclusion:** we concluded that no single test could thoroughly assess clinical reasoning competency, and therefore a battery of clinical reasoning tests is needed. This battery should cover all three parts of clinical reasoning process: script activation, selection and verification. In addition, not only both analytical and non-analytical reasoning, but also both diagnostic and management reasoning should evenly take into consideration in this battery. This paper explains the process of designing and implementing the battery of clinical reasoning in the Olympiad for medical sciences students through an action research.

Key words: Clinical reasoning, medical expertise, clinical reasoning, assessment, olympiad, battery

INTRODUCTION

Besides formal curriculum in medical schools, other educational interventions are required to develop meta-competencies relevant to doctors' professional life. Since, in the contemporary formal exams mostly assess students' knowledge base, but rarely their higher levels of thinking such as problem solving and reasoning, one of these interventions could be planning a competition in which reasoning and problem-solving come into focus. *Medical Students Olympiad* is designed to highlight the importance of reasoning and problem solving in medicine which have been ignored in formal education in respect of education and assessment.[1]

Critical reasoning is a major competency which leads a physician takes steps wisely and purposefully in both diagnosing and managing patients. Clinical reasoning consists of all stages of patient workup: From the history taking to completion of treatment and follow-up. Therefore, it is no

Address for correspondence: Alireza Monajemi,
Assistant Professor, Institute for Humanities and Cultural Studies,
Tehran, Iran. E-mail: monajemi@med.mui.ac.ir

exaggeration to say that clinical reasoning is the practice of medicine per se.

The main questions that were addressed in paper were: What kind of test is appropriate to assess clinical reasoning in undergraduates medical students? Are the contemporary tests suitable for this purpose? Should other tests be designed? This paper shows how these questions were answered through an action research to serve a theoretical basis for this type of clinical reasoning assessment.

MATERIALS AND METHODS

This paper is a qualitative study performed with an action research approach.[2-4] This style of research is performed in a context where authorities focus on promoting their organizations' performance and is carried out in the form of teamwork called participatory research.[2-4] It is based on cooperation and mostly deals with problems challenging organizations and simultaneously focuses on the problems and their solutions.

First of all, literature and databases were extensively reviewed and then analyzed through content analysis.[3] At the next stage, test structure was determined after holding several sessions (about 100 two hours sessions) with experts. Then this structure was reevaluated by some medical education experts, faculty members of clinical departments and senior educational administrations and the final form was achieved after applying changes and amendments. It should be mentioned that two experts in qualitative research analyzed related texts and group discussions. Sessions were held in research division of Isfahan University of Medical Sciences and Chancellery of education of Ministry of Health and Medical Education.

RESULTS

Results are presented in four parts as basic concepts, clinical reasoning assessment, test framework, and scoring as follows:

Basic concepts

A theoretical framework was based on illness script theory because it serves a good basis for novice-expert differentiation. In every patient encounter, doctors perceive features–symptoms, signs and background information of the patient. Therefore, a relevant illness scripts activate which help doctors to rule hypotheses in or out in the diagnostic process, whereas others are used for patient management. In this essence, first clinical reasoning can be described as illness script activation, illness script selection, and illness script verification.[5-10]

Second, all the information about diseases that doctors has is organized in a structure called the *illness script*. It is an integrated knowledge structure consisting of at least four parts: *faults*, *consequences*, *enabling conditions*, and *management*.[10] Faults are pathophysiological malfunctions that constitute the biomedical core of the disease and are usually subsumed under a diagnostic label. Consequences are about the clinical manifestations of a disease such as complaints, signs, and

symptoms. Enabling conditions are the patient's background information (e.g., age, sex, medical history, drug history, family history of diseases, occupation, and living environment) that generally makes the occurrence of a certain disease more or less likely.[11,12]

Most likely, only one illness script pops up. On the basis of the activated illness script, a doctor evaluates the patient's data and reconfirms the diagnosis and finally manages the patient. This is called non-analytical reasoning. If more than one illness scripts are activated for a single patient, or if the patient' data do not fully fit any particular illness script, analytical reasoning process activates.[9,11,12]

Third is the relationship between diagnosis and management in clinical reasoning. The diagnostic (Dx) education should precede management (Mx) education, simply because in order to learn how to manage an illness, one should first know about its clinical presentation and underlying mechanism. However, it seems that these two types of knowledge are crucial for effective patient workup and these both should take into consideration in clinical reasoning assessment in medical students.[10,13-15]

In summary, three major criteria were defined for a clinical reasoning assessment. It should discriminate novice medical students from expert doctors in terms of illness script components, analytical versus non-analytical reasoning, and Dx versus Mx reasoning.

Clinical reasoning assessment

Clinical reasoning assessment is not similar to other common assessments in medical education.[16-18] Clinical reasoning tests are among those tests called alternative assessment. Alternative assessments assess students' knowledge or skills by making assessment condition close to real situations.[16,17] The use of real-life patient scenarios and open book exams would help to reach such condition. To make exam condition more close to the environment in which a physician practices, by using books, handheld computers and consulting with colleagues approximate the exam condition to real situations. Another characteristic of these tests is an emphasis on doing or making decisions in clinical condition instead of merely asking about information and knowledge.[16,17] If problems are designed based on real condition, they certainly cannot have only one correct answer unlike other common tests (multiple-choice tests) which require only one answer for each item. It therefore becomes clear that real condition problem solving provides possible correct answers; this issue is recognized in alternative tests. One of the most important notions of clinical reasoning tests is the flexibility of answers. Questions must be designed in a way that allows flexible answering. This is in contrast with other common tests which dichotomous require absolute and certain answers and do not allow answers other than those specified (i.e., dichotomous approach). In clinical reasoning assessment correct answer could not only indicated by referring to a sentence of a word in a medical textbook. It is clear that medical knowledge is necessary for solving medical problem, but the way this knowledge is used is totally different from when the test is intended to assess a sort

of rote memorization. In other words, clinical reasoning tests should be designed in a way so that texts are necessary but not sufficient; because if they are necessary and sufficient conditions, it means that the question is of the exact text about knowledge and information and not reasoning.

Nowadays there is a trend in clinical reasoning assessment towards assessing through multidimensional measurement." One instrument for one trait" approach has been replaced by "multi-instrument for multiple roles" approach or multiple biopsy and the idea that there should be a test for each field or area is not defensible anymore.[19] Clinical reasoning assessment is not possible with one test and a theoretical framework is required which encompasses multidimensional measurement. Clinical reasoning tests must have two characteristics of process-orientation and expert-novice discrimination.[5-7] Currently clinical reasoning tests are all based on defaults of a test to measure a characteristic; while the test intended to measure clinical reasoning must be multidimensional. Based on the illness script theory, a test battery will be designed in which three skills of data collection, generating hypothesis and its assessment must be measured respectively for the assessment of clinical reasoning skill to get a complete picture of one's clinical reasoning.

Therefore, clinical reasoning test takes the form of a battery of tests and current tests should be used to design such a test battery. In cases there is no suitable test to assess one of the triple skills, new tests should be designed.

Test framework
First contemporary clinical reasoning tests in literature were examined. Based on our theoretical framework, Key Features (KF),[20] Clinical Reasoning Problem (CRP),[21] Script Concordance[22] and Comprehensive Integrative Puzzle (CIP)[23] tests were found appropriate. KF is suitable for examining illness script activation and the accuracy of data collection assessment, whereas, CRP is more likely appropriate test for assessment of illness script selection as well as analytical reasoning. CIP is best fit with non-analytical reasoning, while, SC is suitable for evaluating illness script verification.

Since some aspects such as reliable and valid data collection based on different sources of patient's information in illness script activation were not assessed, another test, called Information Gathering Test (IGT), was designed for data collection skill based on Illness Script Theory (Appendix 1). Another test, called Scenario Formation (SF), was designed to examine the accuracy of components of illness scripts and its connection to diagnoses (Appendix 2).

Scoring
If actions are taken based on presented values and concepts, the rating of such tests will also be different from that of objective tests, because the search is for all possible answers and all the answers which are in the range of correct answers will be rated. Therefore a group of specialists called expert panel hereafter, are responsible for providing answer keys to tests. Expert panel consists of 15 specialists related to the fields of designed ques-

tions and each one answers questions individually, then their views will be gathered according to existing standards. Based on designing standards of each test and expert panels' views, tests were assigned two rating methods namely dichotomous rating and partial weighting of each item, and the score of each question was the sum of items' scores. The total score of each test was the sum of its questions' scores.[22,23]

FINAL DESIGN

At the first stage, to familiarize universities, each medical school was required to introduce two faculty members in order to participate in a workshop for designing questions in clinical reasoning domain. In this workshop, faculty members got familiar with clinical reasoning concepts and assessment. Several standard styles of designing clinical reasoning test (such as KF, CRP and SC) were presented in the workshop and all participants were asked to design questions based on this framework and send them to the Olympiad Secretariat, so that designers whose questions were standard and close to the framework were selected as team members for designing the Olympiad questions. Questions delivered to the Secretariat were meticulously analyzed. Some universities did not observe the main structure and provided multiple choice questions. Those faculty members who complied with the standards and their questions were of good quality were invited to the final team of designers.

After formatting the final team, several sessions were held about test overview, questions budgeting, time limits, the order of tests and the number of questions of each test; and finally the first framework of clinical reasoning test was changed as follows. After final modifications of tests' framework and structure, the test battery of clinical reasoning was finalized in the form shown in Table 1.

Designing questions
After finalizing tests' framework, the most important issue was to provide specifications for selecting the domains of problems. As the Olympiad participants were medical students internal medicine was chose and consequently seven internal medicine subspecialists were selected as a defensible basis for distributing problems in primary care setting.

The proportion of patients assigned to each of these seven sections was determined by the distribution in primary care set-

Table 1: Tests classified according to the number of questions and the time devoted to each test			
Devoted time	No. of question	Time table	Name of the test
60 minutes	42	1st test-1st day morning	SC
60 minutes	40	2nd test-1st day morning	IGT
90 minutes	8	1st day afternoon-3rd test	SF
60 minutes	50	2nd day morning-4th test	KF
60 minutes	40	2nd day morning-5th test	CRP
90 minutes	60	2nd day afternoon-6th test	Puzzle

ting. Then signs, symptoms and diseases were selected in each section based their importance and prevalence in this field. Similarly and given the budgeting, questions were designed in two sessions at the center of medical education assessment, Ministry of Health and Medical Education. Test scheduling, number of questions and the style of writing were in accordance with scientific references of the technical committee.[20-23]

Scoring system

A scoring key consists of a list of correct answers and a system assigning to these keyed responses. Providing scoring keys required the expert panel, therefore, the expert panel made up of faculty members of medical universities including internal medicine specialists and medical education experts. Each of the members individually answered the questions, and then their answers were collapsed to build the keys. All test scores were entered into Excel software for weighting and calculation. Given the different nature and characteristics of these tests, scoring of each test was explained as follows:

1. KF test: In this test, the score allocated to each item was equal to the weight given by the expert panel, for example if 12 of 15 members voted for item A, its weight would be 12/15. The sum of selected items' scores constituted the score of each question and the final score of KF test was the sum of questions' scores. If a student chose more than 5 items in each question, a negative weight was assigned.[20]

2. IGT test: This included choice of items and short answer tests. The "choice of items" was similar to KF test, and for short answer tests, the expert panel selected a set of answers as correct responses, and each answer would receive a score if written by student. The total scores of items constituted the score of each question, and the sum of questions scores determined the final test score.[20-23]

3. CRP test: Items were weight equally in CRP test and the expert panel selected a set of answers with each answer equal to one score. Wrong items did not receive any score. Scores allocated to correct diagnosis and related findings were considered equal. For example if 1.2 is allocated to the parts 1 and 2 of the CPR test (Appendix 2), 0.2 is for correct diagnosis and 1 (5 to 0.2) is for correct findings. In the case that more than five findings were selected, one of five findings was eliminated for each extra finding. If the diagnosis was wrong, selected findings did not receive any score; if the diagnosis was correct but wrong findings were selected, the student would only receive diagnosis score (0.2). The score of each question was the sum of correct items and the final score was the sum of questions' scores.[21]

4. SC test: In this test, the weight of each item was based on the weight assigned by the expert panel. For example if 14 members chose item -1 and one member chose item 0 (zero), the weight of item -1 was 14/15 and 1/15 for the item 0. The score of each case was the total score of its 3 related questions and the final score consisted of the total case scores. In a case that there was a controversy between an answer provided by one member and the rest of the expert panel, for example when all members chose -1 and -2 but one member chose +2, he or she was asked to explain his or her reasons. In most cases, the reason for this controversy was a misunderstanding. If so, one more vote was considered for the item -2 and the score of the item +2 became zero; but if he or she defended his or her choice with reasonable arguments, +2 was considered correct. Therefore each item chosen by a member of the expert panel is not necessarily the correct answer.[22]

5. Puzzle test: Here, answers were not weighted and a combination of items in four parts of patient's history, physical examination, Para clinic and clinical reasoning was considered as the correct answer which was allocated full score. Where two or three pieces were matched, part of the full score (four pieces matched) was allocated. For example, by referring to two correct pieces and three correct pieces, 0.3 and 0.6 of the full score were allocated provided that one of the pieces would be patent's history.[23]

6. Scenario Formation test: In this test, two members reviewed and rated each scenario according to the standard checklist and then discussed on disputable points to reach an agreement. The use of principle of parsimony, balanced use of clinical symptoms and enabling conditions and the appropriateness of diagnosis and management with the written scenario were rated.[10-12]

LESSONS LEARNT: FEEDBACK AND ASSESSMENT

Clinical reasoning test battery which was designed and performed by faculty members of medical universities is a suggestion for a comprehensive tool to assess clinical reasoning among medical students. This test is intended to offer an opportunity to focus on clinical reasoning education and assessment. Since action research is performed while acting and no action can be taken perfectly, some problems were noticed in designing and performing the Olympiad which demand more attention and research. The summary of these problems is as follows:

1. Unfamiliarity with tests: Most objections to this test were due to its differences with other common tests and unfamiliarity with its concepts. It seems that there is a need for more training in order to familiarize students and faculty members with such tests. Their view about these tests and the belief that they are biasedly rated is one the main reasons for their objections; so there is a need for more transparency about our scoring systems. Those universities with members in the expert panel were less opposing because of their familiarity with rating procedures.

2. Modifications in some tests: The rating of short answer part of IGT test was so difficult and time consuming which also needs modification. This part can be changed into multiple choice questions so that students can choose a number of possible answers.

   Both designing and scoring of SC were complicated and should be modified for a more feasible format. The major problem with SC was that in many situations the link between the scenario and the following question was so loose that for answering the question there is no need to read the scenario.

Some experts believed that the scoring of Scenario Formation test was not reliabale among the expert panel and the key needed more revision and transparency.

3. *Implication for medical education and research:* This battery should take into account with two purposes in mind. First, the battery should be assessed in terms of reliability and feasibility that needs more research. Second is to address these issues: its educational implication, how it could be introduced into formal exams and how both faculty staff and medical students become familiar with these new types of exam.

ACKNOWLEDGEMENT

Appendix 1
A sample of IGT

In the following case, which part of the patient's information need verification? Please write down the number of the item and explain the way of verification.

A 65 years old comatose man was brought to ER. He is retired and has lived alone. His son found him unconscious on the floor (1) a couple of hours ago. His son explained that he has had high blood pressure (2) that has been on treatment (3). He also used an eye drug (4), but he does not know its name. In physical exam, a comatose man with cold extremities is observed. His vital signs are as below:

T = 36.7 C (5) - PR = 110/min- RR = 22 /min- BP = 13/8 cm Hg (6)

Miotic pupils (7) and the sign of the head trauma (8) are also detected.

Appendix 2
A sample of SF

Please write down two separate scenarios with the signs and symptoms in the below bow in a way that each scenario cover all of them. Each scenario must consist of up to 200 words. Please put the diagnosis of each scenario on the top of it. High blood pressure – Increased vocal fremitus – Increased tactile fremitus – Dyspnea - Cough – Fever.

REFERENCES

1. Adibi P, Hadagar A, Hadizadeh F, Haghjoo SH, Monajemi A. Medical sciecne olympiad: Concepts, discieplines and methods. Isfahan: Isfahan University of Medical Sciences Publication; 1998.

2. Hatch JA. Doing Qualitative Research in Educational Settings. Albany, NY: State University of New York; 2002.

3. Insch GS, Moore JE, Murphy LD. Content analysis in leadership research: Examples, procedures, and suggestions for future use. Leadersh Q 1997;8:1-25.

4. Denzin NK, Lincoln YS. Handbook of Qualitative Research. 1st. Thousand Oaks, CA: Sage Publications; 1994.

5. Higgs J, Jones MA, Loftus S, Chiristensen N. Clinical reasoning in the health professions. Second ed.

6. Gruppen LD, Frohna AZ. Clinical reasoning. In: Norman GR, Ven Der Vleuten CP, Newble DI, editors. International handbook of research in medical education. Great Britain: Kluwer Academic Publishers; 2002.

7. Charlin B, Boshuizen, HP, Custers EJ, Feltovich PJ. Scripts and clinical reasoning. Med Educ 2007;41:1178-84.

8. Schmidt HG, Norman GR, Boshuizen HP. A cognitive perspective on medical expertise: Theory implications. Acad Med 1990;65:611-21.

9. Norman G. Research in clinical reasoning: Past history and current trends. Med Educ 2005;39:418-27.

10. Monajemi A, Rikers RM. The role of patient management knowledge in medical expertise development: Extending the contemporary theory. IJPCE 2011;1:109-14.

11. Custers EJ, Boshuizen HP, Schmidt HG. The influence of medical expertise, case typicality and illness script component on case processing and disease probability estimates. Mem Cognit 1996;24:384-99.

12. Custers EJ, Boshuizen HP, Schmidt HG. The role of illness scripts in the development of medical diagnostic expertise: Results from an interview study. Cognit Instruct 1998;16:367-98.

13. Monajemi A, Rikers RM, Schmidt HG. Clinical case processing: A diagnostic versus a management focus. Med Educ 2007;41:1166-72.

14. Monajemi A. Clinical Reasoning: Concepts, education and assessment. Isfahan: Isfahan University of Medical Sciences Publication; 2011. problems. Acad Med 1997;72:173-9.

15. Coderre S, Mandin H, Harasym PH, Fick GH. Diagnostic reasoning strategies and diagnostic success. Med Educ 2003;37:695-703.

16. Van der Vleuten C, Newble D. How can we test clinical reasoning? Lancet 1995;345:1032-4.

17. Newble D, Norman G, van der Vleuten C. Assessing clinical reasoning. In: Higgs J, Jones M, editors. Clinical reasoning in the health professions. 2nd ed. Oxford: Butterworth-Heinemann; 2000.

18. Brener E, Hamilton LA, Best WR. A new approach to evaluating problem solving in medical students. J Med Educ1974;49:666-72.

19. Schuwirth L. Is assessment of clinical reasoning still the Holy Grail? Med Educ 2009;43:298-300.

20. Page G, Bordage G, Allen T. Developing key-feature problems and examinations to assess clinical decisionmaking skills. Acad Med 1995;70:194-201.

21. Grooves M, Scott I, Alexander H. Assessing clinical reasoning: A method to monitor its development in a PBL curriculum. Med Teach 2002;24:507-15.

22. 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;12:189-95.

23. Ber R. The CIP (comprehensive integrative puzzle) assessment method. Med Teach 2003;25:171-6.