Trends in Medicine

Research Article ISSN: 1594-2848

Key feature, clinical reasoning problem. Puzzle and scenario writing: Are there any differences between them in evaluating clinical reasoning?

Elham Boushehri[1], Alireza Monajemi Kamran[2*] and Soltani Arabshahi[3]

[1]Faculty of Health, Hormozgan University of Medical Sciences, Bandar Abbas, Iran

[2]Department of Philosophy of Science, Institute for Humanities and Cultural Studies, Tehran, Iran

[3]Center for Educational Research in Medical Sciences (CERMS), Iran University of Medical Science, Tehran, IranMinina Str., 23, Russia

Abstract

Introduction: At the present time, four tests including key features, clinical reasoning, puzzle, and scenario writing are among the most well-known tests frequently used in training and assessing medical students. Medical Students who are spending their internship are the main targets of the clinical reasoning tests. Research has shown that, in terms of expertise level, the mental scripts of this group of students are categorized as intermediate (compared to two levels of students and residents). That is, different forms of tasks could activate different aspects of their mental scripts. The purpose of this study is to investigate the script-based reasoning of the intermediate group when facing different clinical reasoning tests and compare it with the reasoning of the expert and novice groups.

Method: The way of reasoning in 45 students in different levels of expertise (15 residents as experts, 15 interns as intermediate individuals, and 15 students as novices), when facing clinical reasoning tests, was explored using the "Think aloud method".

Results: The results of exploring the script-based clinical reasoning of the intermediate, expert, and novice groups when facing clinical problems showed that the script-based clinical reasoning of the intermediate group is sensitive to the type of the test.

Discussion: Currently, many efforts have been made to design different forms of clinical reasoning tests. This result of this study could introduce a new approach in designing clinical reasoning tests based on which the examinee's characteristics are taken into account by those who are interested in this area.

Introduction

The first studies on specifying the nature of clinical reasoning began with the question of how expert physicians, compared with the novices, reason. In other words, the question “how different expert physicians are from novice physicians” has always been related to the issue of clinical reasoning [1-7].

Along with the efforts taken by researchers in explaining the clinical reasoning by experts and novices, attention was drawn to a new concept of reasoning and decision making which was previously introduced by the researchers in psychology. This concept, known as “script”, explains how the knowledge structure is stored in mind [8,9]. Based on the theory of script, clinical reasoning means the ability of a physician to recover the mental scripts related to the clinical condition he/she faces with. The formation of mental scripts begins immediately after the first clinical confrontations of a medical student. Therefore, the qualitative and quantitative promotion has a direct linear relationship with expertise [8,10] so that the variety and content quality of the scripts in experienced physicians are more than the novices [10-16].

It was also found that recovery of a script content of the intermediate group was obviously affected by the type of task it faced so that every task led to activation of the type of script which was different from the activated script in terms of content [17,18]. This phenomenon known as scattered knowledge [16] means that all the illness-related knowledge that are necessary to diagnose and manage the illness exist in the mind of an individual from the intermediate group, but since they are saved apart from each other and not related to each other, when confronting with the illness, they are not recovered simultaneously, because different parts of the existing clinical knowledge in the minds of individuals in the intermediate group are not integrated and are stored in their minds as separate pieces and unrelated to each other [18]. In other words, their scripts are not tuned to confront the real clinical conditions. In contrast, when an expert confronts with a related clinical condition, all the required pieces of information related to each illness are recovered simultaneously and this is what is expected from the concept of a rich script. In other words, the formation of a clinical script results from the relationship between the necessary pieces of information about a clinical condition so that when necessary, all this information is activated and recovered simultaneously [8]. Therefore, the recovery and activation of the expert’s script is not affected by the type of the task he/she is dealing with.

Correspondence to: Alireza Monajemi Kamran, Assistant Professor, Department of Philosophy of Science, Institute for Humanities and Cultural Studies, Tehran, Iran, E-mail: monajemi@med.mui.ac.ir

Key words: *clinical reasoning tests, key feature test, clinical reasoning problem test, scenario writing test, puzzle test, intermediate effect*

Received: March 19, 2019; **Accepted:** May 06, 2019; **Published:** May 09, 2019

Evaluation of clinical reasoning

Along with the explanation of the nature of the clinical reasoning and the expertise process, the ways to evaluate the ability of the clinical reasoning was also considered. This attention is due to the fact that the clinical performance of physicians as the consequence of their clinical reasoning method and their role to increase or decrease the diagnostic errors were considered [1,19-22]. The ways to evaluate this skill were invented [1,19]. These studies led to designing some clinical reasoning tests some of the most frequently used of which include: Key Features Test [19,23-28]. Clinical Reasoning Problem (CRP) test [28-31], Comprehensive Integrative Puzzle (CIP) test and [30,32] Scenario test [28,30]. These tests are used on large scale with the aim of promoting the capability of the students' clinical reasoning in at university and at national level. The ability of the tests in creating distinction between different levels of expertise has frequently been proved in different studies [19,28,30,31,33,34].

The point which is still unanswered in designing this type of tests was that how much attention is paid to the characteristics of the audience group of the clinical reasoning tests, who are usually medical students in intermediate levels? As mentioned in the introduction, the result of the studies investigating the intermediate group's clinical scripts aimed to explain the expertization process and showed that the clinical knowledge structure stored in the mind of this group of students is scattered [17,18]. This feature is responsible for the condition that the script of the students at this level of expertise is represented in a different way in terms of the type of the question or the task they face with [17,30]. Since the clinical reasoning tests have different formats and types, it is possible that, like different tasks, they recover different parts of the group's clinical reasoning's script. Therefore, it seems that the clinical reasoning tests do not have a similar role in representing the mental scripts of the intermediate group.

One way to test this claim is to explain the way the intermediate group performs the script-based clinical reasoning when facing different reasoning tests and compare the explained scripts patterns resulted from different tests. Such a study must be planned in a way that two groups from higher and lower levels, Saied the experts and the novices, participate in the study at the same time. The reason the two groups of novices and experts are mandated to be present along with the intermediate group is simultaneous comparison of the way each group performs clinical reasoning. According to the mentioned studies, it is predicted that changing the test will not have any effects on the way of clinical reasoning and consequently on the quality of the mental scripts of the two groups of novices and experts. Regarding the explanations about the intermediate group's characteristics, it is predicted that the different clinical reasoning tests used in this study have a different impact on the representation of the mental scripts of the intermediate group, and proving this phenomenon can signify the necessity of using more than one test in the evaluation of clinical reasoning of this group of physicians

In the current study, the researchers intend to explain the pattern of the clinical reasoning of the intermediate, novice, and expert groups in confronting with different formats of the clinical reasoning tests, compare the result with the pattern of the expected clinical reasoning in these three groups, and finally explain the role of changing the test format on the quality of the clinical script of different groups of expertise.

Research population

In Iran, the General Medical Professional Doctorate Program has a 7-year training program divided into two phases, preclinical and clinical. The preclinical phase and the clinical phase last 4 and 3 years, respectively. The graduation condition for the students is to pass the two phases mentioned above and present their theses at the end of the clinical phase [35]. There are also academic programs for general physicians who like to continue their education in a special medicine course. The general physician who studies internal medicine residency program will be considered as an internist [36].

The participants in this study included the first-year students of clinical phase as the novice group, the students of the last year of clinical phase as the intermediate group, and the students of the third or the fourth year of residency of internal diseases of Medical University of Hormozgan as the expert group [8,14-18]. The reason the internal diseases expertise was chosen is that the educational area of the internal diseases is wider than that of the other expert courses and the clinical issues are similar to the issues taught and practiced in levels of general physician program [8,14,15,17,18,36-44]. The sampling was performed on a voluntary basis [45] from among the students which had no problems regarding their academic achievement [37,38,46,47]. In all three groups of expert, intermediate, and novice, 15 individuals were interviewed.

The source of the clinical reasoning tests used in the study

The questions used in the study were chosen among the standard questions of key features, script concordance test, clinical reasoning problem, Comprehensive Integrative Puzzle (CIP), and scenario writing [48]. The overall reliability index of the questions was 91.0. The validity of this class of questions has been investigated based on the agreement of an elite's panel consisting of 15 different expertise from 15 medical education universities in Iran. In order to select the questions among the existing questions in this source, the item difficulty level index was used. Since the most appropriate difficulty index for a question is recommended to be 5.0 [40,49,50], therefore, among the existing tests in the mentioned test bank, two KF questions, two CRP questions, two SCT questions, one Scenario question and one Puzzle question with the difficulty level between 5.0 and 56.0 were selected and used.

Process

In this study, the 'thinking aloud' method was used in order to collect and analyze the information related to participants' mental script [46]. Think aloud method is one of the most effective methods to evaluate and explain the mental processes in high levels of cognitive areas such as problem solving and reasoning [46,51]. Doing these processes requires involving the active memory whose content can be verbalized [46]. The verbal product of this part of memory which is produced as the interviewee is solving the problem is called 'protocol'. By analyzing the protocol, the thinking process during problem solving could be explored [46,51-54]. Besides explaining the thinking process during problem solving and reasoning, it could be also possible to compare the thinking process of different expertise groups during the reasoning about a common issue [15,46,47,51,55]. Therefore, every participant in the three levels of students (novice), intern (intermediate), and resident (expert) was given a set of clinical reasoning questions including scenario writing, Puzzle, KFT CRP, and SCT and asked to read every question aloud and say his/her thinking while solving the test. Before beginning, the process of think aloud was practiced will

every one of the participants with a non-clinical test. All the protocols were recorded, transcribed and typed for content analyzing.

Content analysis

In the next stage, it was decided that the protocols would be compared with a psychological model. The base pattern used in this comparison, the Illness Script Processing, was adapted from the script theory [7,8,33,34,56]. The model of script processing shows that the selection of every script by a physician is done in a three-stage process [57]: Activation [33,34]; Verification; [57]. Confirmation; [57]. So, in order for the protocol to be comparable with the base model, a coding scheme was prepared (Appendix 1).

In the next stage, the amount to which the protocol matches the pattern was quantified (Appendix 2). Two main goals the researchers had in mind by quantification were:

- Making objective the amount of adaptation of the protocol with the base pattern;
- Comparability of the protocols resulted from different groups with each other;

After quantifying the amount, the protocol matches the base model, using statistical comparative methods to know the impact of "expertise level" and " the type of clinical reasoning test (task) on the way of participants' reasoning (166, 168 and 169) [49,58,59].

Results

In the present study, 45 medical students in different levels of expertise (15 residents considered as experts, 15 interns as intermediate level, and 15 trainees as novices) were faced with clinical reasoning tests and their mental process was analyzed.

Table 1 shows the script process average score resulted from different clinical reasoning tests taken by the intern (intermediate) group. The lowest and highest scores belong to the scenario test and puzzle test, respectively.

Table 1 also shows the average script process score of the novice and expert groups when faced with the clinical reasoning tests are not significantly different (P-value>0.05). However, it was shown that the average script process score of the intermediate and expert group when faced with the clinical reasoning tests are significantly different (P-value= 0.03) and the paired comparison of the script process score of the intermediate group, when faced with the clinical reasoning tests, showed that the script process average scores of KF, CRP, puzzle and scenario range from 2.93(0.25) to 5.5(0.57) and are not significantly different from each other (Table 2).

Till now, results showed that just three tests, CRP, KF and puzzle induce higher quality of illness script in comparison to scenario test. In continuation to the analysis plan, it was decided to determine which of these tests are more appropriate for the group. So, the test score achieved by intermediate group in each three test was compared to that of the other two groups to know which of them is/are able to distinguish intermediate participants from the other groups.

As observed in Table 3, the average scores of KF test are close in all three groups. The results of the Post hoc stage after the variance analysis showed that the average scores of the two groups novice and intermediate are equal (p=0.12) and lower than that of the expert group (P= 0.012) (Table 4).

Table 5 shows the CRP test average score in novice, intermediate, and expert groups. The results of the Post hoc phase after the variance analysis showed that the average scores of these three groups are significantly different from each other (Table 6).

Table 7 shows the average score of different expertise groups resulted from the puzzle test. The results of the Post hoc phase after the variance analysis showed that the average scores of these three groups are significantly different from each other (Table 8).

Table 1. The mean and standard deviation of script process score of novice, intermediate and expert groups when faced with the clinical reasoning tests

Clinical reasoning tests	Mean (SD) in Novice	Mean (SD) in Intermediate	Mean (SD) in Expert	Total Mean (SD)
Scenario	0.56 (0.9)	2.93(0.25)	6.9 (0.75)	2.93(0.75)
Puzzle	1.01(0.89)	5.5(0.87)	7.4(2.3)	5.54(0.25)
CRP	1.03(0.3)	6(0.71)	8(0.22)	6(0.87)
KF	1.41(0.17)	5.5(0.57)	8(2.2)	5.52(0.71)
P-value	0.9	0.03	-	0.09

Table 2. Compression of means of script process score resulted from clinical reasoning tests in intermediate group

Clinical reasoning tests	KF	CRP	Puzzle	Scenario
Scenario	0.001	0.001	0.001	-
Puzzle	0.45	0.65	-	0.001
CRP	0.1	-	0.65	0.001
KF	-	0.1	0.45	0.001

Table 3. Mean and standard deviation of KF score in expert, intermediate and novice groups

Groups	Mean	SD
Expert	98.43	26.95
Intermediate	86.8	19.35
Novice	52.66	10.69

Table 4. Compression of means of score resulted from KF in expert, intermediate and novice groups

Groups	Expert	Intermediate	Novice
Expert	0.012	0.08	-
Intermediate	0.12	-	0.08
Novice	-	0.12	0.012

Table 5. Mean and standard deviation of CRP score in expert, intermediate and novice groups

Groups	Mean	SD
Expert	63	14
Intermediate	28	11.49
Novice	19	11.38

Table 6. Compression of means of score resulted from CRP in expert, intermediate and novice groups

Groups	Expert	Intermediate	Novice
Expert	-	0.019	0.04
Intermediate	0.019	-	0.017
Novice	0.04	0.017	-

Table 7. Mean and standard deviation of Puzzle score in expert, intermediate and novice groups

Groups	Mean	SD
Expert	9.87	1.29
Intermediate	4.47	1.21
Novice	2.09	0.64

Table 8. Compression of means of score resulted from Puzzle in expert, intermediate and novice groups

Groups	Expert	Intermediate	Novice
Expert	-	0.011	0.02
Intermediate	0.011	-	0.001
Novice	0.02	0.001	-

Discussion

First of all, the results of the study are summarized as follows:

1- The script process average scores of the expert group in all of the four tests do not show any significant difference.

2- The script process average scores of the intermediate group in all of the four tests do not show any significant difference.

3- The script process scores of the intermediate group in the CRP, KF and puzzle tests are not significantly different, but they are significantly higher than the script process average score of the scenario test.

4- From the four clinical reasoning tests of the study, the only test that was not able to distinguish two groups of participants (intermediate group and novice group) from each other was the KF test.

The results are discussed more as follows.

The expert group (residents of internal medicine)

As the results related to this group of physicians faced with different clinical reasoning tests shows, they have a similar performance when facing with different clinical reasoning tests.

The script process average scores of the expert group did not prove any significant difference when they were faced with different formats of the clinical reasoning tests. In other words, Experts show the script-based reasoning feature in a higher quality compared to other levels of expertise. It was observed in the findings that the expert's script process average score ranges from 9.6 to 8 when faced with different clinical reasoning tests. ¬ Since the minimum and maximum scores defined fir the script extracted from the participants' thinking are respectively 0 and 9, it can be easily realized that the script score quality resulted from the expert's thinking, when faced with different clinical reasoning tests, it was close to 9. In a study carried out on 235 novices and 90 experts aimed at investigating the effects of concurrent "complexity of the method of presenting the clinical case", "reward", and "clinical experience" on the accuracy of doctors' diagnosis, it was concluded that the expert physicians' performance is only affected by clinical experience in medicine, and such factors as rewards and complexity of the method of presenting the clinical case (task transformation, as exists in the present study) can not affect the experts' performance [60-62]. On the other hand, the expert's diagnosis accuracy and speed do not change even when the task format changes [14].

In present study it was also proved that the variety of clinical reasoning tests design does not prove any impact on the performance of the expert in terms of the quality of the introduced script. This finding shows that the expert has a feature called "performance consistency". This feature has been referred to as "Robustness" in some studies related to expertise [60]. Performance robustness means that expertise is not simply affected by the conditions. On the other hand, not only does the consistency feature indicate that the expert's performance is not affected by the conditions he/she is faced in decision-making situations, but also it shows the high quality of the expert's performance in different situation compared to other levels of expertise [1,7,8,13,17,63].

Novice group (clerkships)

The script scores of the novice group are equal in all of the four tests. As the results related to this group faced with different clinical reasoning tests shows, they have a similar performance when faced with different clinical reasoning tests.

As mentioned in the results section, the novices' script process average score ranges from 56.0 to 4.2 when faced with different clinical reasoning tests. This range, compared to the variation range of the script process average score being 9.6 to 8, is significantly different. This study showed that the novice's information lacks a sufficient amount of clinical knowledge. Therefore, when the novices are confronted with the clinical reasoning conditions, they cannot relate the existing indications in the aforesaid situations to the little amount of clinical information in their script. In other words, considering the scripts which are clinically weak, the process of the novice's mental scripts will not be of sufficient quality. This study showed that increasing experience through increasing the clinical content of mental scripts can enhance the process of novice groups' scripts [13,16,64]. Besides, the average scores and components of the script process of the novice group did not prove any significant difference when they were faced with different formats of the clinical reasoning tests. In other words, different designs or tasks do not affect the novice's performance in terms of script process quality presented. This finding in apparently similar to the result achieved for the expert group. The performance of the expert group was not affected by changing the test type. It must be noted that the quality of the expert group's reasoning is totally different from that of the novice group's reasoning and they are indeed in contrary of each other.

Intermediate group (Interns)

After exploring script-based clinical reasoning pattern of the intermediate group when faced with the clinical reasoning tests, it was observed that the script process scores of the intermediate group in the three CRPT, KFT, and puzzle tests are equal to each other but higher than the score resulted from the scenario test. This result is of importance in two parts:

Part A: In the present study, it is concluded so far that the test type does not impact the script process score of expert and novice groups. So, expert and novice individuals act the same when faced with different tests, except that the script process score are high in the expert group and low in the novice group. In contrary, it was observed that the script process score of the intermediate group, when faced with the four CRP, KF, scenario, and puzzle tests, is between 93.2 and 6 based on the results of Table 2 which is, however, less than that of the expert group and more than that of the novice group when faced with the same three tests. In a study where the script quality of three novice, intermediate and expert groups was investigated through counting the essential clinical content in each group, it was found that the intermediate group members selected from among the second-year trainee students with an average of 5.5 years of clinical experience had a quality between the two poles of expert and novice groups [17]. Another significant result of the study was proving the direct positive relationship between the scripts of two novice and intermediate groups and the number of cases they had in real setting with the same conditions. This finding reaffirms the significant role that experience plays in forming and enhancing mental scripts and in developing novices to the same status where the intermediate and expert groups are. On the other hand, this proves the finding of the present study where the script process average score of the intermediate group is higher than the novice group and lower than the expert group [5,13].

Part B: Another result achieved in in the present study is the multiple performance of the intermediate group when faced with different types of test [5,17,18]. That is, statistically speaking, there was no significant difference between the intermediate group's script process average scores when faced with the three CRP, KF, and puzzle tests. In other words, the script process quality of the intermediate group when faced with these three tests was the same and the script process average score of this group when faced with these three tests was in between the script process average scores of expert and novice groups when faced with the three tests. On the other hand, from the cognitive point of view, these three test types are considered to be task. Moreover, it was observed that the intermediate group's script process average score when faced with the scenario test is significantly different from the script process average score when faced with the three tests and is lower. This is the fact while as shown, neither the expert group nor the novice group had such a dual performance when faced with the tests, and functioned the same when faced with the tests.

Conclusion

Since the intermediate groups are mainly considered as the primary targets of evaluation in clinical reasoning tests, any results which indicates this group is affected by the test type is of great importance. Three CRP, KF, and puzzle tests are at the same level in terms of script process score induced in this group. In other words, the script process qualities of the intermediate group resulted from each CRP, KF, and puzzle tests are similar to each other. In principal, in cognitive point of view, they are considered a similar task. On the contrary, the script process quality resulted from the scenario test in this group is significantly less than the other tests. It must be noted that besides the low ability of the scenario test in inducing the intermediate group's mental script, the script process average score of the intermediate group achieved in this test is again significantly higher compared to that of the novice group.

In this section, two important questions are posed:

A) The first question is that besides the poor ability of the scenario test in inducing the mental script of the intermediate group, whether this test could be recommended as an appropriate test to evaluate the intermediate group?

As seen in the results, the script process average score for the intermediate group when faced with the scenario test tend to the lowest number in the scoring scale. Maybe if the script average score of this test were investigated alone and not at the same time with the other three tests, this conclusion that the script average score resulted from this test is increasing from the novice group to the expert group would confirm the appropriateness of this test for the intermediate group and would introduce it as a good test in terms of having distinguishing features among the scripts of three novice, intermediate, and expert groups. However, one feature of this study is to concurrently investigate the reasoning method of the three groups when faced with the four clinical reasoning tests, and the results have shown that the script process quality induced by the three CRP, KF, and puzzle tests into this group is significantly higher than that of the script induced by the scenario. Considering the duality of the intermediate group behavior against the different formats of clinical questions, it can be concluded that the specific format of the scenario test cannot activate and process the intermediate group script into the quality expected. This format is not appropriate for the intermediate groups in which the disease-related information is scattered. So, it is not able to activate all parts of the mental scripts of the intermediate group. However this does not mean the inappropriateness of this test for this group because the script process score resulted from this test is between the two novice and expert levels. Besides, it seems that this test can be used depending the main goal of clinical education. For example, if the purpose of clinical reasoning evaluation is to determine the student's position in the evolving process of clinical scripts, it is better, while understanding the script non-uniformity theory at this stage of education, to use tests which are able to induce the students' mental script such as puzzle and CRP. But if the purpose of evaluation is to facilitate and accelerate the students' mental script at this level of education, it is better to use the scenario test. The results of the study may be considered as to confirm the necessity of using more than one test in assessing the clinical reasoning of intermediate group of physicians as it were previously suggested by the other researchers [65,66].

B) The second question is that whether one test, among the tests at the same level in the intermediate group i.e. CRP, KF, and puzzle tests, could be selected as the most appropriate test for this group?

It would be expected that the average scores in CRP, KF, and puzzle tests in the intermediate group and the novice group is significantly different. It means that a test could differentiate a level of expertise from its lower level. So, if it would be found that a test has no the ability, it should not be considered as an appropriate test because it cannot distinguish this group from other lower ones.

The results of this study could be also academically useful, especially for clinical teachers. Clinical teachers are well aware of the fact that designing the clinical reasoning tests is a process requiring a lot of time and accuracy. A clinical professor who intends to design the questions of a clinical reasoning test in any format needs to receive related and complicated training which still does not completely guarantee his success in designing this type of tests. On the other hand, the number of studies which frequently publish in order to verify different kinds of the existing tests or to design new types is increasing [19,23-32,67]. It seems that studies similar to the present study can limit the number of choices a professor faces as the clinical evaluator and therefore, help him overcome the confusion in choosing the best format and understand what role each test type has in representing the clinical knowledge of this group of test takers. Also, if the studies prove that all the existing tests are similar in terms of the type of reasoning they induce, the doubts on possible superiority of one test over another in the evaluators' mind will be eliminated. It could also introduce a new approach in designing clinical reasoning tests based on which the examinee's characteristics are taken into account by those who are interested in this area.

Suggestions

The results of student evaluation, as the final stage of the planning process of teaching, are a reflection of education quality of different aspects of the curriculum. On the other hand, the results of evaluations are considered a point of departure for making change and enhancing the quality of different components of curriculum. Using the clinical reasoning tests in the curriculum of students in the clinical level indicates the fact that the students' clinical reasoning is one of the important goals of their clinical education. Therefore, based on the role the evaluation of clinical reasoning tests in activating different parts of the clinical knowledge of the intermediate groups (interns), suggestions can be made as to choose an appropriate educational approach in this group of students at this clinical level. The educational suggestions made based on the results achieved in this study must indeed include strategies to increase the consistency of the clinical knowledge of the intermediate group. Different studies have suggested that increasing

the number of clinical confrontations in the real educational settings is the key to develop from the novice level to the expert level [1,17,68]. The consistency of the student's mental script will only be achieved by his/her frequent clinical experience. In this regard, some studies have suggested the students' early exposure to the clinical context as an accelerating way of reaching expertise. Based on the present study, it seems that the facilitating and accelerating the way of reaching expertise in medical students is not possible through providing early exposure only. However, confronting the appropriate clinical reasoning tests is also a good way to activate all necessary component of a related mental script and develop its consistency.

References

1. Norman G (2005) Research in clinical reasoning: past history and current trends. *Med Educ* 39: 418-427. [Crossref]
2. Elstein AS, Shulman LS, Sprafka SA (1978) Medical problem solving: An analysis of clinical reasoning. Cambridge, MA: Harvard university press.
3. Neufeld VR, Norman GR, Barrows HS, Feightner JW (1981) Clinical problem-solving by medical students: a longitudinal and cross-sectional analysis. *Med Educ* 15: 315-322.
4. Norman GR, Brooks LR, Allen SW (1989) Recall by expert medical practitioners and novices as a record of processing attention. *J Exp Psychol Learn Mem Cogn* 15: 1166-1174. [Crossref]
5. Verkoeijen PP, Rikers RM, Schmidt HG, van de Wiel MW, Kooman JP (2004) Case representation by medical experts, intermediates and novices for laboratory data presented with or without a clinical context. *Med Educ* 38: 617-627. [Crossref]
6. Patel VL, Groen GJ (1986) Knowledge based solution strategies in medical reasoning. *Cogn Sci* 10: 91-115.
7. Charlin B, Tardif J, Boshuizen HPA (2000) Scripts and medical diagnostic knowledge: Theory and applications for clinical reasoning instruction and research. *Acad Med* 75: 182-190.
8. Schmidt HG, Norman GR, Boshuizen HP (1990) A cognitive perspective on medical expertise: theory and implication. *Acad Med* 65: 611-621. [Crossref]
9. Coderre S, Wright B, McLaughlin K (2010) To think is good: Querying an initial hypothesis reduces diagnostic error in medical students. *Acad Med* 85.
10. Ericsson KA, Charness N (1994) Expert performance: Its structure and acquisition. *Am Psychol* 49: 725-747.
11. Hobus PPM, Schmidt HG, Boshuizen HPA, Patel VL (1987) Contextual factors in the activation of first diagnostic hypotheses: Expert-novice differences. *Med Edu* 21: 471-476.
12. Schmidt HG, Boshuizen HP (1993) On the origin of intermediate effects in clinical case recall. *Mem Cognit* 21: 338-351. [Crossref]
13. Rikers RM, Schmidt HG, Boshuizen HP (2000) Knowledge encapsulation and the intermediate effect. *Contemp Educ Psychol* 25: 150-166. [Crossref]
14. Verkoeijen PP, Rikers RM, Schmidt HG, van de Wiel MW, Kooman JP (2004) Case representation by medical experts, intermediates and novices for laboratory data presented with or without a clinical context. *Med Educ* 38: 617-627. [Crossref]
15. Hobus PPM, Boshuizen HPA, Schmidt HG (1989) Mental representation of prototypicalpatients: Expert-novice differences. Paper presented at the first european congress of psychology, Amsterdam, and The Netherlands.
16. Boshuizen H, Schmidt H (1992) On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cog Sci* 16: 153-184.
17. Custers E, Boshuizen, Henny PA, Schmidt H (1998) The role of illness scripts in the development of medical diagnostic expertise: Results from an interview study. *Cogn Instr* 16: 367-439.
18. Monajemi A, Rostami ER, Savaj S, Rikers RM (2012) How does patient management knowledge integrate into an illness script? *Educ Health (Abingdon)* 25: 153-159. [Crossref]
19. van der Vleuten CP, Newble DI (1995) How can we test clinical reasoning? *Lancet* 345: 1032-1034. [Crossref]
20. Weingart SN, Wilson RM, Gibberd RW, Harrison B (2000) Epidemiology of medical error. *BMJ* 320: 774-777. [Crossref]
21. Zhang J, Patel VL, Johnson TR, Shortliffe EH (2004) A cognitive taxonomy of medical errors. *J Biomed Inform* 37: 193-204. [Crossref]
22. Croskerry P (2009) Clinical cognition and diagnostic error: applications of a dual process model of reasoning. *Adv Health Sci Educ Theory Pract* 14 Suppl 1: 27-35. [Crossref]
23. Page G, Bordage G, Allen T (1995) Developing key-feature problems and examinations to assess clinical decision- making skills. *Acad Med* 70: 194-201.
24. Page G, Bordage B (1995) The medical council of Canada's key features project: a more valid written examination of clinical decision -making skills. *Acad Med* 70: 104-110.
25. Page G, Bordage B, Allen T (1995) Developing key-feature problems and examinations to assess clinical decision making skills. *Acad Med* 70: 194-201.
26. Monajemi A (2006) Paradigm shift in clinical reasoning assessment. 3rd Asia pacific medical education conference, Singapore.
27. Norman G (2000) The epistemology of clinical reasoning: Perspectives from philosophy, psychology, and neuroscience. *Acad Med* 75: 127-133.
28. Adibi P (2008) The national scientific Olympiad for medical students. Esfahan university of medical science publisher.
29. Santrock JW (2001) Educational psychology. Mc Graw-hill higher education companies.
30. Monajemi A (2010) Clinical reasoning: concepts, education and assessment. Esfahan university of medical science publication.
31. Groves M, Scott I, Alexander H (2002) Assessing clinical reasoning: a method to monitor its development in a PBL curriculum. *Med Teach* 24: 507-515. [Crossref]
32. Monajemi A (2005) CRP test: an instrument for clinical reasoning evaluation in medical students, The 7th national congress of medical education. Tabriz, Iran.
33. Schmidt HG, Boshuizen HPA (1993) On acquiring expertise in medicine. *Educ Psychol Rev* 5: 3.
34. Amini M, Moghadami M, Kojuri J, Abbasi H, Abadi AA et al. (2011) An innovative method to assess clinical reasoning skills: Clinical reasoning tests in the second national medical science Olympiad in Iran. *BMC Research Notes*. [Crossref]
35. Medical education curriculums for general practitioner program, Iranian ministry of health and medical education website. http://hcmep.behdasht.gov.ir/uploads/Ayeen_Dromomi89.pdf
36. Medical education curriculums for residency programs, Iranian ministry of health and medical education website. http://cgme.behdasht.gov.ir/index.aspx?siteid=264&pageid=22758
37. Firouznia S, Yousefi A, Ghassemi G (2009) The Relationship between academic motivation and academic achievement in medical students of Isfahan university of medical sciences. *Iranian Journal of Medical Education* 9: 79-84.
38. Baghani M (2011) The effect of educational motivation on self-efficacy and study approaches in students, The first national congress on cognitive science in education, Iran.
39. Johnstone CJ, Bottsford-Miller NA, Thompson SJ (2006) Using the think aloud method (cognitive labs) to evaulate test design for students with disabilities and English language learners (Technical Report 44). Minneapolis, MN: University of Minnesota, National center on educational outcomes. http://education.umn.edu/NCEO/OnlinePubs/Tech44/
40. Charters E (2003) An introduction to think-aloud methods. *Brock Education* 12.
41. Cullum L (1998) Encouraging the reluctant reader: Using a think-aloud protocol to discover strategies for reading success. Report for department of English, Indiana university of Pennsylvania (ERIC Document Reproduction Service).
42. Nielson J (1994) Estimating the number of subjects needed for a thinking aloud test. *Int J Human-Computer Studies* 41: 385-397.
43. Gall MD, Borg WR, Gall JP (1996) Educational research: An introduction. White Plains, NY: Longman Publishers, USA.
44. Denzin N, Lincoln Y (1998) Collecting and interpreting qualitative materials. *Sage Publications*.
45. Coyne IT (1997) Sampling in qualitative research. Purposeful and theoretical sampling; merging or clear boundaries? *J Adv Nurs* 26: 623-630. [Crossref]
46. Ericsson KA, Simon HA (1993) Protocol analysis: Verbal reports as data (revised edition). Cambridge, MA: MIT Press.

47. Van Someren M, Barnard Y. Sandberg J (1994) The thinks aloud method. A practical guide to modelling cognitive processes. Academic Press, London.

48. The second Iranian national medical Olympiad tests (2010). http://medolympiad.behdasht.gov.ir/index.aspx?fkeyid=&siteid=280&pageid=51454

49. Akyel A, Kamisli S (1996) Composing in first and second languages: Possible effects of EFL writing instruction. Paper presented at the Balkan conference on English language teaching of the international

50. association of teachers of English as a foreign language, Istanbul, Turkey. ERIC document reproduction service.

51. Johnson KE (1992) Cognitive strategies and second language writers: A re-evaluation of sentence combining. *J Second Lang Writ* 1: 61-75.

52. Atkinson RC, Shiffrin RM (1971) The control of short-term memory. *Sci Am* 225: 82-90. [Crossref]

53. Patel VL, Arocha JF, Zhang J (2004) Thinking and reasoning in medicine. In: Keith Holyoak: Cambridge handbook of thinking and reasoning. Cambridge, UK: Cambridge university press.

54. Patel V, Kaufman D. Arochab JF (2002) Emerging paradigms of cognition in medical decision-making. *J Biomed Inform* 35: 52-75.

55. Baumann A, Bourbonnais F (1982) Nursing decision making in critical care areas. *J Adv Nurs* 7: 435-446. [Crossref]

56. Kieras DE, Just MA (1987) New methods in reading comprehension research. Hillsdale, NJ: Erlbaum, pp: 253-286.

57. Coderre S, Mandin H, Harasym PH, Fick GH (2003) Diagnostic reasoning strategies and diagnostic success. *Med Educ* 37: 695-703. [Crossref]

58. Charlin, B, Tardif J, Boshuizen HPA (2000) Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. *Acad Med* 75: 182-190.

59. Seif A (2011) Measurement, assessment and evaluation in Education. Dowran publisher, Iran.

60. Daniel W (2008) Biostatistics: A foundation for analysis in the health sciences. 9th Ed. Wiley, John & Sons.

61. Rikers RM, Schmidt HG, Boshuizen HP, Linssen GC, Wesseling G, et al. (2002) The robustness of medical expertise: clinical case processing by medical experts and subexperts. *Am J Psychol* 115: 609-629. [Crossref]

62. Rikers R, Winkel WT, Loyens S, Schmidt H (2003) Clinical case processing by medical experts and subexperts. *J Psychol* 137: 213-223. [Crossref]

63. Ferguson E, Starmer C (2013) Incentives, expertise, and medical decisions: testing the robustness of natural frequency framing. *Health Psychol* 32: 967-977.

64. Brazeau-Lamontagne L, Charlin B, Gagnon R, Samson L, van der Vleuten C (2004) Measurement of perception and interpretation skills during radiology training: utility of the script concordance approach. *Med Teach* 26: 326-332.

65. Patel VL, Groen GJ, Scott HM (1988) Biomedical knowledge in explanations of clinical problems by medical students. *Med Educ* 22: 398-406. [Crossref]

66. Custers E, Boshuizen HPA, Schmidt H (1996) The influence of medical expertise, case typicality, and illness script component on case processing and disease probability estimates. *Mem Cognit* 24: 384-399.

67. Nendaz MR, Tekian A (1999) Assessment in problem-based learning medical schools: A literature review. *Teach Learn Med* 11: 232-243.

68. Ber R (2003) The CIP (comprehensive integrative puzzle) assessment method. *Med Teach* 25: 171-176.

69. Schmidt HG, Rikers RM (2007) How expertise develops in medicine: Knowledge encapsulation and illness script formation. *Med Educ* 41: 1133-1139.