Journal of Advances in Medical Education & Professionalism

Combination of different clinical reasoning tests in a national exam

ANAHITA SADEGHI[1], MD; ALI ALI ASGARI[1], MD; NEZARALI MOULAEI[2], MD; VAHID MOHAMMADKARIMI[3], MD; SOMAYEH DELAVARI[4], PhD; MITRA AMINI[5*], MD; SETAREH NASIRI[6], MD; ROGHAYEH AKBARI[7], MD; MOJGAN SANJARI[8], MD; IRAJ SEDIGHI[9], MD; PARISA KHOSHNEVISASL[10], MD; MANOUCHEHR KHOSHBATEN[11], MD; SAEED SAFARI[12], MD; LEILY MOHAJERZADEH[13], MD; PARISA NABEIEI[5], MS; BERNARD CHARLIN[14], MD

[1]Digestive Disease Research Institute, Tehran University of medical Sciences, Tehran, Iran; [2] Infectious Diseases and Tropical Medicine Research Center, Resistant Tuberculosis, Institute, Zahedan University of Medical Sciences, Zahedan, Iran; [3]Department of Internal Medicine, School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran; [4]Center for Educational Research in Medical Sciences (CERMS), Department of Medical Education, School of Medicine, Iran University of Medical Sciences, Tehran, Iran; [5]Clinical Education Research Center, Shiraz University of Medical Sciences, Shiraz, Iran; [6] Department of Gynecology Oncology, Iran University of Medical Sciences, Tehran, Iran; [7]Department of Internal Medicine, Babol University of Medical Sciences, Babol, Iran; [8]Endocrinology and Metabolism Research Center, Institute of Basic and Clinical Physiology Sciences, Kerman University of Medical Sciences, Kerman, Iran; [9]Department of Pediatric, School of Medicine, Hamedan University of Medical Sciences, Hamedan, Iran; [10]Zanjan Social determinants of health research center, Zanjan University of Medical Sciences, Zanjan, Iran; [11]Medical Education Research Center, Health Management and Safety Promotion Research Institute, Tabriz University of Medical Sciences, Tabriz, Iran; [12]Department of Surgery, Firoozgar general hospital, Iran University of Medical Sciences, Tehran, Iran; [13]Pediatric Surgery Research Center (PSRC), Research Institute for Children Health (RICH), Mofid Children's Hospital (MCH), Shahid Beheshti University of Medical Sciences (SBMU), Tehran, Iran; [14]Medical School, University of Montreal, Montreal, Canada

▶ **Abstract**

Introduction: Clinical reasoning as a critical and high level of clinical competency should be acquired during medical education, and medical educators should attempt to assess this ability in medical students. Nowadays, there are several ways to evaluate medical students' clinical reasoning ability in different countries worldwide. There are some well-known clinical reasoning tests such as Key Feature (KF), Clinical Reasoning Problem (CRP), Script Concordance Test (SCT), and Comprehensive Integrative Puzzle (CIP). Each of these tests has its advantages and disadvantages. In this study, we evaluated the reliability of combination of clinical reasoning tests SCT, KF, CIP, and CRP in one national exam and the correlation between the subtest scores of these tests together with the total score of the exam.

Methods: A total of 339 high ranked medical students from 60 medical schools in Iran participated in a national exam named "Medical Olympiad". The ninth Medical Olympiad was held in Shahid Beheshti University of Medical Sciences, Tehran, Iran, under the direct supervision of the Ministry of Health and Medical Education in summer 2017. The expert group designed a combination of four types of clinical reasoning tests to assess both analytical and non-analytical clinical reasoning. Mean scores of SCT, CRP, KF, and CIP were measured using descriptive statistics. Reliability was calculated for each test and the combination of tests using Cronbach's alpha. Spearman's correlation coefficient was used to evaluate the correlation between the score of each subtest and the total score. SPSS version 21 was used for data analysis and the level of significance was considered <0.05.

Results: The reliability of the combination of tests was 0.815. The reliability of KF was 0.81 and 0.76, 0.80, and 0.92 for SCT, CRP, and CIP, respectively. The mean total score was 169.921±41.54 from 240. All correlations between each clinical reasoning test and total score were significant (P<0.001). The highest correlation (0.887) was seen between CIP score and total score.

Conclusion: The study showed that combining different clinical reasoning tests can be a reliable way of measuring this ability.

Keywords: Education, Medical assessment, Medical students

230

J Adv Med Educ Prof. October 2019; Vol 7 No 4

Introduction

Clinical reasoning is the method by which clinicians collect information about the patients' problems and develop a plan to solve the problem and manage the patients (1-3). Clinical reasoning is a critical and high level clinical competency, and it should be acquired during medical education, and medical educators should attempt to teach this ability to medical students (4, 5).

Although special medical care is provided by specialists, most of the general medical care is performed by general practitioners; Then, improving the ability of clinical reasoning and decision-making can play a major role in reduction of the incidence of adverse events in the clinical care and the promotion of health indicators (6). Assessing this ability is also important; different clinical reasoning tests developed for measuring this ability include Key Features (KF), Script Concordance Test (SCT), Comprehensive Integrative Puzzle (CIP), and Clinical Reasoning Problem (CRP) (4).

One of the tests used for assessing clinical reasoning is clinical reasoning problems (CRP). In this test, a scenario is presented, and students should choose the two diagnoses they consider most likely for the scenario; they should also mention the features of the case that are important for the correct diagnosis and indicate whether these features positively or negatively predict each diagnosis (7-10).

KF test was introduced by Bordage and Page (10); its questions focused on critical steps in clinical problem solving and may pertain to aspects that learners generally find difficult or that are necessary for the patients' management. Its focus is on the key features of diseases (4). Several studies investigated the reliability, generalizability, construct , and predictive validity of the KF test (11).

SCT was introduced by Bernard Charlin in 2000 (12, 13). In SCT, a short patient scenario ensues with three questions. Every scenario has three columns, which include diagnostic hypothesis, new clinical finding, and scale from −2 to +2 (4, 12, 13). Scoring in this test is according to an expert panel answering the question (4, 12).

Another clinical reasoning test is the CIP Test that introduced Case and Swanson for assessing clinical reasoning in routine situations (14, 15). This test was developed according to illness script theory (16). CIP is a kind of pattern recognition test such that its format is similar to extended matching test (14, 17).

Each of the clinical reasoning tests measures a different aspect in the clinical reasoning domain. Sometimes, it seems necessary to measure all abilities by using all the above-mentioned tests.

In this study, we evaluated the reliability of combining various clinical reasoning tests (SCT, KF, CIP, and CRP) in one national exam. We also evaluated the correlation between these tests and the total score of the exam.

Methods

A total number of 339 high rank medical students from 60 medical schools participated in a national exam named Medical Olympiad. The ninth Medical Olympiad was held in Tehran, Iran, in Shahid Beheshti University of Medical Sciences under the direct supervision of the Ministry of Health and Medical Education in summer 2017.

An expert group with members from all Iranian medical schools developed a question bank of specialized clinical reasoning test items in Internal medicine, General Surgery, Pediatrics, Obstetrics, and Gynecology. The expert group designed a combination of four types of clinical reasoning tests to assess the clinical reasoning based on the methodology described in former articles (8, 10, 13, 16, 18-22).

Fifteen experts (the expert panel) in the fields of Internal medicine, Surgery, Pediatric, Obstetrics, and Gynecologists were asked to answer the tests without using textbooks or consulting with each other. Then, each answer was weighted according to their scores.

In each KF test, a case was described and subsequently followed by 16 questions. Students would choose four correct answers for each KF test. The answers were weighted by the expert panel's (answers to the same questions). To increase the discriminating power, we calculated the partial credit score (18, 23-27).

To measure clinical data interpretation in ill-defined cases, we presented the clinical situations as vignette cases that did not include all the data necessary to provide a diagnosis. Then, a series of related items with different formats (diagnosis, investigation, or treatment) in three parts were designed. The first part included a clinical scenario or hypothesis, the second part succinctly gave more information (clinical or paraclinical data) that might have positive or negative effects on the first, and the third part was a 5-point Likert scale type. SCT was introduced by Bernard Charlin in 2000 (12, 13). In it, a short patient scenario follows three questions. Every scenario has three columns, which include diagnostic hypothesis, new clinical finding, and scale from −2 to +2 (1, 12, 13). Scoring in this

test was according to the expert panel's answers to the question (1, 12, 15).

The CRP tests were scored based on the binary method in which the correct answer had one score, and the wrong answer had zero score. By using the summative method, the total score (sum of the scores of questions) was calculated (9).

In this test, the answers were not weighted, and a combination of items in four parts (patient's history, physical examination, paramedic(s) and treatment) was considered as the correct answer. CIP scores were calculated from the answers given by the reference panel. For each of the four columns, four correct responses out of 4 (4/4) questions were graded as 100%, 3/4 as 75%, 2/4 as 50%, and 1/4 was graded as 0%, respectively. The grade of the CIP exam was determined by the sum of grades (15).

The total exam score obtained by the summation of four test scores; each test accounted for 25 percent of the total score. The total score was 240, and each test score was 60.

Data analysis was performed using SPSS software version 21. Mean scores of SCT, CRP, KF, and CIP were expressed using descriptive statistics. Reliability was calculated using Cronbach's alpha. Spearman's correlation coefficient was used to evaluate the correlation between the score of each clinical reasoning test and the total score. Student t-test was used to measure the difference between the scores of male and female students. The significance level was considered <0.05.

Results

Reliability was calculated through Cronbach's alpha for the combination of tests, which was 0.815. The reliability of KF was 0.81, and this measure was 0.76 for SCT, 0.80 for CRP, and 0.92 for CIP. 56% of the participants were female (190), while 149 (44 %) people were male. The mean total score was 169.92±41.54 from 240, and the total mean scores of female and male students were 164.87±42.85 and 176.353±9.01, respectively. Table 1 shows clinical reasoning subsets scores according to gender. Male students obtained better total scores than female students, and this difference was statistically significant (P<0.001).

Table 2 shows the correlation between each of the clinical reasoning and total score. According to this Table, all of the correlations were significant (P=0.001). The highest correlation (0.887) was between the CIP score and total score, and the lowest correlation (0.473) was between CRP and SCT.

Discussion

Clinical reasoning is a cognitive process for the diagnosis of the patient problem, and it has an important role in clinical problem solving and patient management (1, 2, 19). Good assessment methods for measuring this ability are important to determine weak points in this field and try to improve this ability (4). There are various methods for assessing this ability, including PT, KF, SCT, and CRP (19, 20). Clinical reasoning assessment methods addressed the underlying construct of the clinical reasoning process and focused on specific sub-tasks, such as data gathering, activating diagnostic hypotheses, and prioritizing diagnostic alternatives (19). This study was designed and implemented to evaluate the reliability of the combination test of clinical reasoning (SCT, KF, PT, and CRP), and the correlation between the score of each test and with the total score.

The reliability of combination of these four tests was excellent and equal to 0.815. Amini et

Table 1: Clinical reasoning total and subset scores according to gender

Sex	Mean±SD Total score (from240)	Mean±SD KF (from 60)	Mean±SD SCT (from 60)	Mean±SD CRP (from 60)	Mean±SD CIP (from 60)	N
Female	164.87±42.85	37.39±7.93	18.77±5.87	50.80±17.65	57.90±17.38	190
Male	176.35±39.01	38.76 7.43	20.89±4.64	55.67±16.65	61.03±16.05	149
Total	169.92±41.54	37.99±7.73	19.70±5.46	52.94±17.36	59.28±16.86	339

Table 2: Correlation of clinical reasoning tests with each other and with the total score

	Variable	KF	SCT	CRP	CIP
Spearman's rho	KF	1	0.545**	0.675**	0.640**
	SCT	0.545**	1	0.473**	0.523**
	CRP	0.675**	0.473**	1	0.647**
	CIP	0.640**	0.523**	0.647**	1
	Total Score	0.803**	0.631**	0.885**	0.887**

**Correlation is significant at the 0.01 level (2-tailed)

232

J Adv Med Educ Prof. October 2019; Vol 7 No 4

al. in their study used four methods of clinical reasoning (KF, SCT, CRP, and CIP) and reported the reliability 0.91 for a combination of the tests (15). Khsohbaten et al. in their study used four methods of clinical reasoning (KF, Scenario, CRP, and puzzle); the combined reliability was equal to 0.86 (21). KF reliability in this article was 0.81; previous studies reported the different range of reliability for the KF: 0.75-0.95 (1, 15, 20, 22-24). A good reliability range for SCT was 0.7 to 0.8 (25-27); the present study showed the reliability of 0.76 for SCT. CRP reliability in this article was 0.81; previous studies reported a different range of reliability for the CRP: 0.72 (9), 0.61 (8), 0.83 (7). However, acceptable reliability for CIP was 0.43–0.73 with a mean of 0.60 (28). The reliability of CIP in the Amini et al.'s study was 0.91 (15); the reliability of CIP in the present study was similar to Amini et al.'s (15) results and was equal to 0.92.

In this study, the highest and lowest mean scores were related to CIP and SCT, respectively. These results are contrary to Amini et al.'s study that the highest and lowest mean scores were related to KF and CRP, respectively (15). In Groves et al.'s study, there was a clear difference between the CRP and SCT means (8). The evaluation of the relationship between gender and the scores of the sub-tests and the total score showed that male participants in all four subsets had a higher score than the female participants; the highest and lowest difference was related to CRP and SCT subset, respectively.

In this study, we investigated each clinical reasoning test and the total score; the correlation between each of the clinical reasoning tests and the total score was positive. SCT examined the degree of coherence between the panel judgment of experts and student responses (1, 29), and it was suitable for assessing the depth and breadth of the students' knowledge. Therefore, it focuses on the structure and organization of the knowledge base (12). The main emphasis was on assessing the students' ability to evaluate the diagnostic assumptions by providing new information (13); although CRP and SCT assess data interpretation, clinical reasoning assessment using CRP test provided a more comprehensive vignette by assessing individual ability in the generation of diagnostic hypothesis and information synthesis, as well as data interpretation (8). On this basis, we expected the correlation between the SCT and CRP to be low; the result of the present study verified this assumption. Therefore, the lowest correlation was related to the correlation between CRP and SCT.

The philosophy of developing the KF test is

solving a clinical problem (15). The results of this study showed that the correlation between each clinical reasoning test and the total score was positive, and it was above 0.473. The highest correlation between the subsets was related to that between CRP and KF.

CIP was developed based on illness script theory (16), and it is a kind of pattern recognition (14, 17). The scoring system for CIP is not dependent on experts' panel judgment and each puzzle has definite answers. Therefore, we expected the high correlation between CIP and other subsets. Results of the present study showed that positive and high correlation existed between CIP and each of the other clinical reasoning tests (CRP, SCT, KF).

We expected a high correlation between the total score and all of the clinical reasoning tests that confirmed our expectation by results. The findings of this study and other studies that used combination methods suggested that using different and complementary methods of assessing clinical reasoning provided a more detailed and qualitative evaluation than either clinical reasoning assessment method alone (8).

The strength of the present study is that the study is a national study that was done by the participation of all medical students from all over the country. The limitation of the present study was the fact that data were gathered from top students that may not be a good representative of all medical students.

Conclusion

The combination of different clinical reasoning tests will be a reliable method for measuring clinical reasoning ability in high stake examinations. The use of combination of these clinical reasoning tests is recommended for high stake examinations in medical schools.

Acknowledgment

Conflict of Interest: None declared.

References

1. Zamani S, Amini M, Masoumi SZ, Delavari S, Namaki MJ, Kojuri J. The comparison of the key feature of clinical reasoning and multiple choice examinations in clinical decision makings ability. Biomedical Research. 2017;28(3):1.
2. Monajemi A, Rikers RM. The role of patient

management in medical expertise development: Extending the contemporary theory. Int J Pers Cent Med. 2011;1(1):161-6.

3. Delavari S, Soltani-Arabshahi K, Monajemi A, Baradaran HR, Yaghmaei M, Myint PK. How to develop clinical reasoning in medical students and interns based on illness script theory: An experimental study. Medical Journal of the Islamic Republic of Iran. 2019; 16(4): 367-98.

4. Ten Cate O, Durning SJ. Approaches to Assessing the Clinical Reasoning of Preclinical Students: Principles and Practice of Case-based Clinical Reasoning Education. Springer. 2018; 15: 65-72.

5. Yousefichaijan P, Jafari F, Kahbazi M, Rafiei M, Pakniyat A. The effect of short-term workshop on improving clinical reasoning skill of medical students. Medical journal of the Islamic Republic of Iran. 2016;30:396.

6. Delavari S, Amini M, Sohrabi Z, Koohestani H, Delavari S, Rezaee R, et al. Development and psychometrics of script concordance test (SCT) in midwifery. Medical journal of the Islamic Republic of Iran. 2018;32:75.

7. Groves M, Scott I, Alexander H. Assessing clinical reasoning: a method to monitor its development in a PBL curriculum. Med Teach. 2002;24(5):507-15.

8. Groves M, Dick ML, McColl G, Bilszta J. Analysing clinical reasoning characteristics using a combined methods approach. BMC medical education. 2013;13(1):144.

9. Derakhshandeh Z, Amini M, Kojouri J, Dehbozorgian M. Psychometric characteristics of clinical reasoning problems (CRPs) and its correlation with routine multiple choice question (MCQ) in cardiology department. JAMP. 2018;6(1):37.

10. Farmer EA, Page G. A practical guide to assessing clinical decision-making skills using the key features approach. Med Educ. 2005;39(12):1188-94.

11. Hrynchak P, Glover Takahashi S, Nayer M. Key-feature questions for assessment of clinical reasoning: a literature review. Med Educ. 2014;48(9):870-83.

12. 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.

13. Charlin B, Roy L, Brailovsky C, Goulet F, Van Der Vleuten C. The Script Concordance test: a tool to assess the reflective clinician. Teaching and learning in medicine. 2000;12(4):189-95.

14. Monajemi A, Yaghmaei M. Puzzle test: A tool for non-analytical clinical reasoning assessment. Medical journal of the Islamic Republic of Iran. 2016;30:438.

15. Amini M, Moghadami M, Kojuri J, Abbasi H, Abadi AAD, Molaee NA, et al. An innovative method to assess clinical reasoning skills: Clinical reasoning tests in the second national medical science Olympiad in Iran. BMC research notes. 2011;4(1):418.

16. Sellar B, Murray CM, Stanley M, Stewart H, Hipp H, Gilbert-Hunt S. Mapping an Australian Occupational Therapy curriculum: Linking intended learning outcomes with entry-level competency standards. Australian occupational therapy journal. 2018;65(1):35-44.

17. Ber R. The CIP (comprehensive integrative puzzle) assessment method. Med Teach. 2003;25(2):171-6.

18. Page G, Bordage G. The Medical Council of Canada's key features project: a more valid written examination of clinical decision-making skills. Acad Med: journal of the Association of American Medical Colleges. 1995;70(2):104-10.

19. Gruppen LD. Clinical Reasoning: Defining It, Teaching It, Assessing It, Studying It. Western Journal of Emergency Medicine. 2017;18(1):4.

20. Huwendiek S, Reichert F, Duncker C, de Leng BA, van der Vleuten CP, Muijtjens AM, et al. Electronic assessment of clinical reasoning in clerkships: A mixed-methods comparison of long-menu key-feature problems with context-rich single best answer questions. Med Teach. 2017;39(5):476-85.

21. Khoshbaten M, Rasi Marzabadi L, Gorbani S, Salek Ranjzadeh F, Hassanzadeh S, Ahmadian A. The Management of Holding and Evaluating Clinical Reasoning Exams Using a Comprehensive System of Electronic Clinical Reasoning Exams (Sajab) in the Sixth Nationwide Medical Sciences Students Olympiad. Res Dev Med Educ. 2015;4(2):159-64.

22. Nikendei C, Mennin S, Weyrich P, Kraus B, Zipfel S, Schrauth M, et al. Effects of a supplementary final year curriculum on students' clinical reasoning skills as assessed by key-feature examination. Med Teach. 2009;31(9):e438-e42.

23. Rademakers J, Ten Cate TJ, Bär P. Progress testing with short answer questions. Med Teach. 2005;27(7):578-82.

24. Trudel JL, Bordage G, Downing SM. Reliability and validity of key feature cases for the self-assessment of colon and rectal surgeons. Ann Surg. 2008;248(2):252-8.

25. Charlin B, Gagnon R, Pelletier J, Coletti M, Abi-Rizk G, Nasr C, et al. Assessment of clinical reasoning in the context of uncertainty: the effect of variability within the reference panel. Med Educ. 2006;40(9):848-54.

26. Goos M, Schubach F, Seifert G, Boeker M. Validation of undergraduate medical student script concordance test (SCT) scores on the clinical assessment of the acute abdomen. BMC Surg. 2016;16(1):57.

27. Duggan P, Charlin B. Summative assessment of 5 th year medical students' clinical reasoning by script concordance test: requirements and challenges. BMC Med Educ. 2012;12(1):29.

28. Capaldi VF, Durning SJ, Pangaro LN, Ber R. The Clinical Integrative Puzzle for Teaching and Assessing Clinical Reasoning: Preliminary Feasibility, Reliability, and Validity Evidence. Military medicine. 2015;180(suppl_4):54-60.

29. Charlin B, van der Vleuten C. Standardized assessment of reasoning in contexts of uncertainty the script concordance approach. Eval Health Prof. 2004;27(3):304-19.