OPEN

Does Objective Structured Clinical Examinations Score Reflect the Clinical Reasoning Ability of Medical Students?

Wan Beom Park, MD, PhD, Seok Hoon Kang, MD, Yoon-Seong Lee, MD, PhD and Sun Jung Myung, MD, PhD

Abstract: *Background:* Clinical reasoning ability is an important factor in a physician's competence and thus should be taught and tested in medical schools. Medical schools generally use objective structured clinical examinations (OSCE) to measure the clinical competency of medical students. However, it is unknown whether OSCE can also evaluate clinical reasoning ability. In this study, the authors investigated whether OSCE scores reflected students' clinical reasoning abilities. *Methods:* Sixty-five fourth-year medical students participated in this study. Medical students completed the OSCE with 4 cases using standardized patients. For assessment of clinical reasoning, students were asked to list differential diagnoses and the findings that were compatible or not compatible with each diagnosis. The OSCE score (score of patient encounter), diagnostic accuracy score, clinical reasoning score, clinical knowledge score and grade point average (GPA) were obtained for each student, and correlation analysis was performed. *Results:* Clinical reasoning score was significantly correlated with diagnostic accuracy and GPA (correlation coefficient = 0.258 and 0.380; $P = 0.038$ and 0.002, respectively) but not with OSCE score or clinical knowledge score (correlation coefficient = 0.137 and 0.242; $P = 0.276$ and 0.052, respectively). Total OSCE score was not significantly correlated with clinical knowledge test score, clinical reasoning score, diagnostic accuracy score or GPA. *Conclusions:* OSCE score from patient encounters did not reflect the clinical reasoning abilities of the medical students in this study. The evaluation of medical students' clinical reasoning abilities through OSCE should be strengthened.

Key Indexing Terms: Objective structured clinical examinations; Clinical reasoning; Medical education. [Am J Med Sci 2015;350(1):64–67.]

The objective structured clinical examination (OSCE) is a useful tool for clinical performance assessment and is used worldwide as part of medical licensing examinations, for example, in the U.S. Medical Licensing Examination Step 2 Clinical Skills section. Mounting evidence validating the OSCE has led to increasingly widespread use of this tool to measure the clinical competency of medical students.[1–3] In 2009, this clinical skills examination was introduced in Korea as an independent examination to be performed as part of the Korean Medical Licensing Examination. During an OSCE, the examinee compiles the history and physical examination information for a standardized patient (SP) and documents the relevant findings, differential diagnosis and plan of action in a structured patient note. The OSCE score usually comprises the following evaluation categories: history taking (30%–40%), physical examination (20%–40%), patient education (0%–10%), physician-patient interaction (20%–40%) and patient note (~5%).

Clinical reasoning ability is regarded as an important factor determining a physician's competency and thus should be taught and tested in medical schools. The OSCE is a useful tool for assessing clinical performance, but it remains unknown whether the OSCE score from a patient encounter reflects a student's clinical reasoning ability. Few studies have investigated the efficacy of an OSCE for evaluating medical students' clinical reasoning ability.[4–6] Therefore, here the authors conducted an analysis to determine whether a high OSCE score from a patient encounter was significantly correlated with good clinical reasoning ability.

METHODS

In March 2011, 65 fourth-year students at Seoul National University College of Medicine (Seoul, South Korea) voluntarily participated in this study. The institutional review board approved this study and waived the requirement for written consent. The OSCE consisted of 4 stations presenting the following clinical cases: IgA nephropathy, neurogenic diabetes insipidus, acute pyelonephritis and exercise-induced asthma. Each station involved a 10-minute student-SP encounter, followed by a 5-minute interstation examination. Performances were evaluated by trained SPs using a checklist. The major evaluation components were overall assessment, history taking, physical examination, physician's manner, patient education and physician-patient interaction. After encountering patients, the students received an answer sheet presenting a table designed to evaluate the students' clinical reasoning process. During the 5-minute interstation examination, students were asked to complete the table with the differential diagnoses and the symptoms or signs that were compatible with or differed from each diagnosis. Then, the patient note was independently rated by 2 physician raters, who assessed the table and counted the number of correct findings within each section. Each correct finding counted as 1 point, and the student's clinical reasoning score was calculated as the total sum of points. Diagnostic accuracy score was calculated as the total number of correct diagnoses among 4 cases.

This patient note form was considered to have content validity because its components were consistent with the literature regarding clinical reasoning skills[7–11] and were chosen based on what physicians write on patient notes in clinical practice. The diagnostic accuracy score was literally a numeric index of diagnostic correctness.

For each student, the authors recorded the OSCE scores from SP encounters, including history taking, physical examination and patient-physician interaction, as well as the diagnostic accuracy scores from the 4 cases. The authors also obtained the students' demographic characteristics, grade point average (GPA) and clinical knowledge test score. The clinical knowledge test investigated clinical knowledge (eg, internal medicine, surgery, psychiatry) using the same format as the Korean Medical Licensing Examination, comprising a 6-section 400-multiple-choice question examination. These collected data were subjected to correlation analyses using Pearson's correlation coefficient. Statistical analysis was performed using SPSS software (version 19.0; SPSS Inc, Chicago, IL). A P value of <0.05 was considered statistically significant.

RESULTS

A total of 65 fourth-year students underwent OSCE examination. Table 1 presents the students' demographic data and examination scores, including GPA and OSCE. Clinical reasoning score was not statistically significantly correlated with OSCE score or clinical knowledge test score (correlation coefficient = 0.137 and 0.091; P = 0.276 and 0.472, respectively) but was significantly correlated with GPA and diagnostic accuracy score (correlation coefficient = 0.380 and 0.258; P = 0.002 and 0.038, respectively) (Table 2). The total OSCE score was not significantly correlated with clinical knowledge test score, clinical reasoning score, diagnostic accuracy score or GPA. Among the components of the OSCE score, neither history taking score nor physical examination score was correlated with clinical reasoning score (correlation coefficient = 0.199 and 0.045; P = 0.112 and 0.722, respectively) (Table 2).

DISCUSSION

Clinical reasoning skills may help students to better focus on the efficient history taking and physical examinations that are required for making a correct diagnosis.[12] Appropriate clinical reasoning is more likely to result in appropriate history taking and physical examination, which comprises a large percentage of OSCE checklists. However, the present results showed that clinical reasoning score was not correlated with OSCE score—meaning that OSCE (especially with a checklist scoring system) could not differentiate students who asked appropriate history questions with appropriate clinical reasoning from others who asked history questions with insufficient clinical reasoning. This further suggests that some students could receive a high OSCE score simply by asking and checking memorized items without adequate reasoning.

Additionally, the authors found that diagnostic accuracy was positively correlated only with clinical reasoning score and not with OSCE score or clinical knowledge score. This suggests that clinical information obtained from the patient may not be the only factor that raises the probability of correct diagnosis and that clinical reasoning may be more vital for correct diagnosis than the amount of clinical information. Some students who gained limited clinical information (low OSCE score) during an encounter still made the right diagnosis by having good clinical reasoning ability. The authors also found that GPA was significantly correlated with clinical knowledge score and clinical reasoning score. GPA is regarded as the global indicator of a student's performance across the spectrum.[13,14] The results showed that GPA was positively correlated with clinical reasoning ability (correlation coefficient = 0.380; P = 0.002).

The clinical knowledge test was originally designed to assess problem solving and clinical decision-making abilities; however, here the authors found that the clinical knowledge test score was not correlated with the diagnostic accuracy score or the clinical reasoning score. In contrast to the SP encounters during OSCE, which required students to actively gather the clinical information required for making an appropriate clinical decision, the clinical knowledge test was a paper examination in which students were passively given all clinical information in a paragraph so that they could interpret laboratory or radiological data and make a diagnostic or therapeutic assumption. The practice situation presented in OSCE is closer to real-world clinical reasoning than the paper examination, as physicians can only get clinical information when they properly ask. Overall, the data underline that the evaluation of clinical competency should include not only clinical knowledge and clinical performance but also clinical reasoning ability, which was not adequately reflected by the results of the presently analyzed methods.

Many efforts have been made to develop a valid and reliable measure of clinical reasoning ability. These have included the use of patient management problems, modified essay questions, script concordance tests and other methods.[15–18] Although clinical reasoning must be assessed in educational programs and certification processes, no study to date has uncovered a single best tool. In this study, the authors tried to assess clinical reasoning ability by simply modifying the existing OSCE examination.

There may be some shortcomings of performing the OSCE examination using SPs as evaluators. It is possible that

TABLE 1. Characteristics and examination scores of 65 medical students

	Mean (SD)	Min	Max
Mean age, yr	26.4 (2.0)	23	35
Sex, M:F	46:19		
Grade point average of the last semester	3.03 (0.59)	2.00	4.14
OSCE score[a]			
Total score[b]	68.0 (5.9)	55.8	82.8
History taking	71.6 (6.8)	58.5	88.7
Physical examination	46.6 (10.3)	31.0	72.1
Patient-physician interaction	63.9 (5.7)	48.6	79.0
Mean diagnostic accuracy score[c]	3.40 (0.66)	1	4
Clinical reasoning score[d]	29.1 (9.0)	12	47
Clinical knowledge test score[e]	285.5 (38.6)	208.0	373.0

[a] A perfect score is 100.
[b] Total score is composed of 40% history taking score, 30% physical examination score and 30% patient-physician interaction score.
[c] A perfect score is 4.0.
[d] A perfect score is not limited.
[e] A perfect score is 400.
OSCE, objective structured clinical examination.

TABLE 2. Pearson correlations between clinical reasoning score, GPA, clinical knowledge test score, diagnostic accuracy score and each component of OSCE score

		GPA	Diagnostic accuracy	Clinical reasoning	OSCE total score	OSCE (G)	OSCE (Hx)	OSCE (PE)	OSCE (C)	OSCE (I)	OSCE (PPI)
Clinical knowledge test	r	0.837[a]	0.185	0.242	0.091	−0.053	0.084	0.171	0.049	0.036	−0.035
	P	0.000	0.141	0.052	0.472	0.675	0.505	0.174	0.698	0.776	0.784
GPA	r		0.231	0.380[a]	0.039	−0.068	0.050	0.141	0.036	−0.037	−0.092
	P		0.067	0.002	0.761	0.596	0.693	0.268	0.781	0.774	0.469
Diagnostic accuracy	r			0.258[a]	−0.030	−0.073	0.027	0.031	0.060	−0.238	−0.077
	P			0.038	0.811	0.566	0.828	0.804	0.636	0.057	0.543
Clinical reasoning	r				0.137	0.180	0.199	0.045	−0.163	0.064	0.134
	P				0.276	0.152	0.112	0.722	0.195	0.610	0.288
OSCE total score	r					0.736[a]	0.893[a]	0.634[a]	0.499[a]	0.626[a]	0.698[a]
	P					0.000	0.000	0.000	0.000	0.000	0.000
OSCE(G)	r						0.561[a]	0.271[a]	0.357[a]	0.478[a]	0.908[a]
	P						0.000	0.029	0.003	0.000	0.000
OSCE(Hx)	r							0.431[a]	0.281[b]	0.458[a]	0.494[a]
	P							0.000	0.024	0.000	0.000
OSCE(PE)	r								0.237	0.295[a]	0.205
	P								0.058	0.017	0.101
OSCE(C)	r									0.240	0.383[a]
	P									0.054	0.002
OSCE(I)	r										0.447[a]
	P										0.000

[a] A P value <0.05 was considered statistically significant.

C, clinical courtesy; G, global rating; GPA, grade point average; Hx, history taking; I, information sharing; PE, physical examination; PPI, physician patient interaction; OSCE, objective structured clinical examination.

OSCE score and clinical reasoning score would be positively correlated if the authors evaluated the students' clinical performances using a global rating based on holistic traits (eg, clinical data interpretation, thought process and logic) instead of a checklist system for each item of history taking or physical examination.[19] However, performance evaluation by trained SPs using a checklist is regarded as a reliable method and has been verified by many previous reports.[20–22] For more precise evaluation of a student's clinical performance, a physician's observation of a student-patient encounter for a sufficient duration would be ideal. However, this could be very difficult to accomplish, especially when evaluating a large volume of students. Thus, many medical schools and license examinations use checklists to evaluate clinical performance.

The interstation examination is already available for evaluating a student's ability to interpret obtained information. However, simply entering an assessment and plan was insufficient for evaluating clinical reasoning ability. The present results indicated that the checklist system for clinical performance in OSCE was very limited in its reflection of clinical reasoning ability. Therefore, the authors created a new patient note form for the interstation examination. Although global rating by experts is regarded as the "gold standard" for clinical reasoning assessment,[23] here the authors used analytic scoring to evaluate clinical reasoning ability. Compared with global rating, analytic scoring is known to be an effective method of giving feedback and to have increased reliability over global ratings.[4] Furthermore, because

the analytic score was rated by physicians, the scoring system to evaluate clinical reasoning might be more reliable than other analytic scoring systems.

This study has several limitations. First, this study was performed in a single institution and only included 65 fourth-grade students. Therefore, these results may not be generalizable to other institutions, which might have different clinical clerkship programs and student evaluation systems. Second, the authors tested just 4 cases, which may not be enough to widely generalize these conclusions. Third, the OSCE score used for the present analysis was only based on the patient encounter and did not include the score for the patient note during the interstation examination. It is possible that a different format or proportion of interstation examination could have changed the influence of clinical reasoning ability on OSCE score. Finally, the study analyzed OSCE examination data early in the fourth year. Students' OSCE scores are not usually consistent throughout the year, and the timing of the OSCE might influence the data.

Although a patient note focused on assessment and planning could be considered to reflect clinical reasoning ability, the data showed that the correlation coefficient between clinical reasoning score and diagnostic accuracy (assessment) was too low to suggest that patient note was an indicator of clinical reasoning ability (r = 0.258, P = 0.038). Furthermore, because only ~5% of total score is usually allotted to patient note in OSCE, the correlation between OSCE and clinical

reasoning score would not be significantly changed if the patient encounter score was substituted with the total composite score including patient note.

For assessing a student's clinical reasoning ability, the authors suggest using the presented table, which includes symptoms or signs that are compatible with or differ from each diagnosis. In addition, a structured short essay or schematic visualization that describes the student's clinical reasoning process could be an effective tool for assessing clinical reasoning ability. Furthermore, research is warranted for the development of valid methods to properly evaluate clinical reasoning ability with an OSCE.

In conclusion, the present results suggest that the OSCE score from a patient encounter may not reflect a medical student's clinical reasoning ability. Efforts should be made to improve the evaluation of the clinical reasoning abilities of medical students using OSCE.

REFERENCES

1. **Cuschieri A, Gleeson FA, Harden RM, et al.** A new approach to a final examination in surgery. Use of the objective structured clinical examination. Ann R Coll Surg Engl 1979;61:400–5.

2. **Harden RM, Gleeson FA.** Assessment of clinical competence using an objective structured clinical examination (OSCE). Med Educ 1979;13:41–54.

3. **Harden RM, Stevenson M, Downie WW, et al.** Assessment of clinical competence using objective structured examination. Br Med J 1975;1:447–51.

4. **Berger AJ, Gillespie CC, Tewksbury LR, et al.** Assessment of medical student clinical reasoning by "lay" versus physician raters: interrater reliability using a scoring guide in a multidisciplinary objective structured clinical examination. Am J Surg 2012;203:81–6.

5. **Durning SJ, Artino A, Boulet J, et al.** The feasibility, reliability, and validity of a post-encounter form for evaluating clinical reasoning. Med Teach 2012;34:30–7.

6. **Kreiter CD, Bergus G.** The validity of performance-based measures of clinical reasoning and alternative approaches. Med Educ 2009;43:320–5.

7. **Kassirer JP, Kuipers BJ, Gorry GA.** Toward a theory of clinical expertise. Am J Med 1982;73:251–9.

8. **Neil C, Michael T, Ralf K, et al.** The role of deliberate practice in chess expertise. Appl Cogn Psy 2005;19:151–65.

9. **Chi MTH, Glaser R, Farr MJ.** The Nature of Expertise. New Jersey: Erlbaum; 1988.

10. **Chi MTH, Glaser R, Rees E.** Expertise in problem solving. In: Sternberg RJ, ed. Advances in the Psychology of Human Intelligence. New Jersey: Lawrence Erlbaum Associates; 1982:7–76.

11. **Ericsson KA.** The influence of experience and deliberate practice on the development of superior expert performance. In: Ericsson KA, Charness N, Feltovich PJ, Hoffman RR, editors. The Cambridge Handbook of Expertise and Expert Performance. New York: Cambridge University Press; 2006:683–703.

12. **Higgs J, Jones MA, Loftus S, et al.** Clinical reasoning in the health professions. 3rd ed. Edinburgh: Elsevier (Butterworth Heinemann); 2008:3–14.

13. **Taylor ML, Blue AV, Mainous AG, et al.** The relationship between the National Board of Medical Examiners' prototype of the Step 2 clinical skills exam and interns' performance. Acad Med 2005;80:496–501.

14. **Andriole DA, Jeffe DB, Hageman HL, et al.** What predicts USMLE Step 3 performance? Acad Med 2005;80:S21–4.

15. **Higgs J, Jones MA, Loftus S, et al.** Clinical reasoning in the health professions. 3rd ed. Edinburgh: Elsevier (Butterworth Heinemann); 2008:379–88.

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

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

18. **Humbert AJ, Besinger B, Miech EJ.** Assessing clinical reasoning skills in scenarios of uncertainty: convergent validity for a Script Concordance Test in an emergency medicine clerkship and residency. Acad Emerg Med 2011;18:627–34.

19. **Regehr G, MacRae H, Reznick RK, et al.** Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. Acad Med 1998;73:993–7.

20. **Elliot DL, Hickam DH.** Evaluation of physical examination skills. Reliability of faculty observers and patient instructors. JAMA 1987;258:3405–8.

21. **Vu NV, Marcy MM, Colliver JA, et al.** Standardized (simulated) patients' accuracy in recording clinical performance check-list items. Med Educ 1992;26:99–104.

22. **Woodward CA, McConvey GA, Neufeld V, et al.** Measurement of physician performance by standardized patients. Refining techniques for undetected entry in physicians' offices. Med Care 1985;23:1019–27.

23. **Ben-David MF, Boulet JR, Burdick WP, et al.** Issues of validity and reliability concerning who scores the post-encounter patient-progress note. Acad Med 1997;72:S79–81.