ORIGINAL ARTICLE

Assessment of Clinical Competence of Medical Students Using the Objective Structured Clinical Examination: First 2 Years' Experience in Taipei Veterans General Hospital

Chin-Chou Huang[1,2,6,7], Cho-Yu Chan[5,9,10]*, Chun-Lien Wu[1], Ya-Lin Chen[1], Hui-Wen Yang[1], Chia-Chang Huang[1], Chen-Huan Chen[1,2,5,6], William J. Huang[1,4,5], Fa-Yauh Lee[1,3], Shing-Jong Lin[1,2,6,8], Jaw-Wen Chen[1,2,6,7]*

[1]Department of Medical Research and Education, Divisions of [2]Cardiology and [3]General Medicine, Department of Medicine, and [4]Division of Urology, Department of Surgery, Taipei Veterans General Hospital; [5]Faculty of Medicine, [6]Cardiovascular Research Center, [7]Institute of Pharmacology, and [8]Institute of Clinical Medicine, National Yang-Ming University, Taipei; Departments of [9]Education and [10]Medical Quality Care, China Medical University Hospital, Taichung, Taiwan, R.O.C.

Background: Competence-oriented education is currently the mainstream method of teaching clinical medical education. The objective structured clinical examination (OSCE) is a widely employed and accepted tool to measure the clinical competence of medical students. We describe the first 2 years' experience of OSCE in Taipei Veterans General Hospital.

Methods: At Taipei Veterans General Hospital, every 7th-year medical student has taken the OSCE since 2006. There were 15 stations in the first 2 years' OSCEs. In years 1 and 2, 133 and 132 students were assessed by the OSCE, respectively. The content of the OSCE included internal medicine, surgery, pediatrics, obstetrics and gynecology, communication, and emergency training. All categories and results of examinees' evaluation at each station were recorded inclusively and compared statistically.

Results: The average scores of students from the 15 stations ranged from 47.7 ± 16.4 to 93.7 ± 8.5 in 2007. The score for communication skills was the lowest, whereas the score for Micro-Sim was the highest. Communication skills and electrocardiography interpretation were the 2 categories in which most of the students failed. A reliability analysis was conducted of the 2007 OSCE questions. The overall score and reliability (Cronbach's reliability) was 0.641. The difference between the impacts on reliability after deleting a test item ranged from 0.59 to 0.65 for all stations. This meant that every station had a similar impact on reliability after being deleted. The squared multiple correlation, R^2, of the reliability of each item was between 0.12 and 0.49, with chest X-ray interpretation being the lowest. The item–total correlation was between 0.10 and 0.41, with interactive case being the lowest.

Conclusion: The OSCE is an effective method for assessing the clinical competence of medical students. The OSCE could be improved further by modifying the examination questions and promoting effective training for standardized patients and examiners. [*J Chin Med Assoc* 2010;73(11):589–595]

Key Words: clinical competence, evaluation, objective structured clinical examination, standardized patient

Introduction

The objective structured clinical examination (OSCE), which was developed in 1975,[1] is an assessment instrument that evaluates medical students' performance based on clinical knowledge, skills and attitudes. The objective of the OSCE is to improve the evaluation of clinical education on competence via real-case scenarios that are acted out by standardized patients. Compared with the traditional written test, the OSCE can

589

better examine students' performance levels of clinical competence by combining various aspects of clinical knowledge and competencies into a single examination. It has become an important method that has been adopted by domestic and international medical institutions to evaluate the clinical competence of medical students.[2–10]

Each OSCE script represents a realistic clinical case, in which examinees are given the same problem and asked to execute the same task. The examiners evaluate students' clinical performance during the scenario based on the standardized checklist that includes medical history inquiry and communication skills, physical examination, clinical skills, examination procedures, clinical thinking, emergency training, and medical ethics. OSCE can be used as a formal evaluation method and for clinical teaching purposes. In addition, feedback from students, standardized patients and professors can all contribute to improving the curriculum and content of clinical education.

The term standardized patient refers to an individual instructed and trained by professional health care personnel to act as a patient according to the role-play script.[4] Standardized patients should act in an objective and realistic fashion as asked, and perform in a standardized manner, not to be modified over time or by the interlocutor. Simulation of clinical scenarios and of standard operation procedures helps facilitate an objective evaluation of students' clinical competence,[11] that is, to obtain systematically and skillfully a patient's basic medical history and undertake physical check-ups, to illustrate clinical logic and analytical and decision-making abilities to analyze and solve a patient's problem, and at the same time, to develop and refine communication skills with patients. The application of the OSCE helps students to develop the patient-centered concept so that patients do not need to worry about becoming a guinea pig for students, which reduces their uneasiness and sense of danger.

Within the past decade, evaluation of the clinical competence of medical students with OSCE has become a more advanced concept in Taiwan. To date, the majority of medical institutions in Taiwan have adopted, or are adopting, the OSCE method as an important reference when evaluating medical students' clinical competence.[5–10] Since 2006, Taipei Veterans General Hospital (TVGH) has targeted 7th-year medical student trainees who have received training at the institution as candidates for regular OSCE before the end of the semester. The objective of the present study was to analyze the preliminary experience with the OSCE in the hope of improving the examination quality.

Methods

Research subjects
This retrospective study collected and analyzed the relevant information of OSCEs conducted on 7th-year medical students at TVGH in 2006 and 2007, including the structure and content of the OSCE, students' OSCE scores, establishment of passing standards, examinees, examiners, and basic information of standardized patients. Most 7th-year medical students at TVGH came from medical schools in Taiwan, which included Chung Shan Medical University, China Medical University, Taipei Medical University, National Defense Medical Center, and National Yang-Ming University.

Research methods
At the end of 2005, representatives from the Clinical Skills Training Center and the Departments of Internal Medicine, Surgery, Obstetrics and Gynecology, and Pediatrics at TVGH convened to prepare for the OSCE. The examination was conducted during April and May of 2006 for 8 separate groups, with 18 students per group. The first version of the OSCE test comprised 15 questions, in which the short stations lasted for 10 minutes and the long stations for 20 minutes. Since then, all 7th-year medical students of the institution have been required to take the OSCE. The planning and organizing of the OSCE followed a set of guidelines. To begin, an OSCE executive team was assembled, in which the members had to be familiar with the tested areas and content. The background of the members had to include different medical fields, such as professors of internal medicine, surgery, obstetrics and gynecology, and pediatrics. In addition, the executive team was required to have the ability and influence to mobilize further human resources, financial support and materials, if necessary. The examination procedure had to be discussed and revised by experts. Sufficient numbers of standardized patients and examiners needed to be trained. Finally, the evaluation questionnaire for the standardized patients and examiners was also discussed and revised by the experts. The validity of our study depended on the expert validity.

Test tools
The Clinical Skills Training Center was used as the OSCE location. Volunteers from the institution, students from Chinese Culture University and Taipei National University of the Arts, and other volunteers were invited to participate as standardized patients. The volunteers were trained on how to act based on a common scenario and in a consistent way. Training was also provided for their feedback and evaluation skills.

Host seminars and the use of communication skills courses offered by National Yang-Ming University increased the performance opportunities for standardized patients, with the aim of increasing their performance experience. Assigned or voluntary examiners were recruited from the Departments of Internal Medicine, Surgery, Obstetrics and Gynecology, and Pediatrics at TVGH. They were trained to generate standardized and consistent evaluations and provide appropriate feedback to students. Seed examiners received training through seminars and communication skills courses that were offered by National Yang-Ming University to increase their evaluation and feedback acumen.

Data processing and analysis

All categories and results of examinees' evaluation at each station were recorded inclusively and compared statistically. After each standardized patient test, all examinees, standardized patients and examiners were required to fill out a questionnaire anonymously and provide feedback for the standardized patient test. Each test was analyzed using SPSS version 15.0 (SPSS Inc., Chicago, IL, USA). For descriptive analysis, each test was analyzed based on the number of examinees, maximum value, minimum value, mean value, standard deviation, and percentage to illustrate the distribution of each variable. Item analysis[12] was used to illustrate the following factors that might have influenced the quality of the test: (1) level of difficulty of each test item, calculated as the percentage of the passing students among the total examinees, with a smaller number meaning fewer passing students, and thus, a higher level of difficulty; (2) item–total correlation, in which test items with a lower correlation with the total score were deleted or modified; (3) squared multiple correlation, expressed as R^2, in which test items with a lower R^2 value were deleted or modified; and (4) effect on reliability after the deletion of such items, in which deletion of modification was done if this could have increased the reliability.

Results

OSCE structure and content

From 2006, TVGH instituted the OSCE for all 7th-year medical students. A total of 133 students took the test in 2006 (Table 1). There were 15 stations, and the content of the test preserved a traditional written question for internal medicine, surgery, and obstetrics and gynecology. During the chest X-ray (CXR) and electrocardiography (ECG) interpretations, an examiner was present to supervise the test. Of the 15 stations,

Table 1. Setting and implementation information of the objective structured clinical examination for interns at Taipei Veterans General Hospital, 2006 and 2007

	2006	2007
Number of stations	15	15
Number of SP stations	2	6
Duration of station		
Short (10 min)	12	12
Long (20 min)	3	3
Number of students	133	132

SP = standardized patient.

only 2 used standardized patients, 1 of which was an internal medicine medical history inquiry and physical check-up, and the other focused on communication skills. In 2007, 132 students participated in the test, which still comprised 15 stations. Besides the original 2 stations of standardized patients, surgery also began using standardized patient questions. In addition, the Departments of Obstetrics and Gynecology, and Pediatrics also incorporated interactive questions, with examiners acting as standardized patients. Only 1 written test from the Department of Internal Medicine remained. In the CXR and ECG interpretations, there were no longer examiners present. The contents of the test are shown in Table 2.

OSCE results

After the experience of the 1st OSCE in 2006, numerous reviews and improvements were made. The test results of 133 students from the 2nd year (2007) were analyzed (Table 3). The average scores of students from the 15 stations ranged from 47.7 ± 16.4 to 93.7 ± 8.5. The score for communication skills was the lowest, whereas the score for Micro-Sim was the highest. A holistic method was applied to establish the passing standard. Any score below the borderline of 60% (average score over the highest possible score) was considered as failure. Communication skills and ECG interpretation were the 2 categories that most of the students failed. A reliability analysis was conducted for the 2007 OSCE questions. The overall score and reliability (Cronbach's reliability) was 0.641. The difference between the impacts on reliability after deleting a test item ranged from 0.59 to 0.65 for all stations, which was not significant. This meant that every station had a similar impact on reliability after being deleted. The impact was highest in the CXR and ECG interpretations. The squared multiple correlation, R^2, of the reliability of each item was between 0.13 and

Table 2. Content of stations at Taipei Veterans General Hospital, 2006 and 2007

	2006	2007
Internal medicine		
History taking and physical examination	●	●
Psychomotor skills		
Electrocardiography interpretation	○	□
Chest X-ray interpretation	○	□
Others	○	○
	○	○
Written test	□	□
Surgery		
History taking and physical examination		●
Psychomotor skills	○	○
	○	
Written test	□	
Obstetrics and gynecology		
History taking and physical examination		◎
		◎
Reasoning, differential diagnosis	○	
Written test	□	
Pediatrics		
History taking and communication skills		◎
Reasoning, differential diagnosis	○	
Psychomotor skills	○	○
Others		
Communication skills	●	●
Mega Code		○
Micro-Sim	□	□

● = station with a standardized patient and an examiner; ○ = station with an examiner (not as standardized patient); □ = station without any standardized patient or examiner; ◎ = station with an examiner (also as standardized patient).

0.49, with CXR interpretation being the lowest. The item–total correlation was between 0.10 and 0.41, with interactive case being the lowest.

Discussion

Since the creation of the OSCE in 1975, countries such as Canada, the USA and Japan have integrated it into the national medical examination. Since 2009, Korea has also integrated the OSCE as part of the requirements for medical students to obtain practice licenses.[13] In Taiwan, the National Defense Medical Center[14] and National Yang-Ming University[15] have gradually adopted this test method to evaluate medical students' clinical performance. Kaohsiung Medical University established the standardized patient program and recruited standardized patients for the OSCE in 2003.[5]

Since 2003, National Cheng Kung University has targeted 6th-year medical students to take the OSCE.[6] The National Defense Medical Center trained 6 types of neurological disease standardized patients in 2003[7] and designed 8 types of common emergency diseases in 2004,[8] to evaluate the clinical competence of 7th-year medical students. The Department of Surgery of National Taiwan University Hospital also implemented the standardized patient test for 7th-year students in 2005.[10]

TVGH initiated preparation for the OSCE at the end of 2005, and it was implemented for all 7th-year medical students in 2006. However, unlike other medical institutions and hospitals, the majority of the 7th-year medical students came from different schools and received varied training prior to their internship. In terms of test content, it has been adjusted gradually after several years of experience. For example, in 2006 and 2007, written questions were included, but in the intervening years, there have been no written questions.

At the beginning of the OSCE preparation, due to the lack of human resources, senior residents sometimes have been required to act in the dual role of standardized patients and examiners. Both domestic and international literature has, however, highlighted the shortcomings of such an arrangement.[9,16] The main drawback of this approach is that the standardized patients (examiners) and examinees have mostly teacher–student or senior–junior relationships, and examinees might feel intimated, thus influencing the result of the evaluation. As standardized patient teaching and evaluation has become the norm over the years,[17,18] and more resources and standardized patients are available, this problem no long exists, which contributes to the reliability and validity of the OSCE. Nevertheless, there are still problems with regard to standardized patients. The Association of Standardized Patient Educators and Howley et al[19] conducted extensive searches of the English-language medical literature from 1995 to 2005, which has revealed that, among the 121 relevant studies identified, discussion of age (22%) and sex of standardized patients (27%), test particularity (21%), standardized patient training (15%), and standardized patient performance measurement (38%) has been low, and all these factors could have an impact on test reliability and validity. Another study has found that the sex of standardized patients might influence examinees' performance in the OSCE, which is irrelevant to the sex of the examinee.[20] Therefore, it is suggested that same-sex standardized patients should be used for the same-case scenario, even though the question itself is neutral.

Table 3. Station analysis, 2007

Station	Score (%)	Difficulty (%)	Station analysis		
			Item–total	R^2	αif
Chest X-ray	72.4±16.0	82.6	0.12	0.22	0.65
Electrocardiography	56.0±20.0	56.1	0.20	0.23	0.65
Physical examination	71.3±10.8	87.1	0.13	0.16	0.64
Endotracheal tube intubation	80.4±8.7	100.0	0.34	0.24	0.62
Appendicitis	88.1±10.9	99.2	0.49	0.35	0.59
Interactive case	74.6±10.8	83.3	0.25	0.10	0.63
Suturing	82.8±11.3	97.0	0.45	0.34	0.60
Pneumothorax	71.7±11.2	97.0	0.27	0.23	0.62
Arrhythmia	73.5±12.2	87.1	0.30	0.33	0.62
Neonatal jaundice	75.9±12.2	93.2	0.33	0.40	0.61
Abdominal pain	76.1±9.0	99.2	0.20	0.39	0.63
Vaginal bleeding	76.2±11.9	94.7	0.28	0.41	0.62
Communication skills	47.7±16.4	25.8	0.23	0.28	0.63
Mega Code	81.9±9.8	96.2	0.32	0.17	0.62
Micro-Sim	93.7±8.5	100.0	0.29	0.18	0.62

Difficulty = (passed/total students) × 100%; Item–total = item–total correlation; R^2 (squared multiple R) = variance explained; αif = Cronbach if deleted.

In our institution, there are more female standardized patients than male; whether or not this has an impact on the examinees is not known and is in need of further study.

Given that the OSCE requires extensive human resources, it might succumb to the predicament of examiner shortages. Recently, a study was conducted on examiners by a medical school in Germany.[21] A total of 214 3rd-year medical students were subjected to a 10-station OSCE. Before the test, 25 teaching doctors and 20 4th- to 5th-year students were trained as examiners. At 4 stations, 2 examiners were assigned, including a teaching doctor and a medical student. Although medical students tended to give higher scores, overall, this arrangement did not result in any statistically significant difference. The final result showed that both groups gave similar scores, which were accepted by all examinees. Furthermore, 1 study even assigned standardized patients as examiners.[22] This study demonstrates that the score given by standardized patients is often higher than that given by doctors. The correlation between the total score and the written test score was minimal, which indicates that the score given by standardized patients should not replace that given by doctors.

Previous foreign reports have indicated that OSCE reliability can be affected by the examination time. As the time was prolonged, the OSCE reliability increased from 0.47 at 1 hour, to 0.64 at 2 hours, 0.78 at 3 hours, and 0.88 at 4 hours.[23] Although the OSCE has long been adopted in Taiwan, few studies have addressed its reliability. To date, only Cheng Kung University has analyzed OSCE scores, which involved 73 6th-year students in 2005.[6] The analysis revealed an internal reliability of 0.62. The study has also categorized 3 aspects of measurement, including standard clinical competence, advanced clinical competence, and psychomotor skills. The internal reliability of the measurements was 0.92, 0.81 and 0.77, respectively. The present study analyzed the performance of 132 OSCE examinees at TVGH in 2007 and calculated an overall reliability of 0.641. The overall reliability was acceptable when compared with previous studies. From our current study, we found that the CXR interpretation and the interactive case had the lowest reliability by squared multiple and item–total correlation, respectively. This means that we should try to improve the setting and the checklists in these 2 stations in future examinations. We are still at an early stage; therefore, there is much room for improvement. With an increased understanding of the OSCE over the past few years, we have been able to strengthen our capacities in different areas, including enhancement of education quality, improvement of evaluation forms, and strengthening of standardized patient and examiner training, with the anticipation of raising the reliability of the OSCE in the future.

Two years ago, our passing standard was based on an absolute value approach. All scores under 60% were considered a failure. The OSCE differs from the traditional written test; therefore, the passing standard of 60% might no longer be appropriate. Currently,

there are various ways of setting passing standards, including absolute (criterion-based) or relative (norm-based).[24] An absolute standard determines the outcome (pass/fail) by how well the examinee performs. The examinee is usually judged against an arbitrarily set external standard. Hence, the examinee's success is independent of the performance of the group. A relative standard compares how well the examinee has performed compared with others who took the test. Hence the outcome (pass/fail) is dependent on the performance of the group. At present, there is no consensus on the standard in Taiwan. Although definition of a standard is an important step in developing an advanced OSCE, this will require extensive discussion before a consensus can be reached to enhance OSCE quality in Taiwan.

There were several limitations to the present study. First, the study was retrospective. The impact of standardized patients and examiners on the performance of the medical students could not be assessed correctly. However, the preliminary experience of OSCE will help us to conduct further studies to assess the possible factors that might influence the performance of medical students and improve our OSCE. Second, our study had a cross-sectional design to count the score of a single OSCE. We do not know whether there was improvement in the medical students. In future years, we will develop OSCEs to assess medical students before and after training. Third, we used the total scores for each station. The same or similar items might have been included in different stations, such as communication skills, history-taking, or physical examination. However, we did not separate them into different items. Further study is needed to divide the stations according to different competences.

The OSCE is an important clinical competence evaluation that will, in the future, grant access to the national medical practitioner license test for medical students. How to elevate the quality of the OSCE is currently a pressing issue. The early experience of the TVGH will provide the foundation for future improvement. Through a better understanding of the OSCE, improvement of educational material quality and staff training, and unification of passing standards, the quality of the OSCE can be effectively elevated.

Acknowledgments

References

1. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. *Br Med J* 1975;1:447–51.
2. Fowell SL, Maudsley G, Maguire P, Leinster SJ, Bligh J. Student assessment in undergraduate medical education in the United Kingdom, 1998. *Med Educ* 2000;34(Suppl):1–49.
3. Onishi H, Yoshida I. Rapid change in Japanese medical education. *Med Teach* 2004;26:403–8.
4. Vu N, Baroffio A, Huber P, Layat C, Gerbase M, Nendaz M. Assessing clinical competence: a pilot project to evaluate the feasibility of a standardized patient-based practical examination as a component of the Swiss certification process. *Swiss Med Wkly* 2006;136:392–9.
5. Huang YS, Liu M, Huang CH, Liu KM. Implementation of an OSCE at Kaohsiung Medical University. *Kaohsiung J Med Sci* 2007;23:161–9.
6. Tsai TC, Lin CH, Chen CL, Chao CC, Yeh TL, Tsai JJ, Chang YF. Analysis of OSCE results: experience in National Cheng Kung University Medical College. *J Med Educ* 2006;10:313–23.
7. Chang BY, Shieh JY, Hsu YD, Wang JO, Kao SY, Chang SY, Ho ST. Evaluating the clinical competence of the intern. *J Med Educ* 2005;9:233–40.
8. Chang BY, Shieh JY, Wang JO, Sheen MJ, Kao SY, Chan SY, Ho ST. Evaluating the clinical competence of emergency medicine interns using standardized patients. *J Med Educ* 2007;11:27–38.
9. Tsai MS, Chen JS, Lai IR, Lin MT, Lai HS, Lee PH. Using standardized patients in surgical tests. *J Med Educ* 2006;10:34–44.
10. Tsai SL, Yang CW, Yeh CC, Chang SC. The recruitment and training of standardized patients at National Taiwan University Hospital. *J Med Educ* 2007;11:80–7.
11. Chen PT, Cheng HW, Yen CR, Yin IW, Huang YC, Wang CC, Tsou MY, et al. Instructor-based real-time multimedia medical simulation to update concepts of difficult airway management for experienced airway practitioners. *J Chin Med Assoc* 2008; 71:174–9.
12. Auewarakul C, Downing SM, Praditsuwan R, Jaturatamrong U. Item analysis to improve reliability for an internal medicine undergraduate OSCE. *Adv Health Sci Educ Theory Pract* 2005; 10:105–13.
13. Lee YS. OSCE for the Medical Licensing Examination in Korea. *Kaohsiung J Med Sci* 2008;24:646–50.
14. Shieh JY, Chang BY, Kao SY, Wang JO, Ho ST, Chang SY. Standardized patient and medical education. *J Med Educ* 2004; 8:243–54.
15. Lin MH, Cheng CH, Tsai ST. The assessment of "Clinical Examination" teaching by Group Objective Structured Clinical Examination (GOSCE). *J Med Educ* 2000;4:454–65.
16. Barrows H. Simulated (standardized) patients and other human simulations. *Med Educ* 1998;32:13–24.
17. Stillman PL, Regan MB, Philbin M. Results of a survey on the use of standardised patients to teach and evaluate clinical skills. *Acad Med* 1990;65:288–92.
18. Barrows HS. An overview of the uses of standardized patients for teaching and evaluating clinical skills. *Acad Med* 1993;68: 443–53.
19. Howley L, Szauter K, Perkowski L, Clifton M, McNaughton N. Association of Standardized Patient Educators (ASPE). Quality of standardised patient research reports in the medical education literature: review and recommendations. *Med Educ* 2008;42: 350–8.
20. Humphrey-Murto S, Touchie C, Wood TJ, Smee S. Does the gender of the standardised patient influence candidate performance in an objective structured clinical examination? *Med Educ* 2009;43:521–5.

21. Chenot JF, Simmenroth-Nayda A, Koch A, Fischer T, Scherer M, Emmert B, Stanske B, et al. Can student tutors act as examiners in an objective structured clinical examination? *Med Educ* 2007;41:1032–8.

22. McLaughlin K, Gregor L, Jones A, Coderre S. Can standardized patients replace physicians as OSCE examiners? *BMC Med Educ* 2006;6:12.

23. Petrusa A. Clinical performance assessments. In: Norman G, Van der Vleuten CPM, Newble DI, eds. *International Handbook of Research in Medical Education.* Dordrecht, Boston, London: Kluwer Academic Publishers, 2002: 673–709.

24. Turnbull JM. What is normative versus criterion-referenced assessment. *Med Teach* 1989;11:145–50.