Tzu Chi Medical Journal 26 (2014) 40–43

Contents lists available at ScienceDirect

Tzu Chi Medical Journal

journal homepage: www.tzuchimedjnl.com

Original Article

Objective Structured Clinical Examination (OSCE) including critical simulation: Evaluation of medical student competence

Ming-Chen Hsieh [a,b,c,*], Wei-Chun Cheng [d], Tsung-Ying Chen [a,b,d]

[a] Department of Medical Education, Buddhist Tzu Chi General Hospital, Hualien, Taiwan
[b] Department of Medicine, College of Medicine, Tzu Chi University, Hualien, Taiwan
[c] Department of Curriculum Design and Human Potentials Development, National Dong Hwa University, Hualien, Taiwan
[d] Department of Anesthesiology, Buddhist Tzu Chi General Hospital, Hualien, Taiwan

ARTICLE INFO

Article history:
Received 20 May 2013
Received in revised form
10 June 2013
Accepted 15 July 2013

Keywords:
Assessment
Critical condition
Objective Structured Clinical Examination

ABSTRACT

Objectives: To develop an Objective Structured Clinical Examination (OSCE) station to assess the evaluation skills of medical students in applying evidence and appropriate treatment options in critical situations with a simulated patient. To assess the results using discrimination and reliability comparison of standardized and simulated patient stations.
Materials and Methods: OSCE performance scores of 58 7th-year medical students at the University of Tzu-Chi School of Medicine were analyzed from April 10, 2011 to April 11, 2011 using descriptive statistics and item discrimination. Thirteen OSCE cases were identified for evaluation; we compared the results of all the stations to those of the station with the critical clinical scenario.
Results: Discrimination statistics indicated that only the critical scenario station prepared with a high-fidelity simulator was effective in distinguishing between high-scoring and low-scoring medical students.
Conclusion: Failure to design a skill assessment tool is a missed opportunity to understand more fully and apply the results of the clinical performance of medical students.

1. Introduction

Competency-based education has been popular in medical education for the past decade and is currently the mainstream method of teaching clinical knowledge. It tries to incorporate new models to create medical education objectives [1]. Assessing student clinical skills is also a crucial element in their training. The Objective Structured Clinical Examination (OSCE) is a widely accepted tool to evaluate the clinical competence of medical students [2]. Studies have demonstrated that the OSCE is an effective tool for evaluating areas most critical to the performance of healthcare professionals, such as the ability to obtain information from a patient, establish rapport and communicate, and interpret data and solve problems [3]. Although assessment may be part of an institution or course evaluative process, or have other purposes, teachers use assessment for either summative or formative processes [4]. The OSCE station content varies according to student experience and the nature of the assessment. The types of problems portrayed in an OSCE are those that students would commonly encounter in a clinic or hospital. Standardized patients (SPs) typically have general complaints, although some could present with problems related to emergency conditions [5,6]. Although students in training are familiar with basic practices in critical care medicine, an OSCE is seldom included when evaluating the condition of critically ill patients [7]. During the clinical rotation, medical students have direct patient care responsibilities through which they learn about various forms of critical illness, and how to apply different therapeutic and diagnostic modalities commonly used in critical care medicine [8]. Failure to address critical condition performance is a missed opportunity to understand better and use the results of such examinations for a competence-based evaluation for medical students [9]. Developing an OSCE station for complex critical conditions poses unique challenges. However, current technology allows for critical care scenarios, complete with cardiac and respiratory arrest on a computerized patient simulator in rapid transit stations, such as OSCE [10].

The study design was chosen to allow for collecting quantitative measures of medical student performance in managing a set of

Conflicts of interest: none.

* Corresponding author. Department of Medical Education, Buddhist Tzu Chi General Hospital, 707, Section 3, Chung-Yang Road, Hualien, Taiwan. Tel./fax: +886 3 8462722.
E-mail address: jeany.jeany@msa.hinet.net (M.-C. Hsieh).

1016-3190/$ – see front matter Copyright © 2013, Buddhist Compassion Relief Tzu Chi Foundation. Published by Elsevier Taiwan LLC. All rights reserved.
http://dx.doi.org/10.1016/j.tcmj.2013.08.001

simulated critical shock emergencies. We developed an OSCE station to assess the evaluation skills of medical students in applying evidence and appropriate treatment options in critical situations with a simulated patient. This investigation determined whether critical management OSCE stations play a meaningful role in a summative examination, and assesses the results using discrimination and reliability comparison of standardized and simulated patient stations.

2. Materials and methods

2.1. Study participants

The Department of Medical Education and School of Medicine at Tzu Chi University in Hualien, Taiwan has held routine OSCE examinations since 1996. This retrospective study collected and analyzed relevant OSCE information from 7th-year medical students at Buddhist Tzu Chi General Hospital in 2010. Fifty-eight participants had completed training courses in various subjects, including internal medicine, surgery, pediatrics, and critical care. This study was reviewed and performed by the Research Ethics Committee in accordance with Institutional Review Board (IRB) approval at Buddhist Tzu Chi General Hospital (IRB 101-06).

2.2. Study design

The development of the OSCE examination component was based on a collaborative effort led by faculty members who had experience with case design. The OSCE examined the range of clinical competence in clinical scenarios including interviewing, physical examination skills, critical thinking, clinical judgments, and technical skills. All participants were instructed to perform all appropriate diagnostic and therapeutic actions and verbalize their thoughts and actions. This study focused on assessing the critical thinking abilities of students.

2.3. Scenarios

Students had 1 week of hands-on participation to familiarize themselves with the simulators, represented by an experienced operator prior to the test. During the OSCE, a simulated scenario was conducted in a general ward featuring a high-fidelity simulator. We used the iStan (METI, Medical Education Technologies, Inc., Sarasota, FL, USA), which provides a human-like, full-scale computerized mannequin in a realistic clinical setting. The scenario lasted 15 minutes. Participants were given clear instructions to state the emergency diagnosis and the treatments they were instituting. We presented a 55-year-old man who was admitted to the hospital because of pneumonia complicated by hypotension. Two status respiratory failure and septic shock were shown in the stages. The data, including medication orders, vital signs records, electrocardiograms, and chest radiographs, were collected in the chart. The students needed to assess the patient, including a review of the patient chart, and perform a physical examination. The scenario ended when the patient began a downhill course. Following the station, participants were required to provide a brief summary as a duty note to display the assessment, problems, and plans in an organized format.

2.4. Scoring

Audiovisual recordings were made of each scenario to facilitate scoring and to allow independent review and further analyses. The crisis evaluation and summary of the event and scoring measures are presented in Appendix 1. Five medically qualified educators designed the written sheet for the patient notes, which included four sections: subjective, objective, diagnosis, and plan. The three-part checklist included a history and physical examination, imperative diagnosis with differential diagnosis, and management of septic shock. Reference resources for evaluating the management of severe sepsis and septic shock skills were based on Surviving Sepsis Campaign International Guidelines [11]. A panel of four experienced physicians using a modified Delphi technique selected and prioritized the passing score. For the OSCE, four experienced raters were formally trained in assessing each examination paper and were given specific instructions on scoring.

Scoring was done on a three-point scale ranging from 0 (failed to perform) to 1 (performed poorly or out of sequence) to 2 (properly performed in correct sequence).

2.5. Data processing and analysis

For descriptive analysis, data from a high-fidelity simulator station was analyzed, including the maximum score, minimum score, mean score, and standard deviation. We compared the pass rate, quality estimation between SP stations, and the high-fidelity simulator station (Table 1). The measure of item difficulty (P) – the proportion of participants, who received credit for the item, was based on the average of the two raters' values. A value of 1 indicated that all students received credit. The second measure was item discrimination (D) – the correlation between the item-level score and the total checklist score. Here, higher values (i.e., $D > 0.30$) indicated that the item was able to discriminate between low- and high-ability individuals. In some instances (i.e., all or no students receiving credit), the D value could not be calculated. The third measure was reliability between inter-rater agreement, which was estimated as the Pearson product–moment correlation coefficient between two administrations of the same measure. A value of 1 indicated that the two raters were in perfect agreement on a particular element

Table 1
Difficulty and discrimination of OSCE and simulation test.

Station	Category	HSG[a] score (no. passed/ pass rate, %)	LSG[a] score (no. passed/ pass rate, %)	(P)[b]	(D)[c]
1a	SP	16/100	15/93.75	0.97	0.0625
1b		15/93.75	10/62.50	0.78	0.3125
2a	SP	16/100	16/100	1	0
2b		16/100	11/68.75	0.84	0.3125
3	SP	16/100	16/100	1	0
4	SP	16/100	15/93.75	0.97	0.0625
5	SP	16/100	16/100	1	0
6a	SP	16/100	13/81.25	0.91	0.1875
6b		16/100	11/68.75	0.84	0.3125
7a	SP	16/100	13/81.25	0.91	0.1875
7b		14/87.5	10/62.50	0.75	0.25
8	SP	16/100	14/87.50	0.94	0.125
9	HFS	14/87.5	6/37.50	0.63	0.5

HFS = high-fidelity simulator; HSG = high-scoring group; LSG = low-scoring group; OSCE = Objective Structured Clinical Examination; SP = standardized patient.

[a] Ordered scores of all students from high to low points, select group in first 16 students (27%) as HSG; last 16 students (27%) as LSG.

[b] Difficulty (P) = (Pass rate in HSG + Pass rate in LSG)/2 × 100; (1) $p < 0.25$: too difficult; (2) $0.25 < p < 0.4$: difficult; (3) $0.4 < p < 0.7$: appropriate; (4) $0.7 < p < 0.9$: easy; (5) $p > 0.9$: too easy.

[c] Discrimination (D) = Pass rate in HSG – Pass rate in LSG/100; (1) >0.4: excellent; (2) 0.3–0.39: good, modify probably; (3) 0.2–0.29: acceptable, modify often; (4) <0.19: poor replace with better one or modify.

Table 2
Items estimation in high-fidelity simulator station.

Item	P+	D+	Pbis	Suggestion
Physics				
$1._{1}$	0.34	0.19	0.22	Replace with better one or modify
$1._{2}$	0.62	0.29	0.36	Modify often
$1._{3.a}$	0.88	−0.05	0.03	Replace with better one or modify
$1._{3.c}$	0.64	0.21	0.25	Modify often
Patient's history				
$1._{4}$	0.38	0.37	0.31	Modify probably
Relevant laboratory analysis				
$1._{5.a}$	0.69	0.25	0.21	Modify often
$1._{5.b}$	0.34	0.42	0.3	Modify un-necessarily
$1._{5.c}$	0.29	0.32	0.3	Modify probably
Clinical judgment				
$2._{1}$	0.74	0.21	0.19	Modify often
Differential diagnosis				
$2._{2}$	0.36	0.2	0.28	Modify often
Management plans				
$3._{1}$	0.41	0.11	0.1	Replace with better one or modify
$3._{3.b}$	0.88	0.22	0.22	Modify often
$3._{3.c}$	0.5	0.38	0.35	Modify probably
$3._{4.a}$	0.43	0.2	0.26	Modify often
$3._{4.c}$	0.53	0.07	0.16	Replace with better one or modify
$3._{4.d}$	0.09	0	0.16	Replace with better one or modify

Difficulty (P) = (Pass rate in HSG + Pass rate in LSG)/2 × 100; (1) $p < 0.25$: too difficult; (2) $0.25 < p < 0.4$: difficult; (3) $0.4 < p < 0.7$: appropriate; (4) $0.7 < p < 0.9$: easy; (5) $p > 0.9$: too easy.
Discrimination (D) = Pass rate in HSG − Pass rate in LSG/100; (1) >0.4: excellent; (2) 0.3–0.39: good, modify probably; (3) 0.2–0.29: acceptable, modify often; (4) <0.19: poor replace with better one or modify.

across all items in the SP station. For the high-fidelity simulator station, reliability was estimated using the Kendall coefficient of concordance for raters. A value > 0.9 showed high agreement between raters in individualized performance evaluations. The psychometric analysis shown in Table 2 was done to design each evaluation item in the high-fidelity simulator station. Student results from the National Medical Board were collected 3 months after this study. Analyses were processed by SPSS version 10.0 (SPSS Inc., Chicago, IL, USA). The correlation between significant part and results on the National Medical Board examination were estimated by the Pearson χ^2 test.

3. Results

Among the 39 men and 19 women who participated, five students were retained in the same grade for the internship year. After averaging scores from all raters of individual students in the simulation station, the scores of the 58 students ranged from 12 points (33.33%) to 30 points (83.33%), with a mean ± standard deviation of 19.5 (54.13%) ± 4.032 points (11.20%). Twenty-one of the fifty-eight students (36%) failed the examination. The low-scoring group obtained 16.78 ± 2.95 points and the high-scoring group obtained 21.31 ± 3.59 points. The discrimination statistics (D) indicated that the only station prepared with a high-fidelity simulator was effective in distinguishing between low- and high-ability medical students. Other than the simulator station, discrimination statistics from the SP stations indicated that both low- and high-ability students performed equally with respect to these settings.

Three months later, all students took the National Medical Board qualification in Taiwan. Seven students failed the examination. The pass rate was calculated as 87.9%. Binomial logistic regression found no significant association between results of the National Medical Board examination and the sections physical assessment, taking a history, differential diagnosis, and management, including the average score for the entire test. Data analyzed by the χ^2 test showed no significance ($p = 0.219$).

4. Discussion

There were two main findings in this study. First, quality estimation of each station and each item showed effectiveness in the high-fidelity simulator station, including discrimination between high-scoring and low-scoring student effectiveness. Second, our findings had no significant correlation with the certification examination, proving that we could discern different levels of patient care from individuals in a standardized scenario on a high-fidelity simulator. Quality control is another important issue in test development, particularly for certifying examinations used to classify examinees. The importance of measuring medical professionalism is receiving renewed attention [12].

The OSCE, a tool to assess medical student clinical competence objectively and fairly, has become widely used in medical education worldwide. However, most medical schools in Taiwan have just begun to adopt this assessment method [13]. The Taiwan National Medical Board examination may need to be reformed because it has major problems including inappropriate questions, low discrimination of clinical judgment and management, and a failure rate as high as 50% of participants annually from 2001 to 2006 [14]. The importance of professionalism in medical schools is receiving renewed attention [15]. Written examinations can be used to test student knowledge of clinical and procedural skills, but this method alone may lead students to focus on memorizing skills instead of actually performing them [16]. Clinical skill examinations, like the OSCE, should be implemented in the National Medical Board in Taiwan in 2013, similar to the second step in the United States Medical Licensing Examination. This test focuses on interaction between participants and SPs and mostly expresses communication skills and basic physical assessment [17]. However, there is criticism that the OSCE is limited in assessing the integrated approach to patients and that students are trained to perform [5]. From our study, the OSCE stations demonstrated low discrimination value and need careful revision and critical review. Low-discrepancy effectiveness on student "basic clinical skills" in the OSCE and clinical performance were comparable with the high-fidelity simulation scenario test because actual evaluation and decision-making abilities were all dismissed.

In an OSCE examination, candidates move through numerous short clinical scenarios designed to focus on a range of topics and specific clinical skills [18]. These scenarios mostly range from taking a patient's general history to asking appropriate questions. However, our experience indicates that the "actual patient care" simulation scenario test may be more correlated with individual differences [19]. Several recent illustrations of simulation-based evaluation research have shown training transfer to patient care settings [20]. Critical actions that do not address skill assessment ability signal a missed chance to understand better and apply the results of OSCE examinations to the clinical performance development of medical students [21]. Each OSCE station and checklist item should be carefully selected because it contributes synergistically to the total performance of the clinical skill being assessed [22].

In conclusion, the OSCE is an important tool for clinical competence evaluation that will soon be included in national medical practitioner license tests for medical students worldwide. How to improve the quality of the OSCE and assess student ability, such as actual management of a patient in an emergency situation, is currently a critical issue.

Appendix 1
Summary of events and scoring sheet.

(A) Subjective and objective items:	2	1	0
$1._1$ GCS: G1M1V1	☐	–	☐
$1._2$. Vital signs assessment	☐	–	☐
$1._3$. Pertinent physical assessment: a. breath sounds	–	☐	☐
b. respiratory pattern (rate, depth, symmetry, etc.)	☐	☐	☐
c. sputum color, character, amount	–	☐	☐
$1._4$. Patient history	☐	–	☐
$1._5$ Recent laboratory data: a. chest radiographs	–	☐	☐
b.WBC/DC/BCS	–	☐	☐
c.ABG	–	☐	☐
(B) Assessment:			
$2._1$. Shock	☐	–	☐
$2._2$. *Differential diagnosis (UTI, sepsis, CAD, etc.)	–	☐	☐
(C) Plan:			
$3._1$.Sputum suction/chest care	☐	–	☐
$3._2$.Oxygen supply/maintenance/advanced airway management	☐	☐	☐
$3._3$.**Early goals of septic shock: a. broad-spectrum antibiotics	☐	☐	☐
b. IV fluid challenge	☐	–	☐
c. CVP/PA wedge pressure	☐	–	☐
d. Hemoglobin, lactate levels	☐	☐	☐
e. Inotropic agent use: dopamine or norepinephrine	☐	☐	☐
$3._4$.Appropriate laboratory data/ management:			
a. input/output monitoring or urine output recording (including Foley catheterization)	☐	–	☐
b. Blood or sputum culture	☐	☐	☐
c. Other laboratory data (CRP, urine culture)	–	☐	☐
d. Acid-fast stain, TB	–	☐	☐

0 = not attempted; 1 = attempted but incomplete or inadequate; 2 = performed adequately and completely.
WBC/DC = white blood cell and differential count; ABG = arterial blood gas analysis; BCS = blood chemistry; CAD = coronary artery disease; CRP = C-reactive protein; CVP = central venous pressure; GCS = Glasgow Coma Scale; IV = intravenous; PA = pulmonary artery; TB = tuberculosis; UTI = urinary tract infection.

References

[1] Liu M, Huang YS, Liu KM. Assessing core clinical competencies required of medical graduates in Taiwan. Kaohsiung J Med Sci 2006;2210:475–83.
[2] Townsend AH, McLlvenny S, Miller CJ, Dunn EV. The use of an objective structured clinical examination (OSCE) for formative and summative assessment in a general practice clinical attachment and its relationship to final medical school examination performance. Med Educ 2001;359:841–6.
[3] Mitchell ML, Henderson A, Groves M, Dalton M, Nulty D. The objective structured clinical examination (OSCE): optimising its value in the undergraduate nursing curriculum. Nurse Educ Today 2009;294:398–404.
[4] Crawford III CH, Nyland J, Roberts CS, Johnson JR. Relationship among United States Medical Licensing Step I, orthopedic in-training, subjective clinical performance evaluations, and American board of orthopedic surgery examination scores: a 12-year review of an orthopedic surgery residency program. J Surg Educ 2010;672:71–8.
[5] Collins LG, Schrimmer A, Diamond J, Burke J. Evaluating verbal and non-verbal communication skills, in an ethnogeriatric OSCE. Patient Educ Couns 2011;832:158–62.
[6] Williams RG. Have standardized patient examinations stood the test of time and experience? Teach Learn Med 2004;162:215–22.
[7] Payne NJ, Bradley EB, Heald EB, Maughan KL, Michaelsen VE, Wang XQ, et al. Sharpening the eye of the OSCE with critical action analysis. Acad Med 2008;8310:900–5.
[8] Levy MM, Dellinger RP, Townsend SR, Linde-Zwirble WT, Marshall JC, Bion J, et al. The Surviving Sepsis Campaign: results of an international guideline-based performance improvement program targeting severe sepsis. Intensive Care Med 2010;362:222–31.
[9] Baid H. The objective structured clinical examination within intensive care nursing education. Nurs Crit Care 2011;162:99–105.
[10] Wayne DB, Didwania A, Feinglass J, Fudala MJ, Barsuk JH, McGaghie WC. Simulation-based education improves quality of care during cardiac arrest team responses at an academic teaching hospital: a case-control study. Chest 2008;133:56–61.
[11] Marshall JC. Principles of source control in the early management of sepsis. Curr Infect Dis Rep 2010;12:345–53.
[12] Lin JD, Wu JL, Yen CF. An exploratory study into health care policy for persons with intellectual disabilities in Taiwan. J Intellect Disabil Res 2004;48:252–61.
[13] Huang YS, Liu M, Huang CH, Liu KM. Implementation of an OSCE at Kaohsiung Medical University. Kaohsiung J Med Sci 2007;234:161–9.
[14] Chiu CH, Tsai D. Medical education reform in Taiwan. Med Teach 2010;321:91.
[15] Lai CW. Experiences of accreditation of medical education in Taiwan. J Educ Eval Health Prof 2009;6:2.
[16] Rogers PL, Jacob H, Rashwan AS, Pinsky MR. Quantifying learning in medical students during a critical care medicine elective: a comparison of three evaluation instruments. Crit Care Med 2001;296:1268–73.
[17] Sisley AC, Johnson SB, Erickson W, Fortune JB. Use of an Objective Structured Clinical Examination (OSCE) for the assessment of physician performance in the ultrasound evaluation of trauma. J Trauma 1999;474:627–31.
[18] Falcone JL, Watson GA. Differential diagnosis in a 3-station acute abdominal pain objective structured clinical examination (OSCE): a needs assessment in third-year medical student performance and summative evaluation in the surgical clerkship. J Surg Educ 2011;684:266–9.
[19] Pell G, Fuller R, Homer M, Roberts T. International Association for Medical Education. How to measure the quality of the OSCE: review of metrics – AMEE guide no. 49. Med Teach 2010;32:802–11.
[20] Nackman GB, Bermann M, Hammond J. Effective use of human simulators in surgical education. J Surg Res 2003;115:214–8.
[21] Berkenstadt H, Ziv A, Gafni N, Sidi A. The validation process of incorporating simulation-based accreditation into the anesthesiology Israeli national board exams. Isr Med Assoc J 2006;810:728–33.
[22] White CB, Ross PT, Gruppen LD. Remediating students' failed OSCE performances at one school: the effects of self-assessment, reflection, and feedback. Acad Med 2009;845:651–4.