

# Objective Structured Clinical Examination as an assessment method for undergraduate medical students

Aisha M Al-Osail<sup>1</sup>, Mona H Al-Shiekh<sup>2</sup>, Abir H Al-Said<sup>1</sup>, Emad M Al-Osail<sup>1\*</sup>, Mohannad A Al-Ghamdi<sup>1\*</sup>, Abdulaziz M Al-hawas<sup>1\*</sup>, Abdullah S Al-bahussain<sup>1\*</sup>, Ahmed A Al-dajani<sup>1\*</sup>

<sup>1</sup>Department of Internal Medicine, University of Dammam, Khobar, Saudi Arabia.

<sup>2</sup>Department of Physiology, University of Dammam, Khobar, Saudi Arabia.

Correspondence to: M. Aisha Al-Osail, E-mail: aosail@uod.edu.sa

Received October 31, 2014. Accepted November 13, 2014.

## Abstract

**Background:** One of the most important objectives for University of Dammam is to send competent graduated physicians into the society, which cannot be accomplished without assessments of the clinical skills of the students. The Objective Structured Clinical Examination (OSCE) was recently implemented to improve the quality, reliability, validity, university standards, and international rankings, and to decrease time consumption by the exams.

**Objective:** To study OSCE as an assessment method for undergraduate medical students.

**Materials and Methods:** A pilot study was conducted over one semester. A total of 92 examinees took the OSCE and written exams in three groups. The OSCE comprised 20 clinical stations, which included noting histories, physical examinations, communication skills, and data interpretation. The written exam contained 80 multiple-choice questions.

**Results:** Cronbach  $\alpha$ 's by group were 0.62, 0.79, and 0.85. Correlations for all stations ranged from 0.6 to 0.8, which indicated good stability and internal consistency. The reliability of the written exam was found to be 0.85. The validity of the OSCE was assessed using Pearson correlations, which was found to be 0.6.

**Conclusion:** The OSCE is the gold standard for student assessment and is more reliable and valid than the traditional style of exam.

**KEY WORDS:** Objective Structured Clinical Examination, assessment, clinical skill, Cronbach, reliability, validity

## Introduction

The Objective Structured Clinical Examination (OSCE) is considered the standard method for medical student assessment for both preclinical and clinical courses internationally, especially in high-rank universities of the United States, Canada, the United Kingdom, and Australia. It has also become part of United States Medical Licensing Examination (USMLE),<sup>[1-3]</sup> Medical Council of Canada Evaluating Examination (MCCEE),<sup>[4]</sup> and Professional and Linguistic Assessments Board (PLAB) examination.<sup>[5]</sup>

\*Medical students

In 1979, Harden and Gleeson<sup>[6]</sup> implemented the first OSCE. Many changes were made in its design and structure to make it more efficient and reliable for evaluating the skills of the examinee.<sup>[7]</sup> It is the best way to assess multiple and different clinical skills that can show the students' knowledge, history taking, physical examination, communication skills, investigation interpretation, and management in professional and fair method.<sup>[8]</sup> The OSCE needs a lot of preparation from students because they must read about most of the systems to clear the exam, and they will be examined by a number of faculty members, EXAMINED by same faculty members, which will decrease the discrepancy and they will have similar stations and cases.

Long/short case had many disadvantages, which are summarized as follows<sup>[9,10]</sup>:

1. Time consuming (it will take 45–60 min for interviewing and examining real patients).
2. No direct observation for the students by faculty.
3. Absence of reliability index for evaluation.
4. Difference in patients' distribution among examinee (some will have easy cases and others difficult).
5. Difference in patients' cooperation with examinee.

### Access this article online

Website: <http://www.ijmsph.com>

DOI: 10.5455/ijmsph.2015.3110201437

Quick Response Code:



6. Differences of scoring among the examiner.
7. Discussion focused on few systems because of time limitation.
8. Discussion of investigation and management might be skipped because of time limitation.
9. Possibility of psychological and grading effects due to injustice in the scoring of students.<sup>[11–14]</sup>
10. Discrepancy in exam scores in comparison to clinical exam scores of some students.

The OSCE had many advantages in comparison to long/short cases and written exam or the written essay. The most important points are as follows:

1. It takes less time for assessment of multiple skills.
2. It standardizes patients/cases for all candidates.
3. It is a reliable exam.<sup>[15–17]</sup>
4. It is a valid test.<sup>[3,18]</sup>
5. It is considered as a fair exam.<sup>[6,19,20]</sup>

Many tools exist that are used to measure the reliability of the OSCE, such as Cronbach  $\alpha$  that mainly evaluates “stability.” It can reveal the differences in students’ performance at each station and afford a global rating, which evaluates the overall performance and if the checklist used is appropriate for the skill level of the students. The  $R^2$  coefficient is used to measure the proportional change in the dependent variable (the checklist score) by changes in the independent variable (the global grade); this is a marker of internal consistency.<sup>[21–26]</sup> The final method for establishing the validity of the exam is the comparison of the OSCE scores with written exam scores using Pearson correlation.<sup>[27]</sup>

In Saudi Arabia, King Saud University and King Abdulaziz University were the first Saudi universities to published their experiences about using the OSCE in the field of surgery, which was reported in detail by Al-Naami<sup>[28]</sup> at King Saud University (for students in their final year). This report concentrated on reliability and validity. The use of the OSCE in family medicine at King Saud University was reported by Raheel and Naeem<sup>[29]</sup>; the undergraduate perceptions of the exams were positive. For the field of dermatology, it was concluded that the OSCE was a gold standard exam, but again, there were no details about internal medicine.<sup>[28,29]</sup>

In 2014, the Internal Medicine Department at University of Dammam decided to shift from the old-style long/short cases

exams to the OSCE style after becoming aware of the more obvious disadvantages of the former exam style. These included the lack of establishing the reliability and validity of the exam, especially with the increasing number of medical students accepted and many students entering a residency or fellowship program in the United States, Canada, the United Kingdom, or Australia, where the OSCE is part of the licensing requirements (e.g., the USMLE, MCCEE, and PLAB). In addition, the OSCE will assist in the goals of sending competent physicians into the community and evaluating the quality and contents of courses.

The aims of this study were the following:

- To evaluate the reliability and validity of the OSCE.
- To assess if different reliability results affect the validity of the exam.
- To develop a standard for all examinees.
- To ensure the competency of graduates.

## Materials and Methods

### Participants

This pilot study was conducted during one semester (February to May) with 92 medical students, who took the exam in three groups (March, April, and May 2014). At the end of the semester, the students took the written exam, which comprise 80 multiple-choice questions.

### Procedure

Orientation lectures for the faculty were held about the OSCE; stations, the importance of the rubric for the checklist, and global ratings were explained. Meticulous and lenient consultants were excluded. An introductory orientation about the OSCE was given for each student group on the first day of the course.

The blueprint was established for each exam, and there were no repeated stations in the exam. The OSCE exam had 20 clinical stations and covered history taking, physical examinations, communication skills, and data interpretation. Each station took 7 min to complete. Students were divided in groups as shown in Table 1. The blueprint for each group covered all the systems in internal medicine, including communication skills, cardiology, the respiratory system, gastroenterology, endocrinology, hematology/oncology, nephrology,

**Table 1:** Reliability measures for the fifth-year OSCE

Day/year	Gender	Students/day	Stability <sup>a</sup>	Internal consistency <sup>b</sup>	p-Value	Internal consistency <sup>c</sup>
First group	Female	29/1	0.621	0.60	<0.001	0.61 (61%)
Second group	Female	21/1	0.799	0.621	<0.001	0.80 (80%)
Third group	Male	42/1	0.854	0.75	<0.001	0.85 (85%)

<sup>a</sup>Cronbach  $\alpha$

<sup>b</sup>Spearman rank correlation

<sup>c</sup> $R^2$  coefficient determinants

**Table 2:** Blueprint for the three groups

System	History			Examination			Data interpretation		
	First group	Second group	Third group	First group	Second group	Third group	First group	Second group	Third group
Cardiovascular	X	X	X	X	X	X	X		X
Pulmonary	X	X	X				X	X	X
Gastroenterology	X	X	X	X	X		X	X	
Rheumatology	X	X		X	X	X		X	X
Nephrology	X	X	X	X	X	X			
Infectious disease	X	X	X				X	X	
Hematology/ oncology	X	X	X			X	X	X	X
Endocrinology	X	X	X	X	X	X			X
General medicine			X	X	X	X			
Communication skills	X	X	X						

infectious disease, rheumatology, and general medicine, which are shown in Table 2. The exams were conducted for over 3 days for all the three groups for 5–7 h/day.

The highest total score was 100%; the OSCE exam accounted for 50%, a continuous assessment for 10%, and the written exam for 40%. All 92 students took the clinical and written exam; after each exam, the coordinator of the course met with faculty and students to assess and correct any problems with the OSCE to ensure better reliability in the future.

### Ethical Considerations

The study was approved by the Institutional Review Board of the University of Dammam (approval number: IRB-2014-01-317). Informed consent was obtained from all participants.

### Data Analysis

The exam reliability was assessed using Cronbach  $\alpha$ , the global rating (our rating was clear pass, borderline, and clear fail), and the coefficient of determination,  $R^2$ . Spearman rank correlation was used to evaluate the correlation between the checklist score and the global rate score. At the end of the semester, each student took the written exam, which was analyzed (mean, median, mode) separately for each year.<sup>[21–27]</sup> The validity was measured using Pearson correlation. Each system was analyzed to understand any deficits in the courses.

### Results

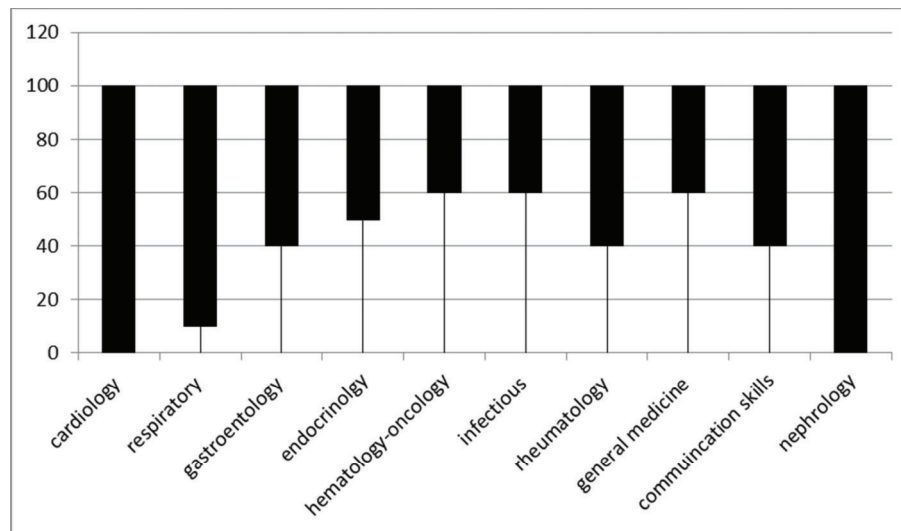
The reliability for the OSCE was evaluated using Cronbach  $\alpha$ , which indicated the stability of the stations on the three exams for the fifth-year students. The  $\alpha$ 's were 0.621, 0.799, and 0.854. Spearman rank correlation and  $R^2$  coefficient

determinants were used for correlating the checklist and the global score to arrive at an internal consistency score. The correlations were 0.6, 0.621, and 0.75 ( $p < 0.001$ ), which indicated a strong correlation between the checklist score and the global rating on all days of the exam. The  $R^2$  coefficient determinants, which were used to examine the linear correlation between the checklist and the global score, were 61%, 80%, and 85% for the fifth-year students, with the highest value in the male group. Spearman rank correlation and  $R^2$  coefficient determinant values did not differ much, which indicated very good internal consistency [Table 1].

Cronbach  $\alpha$  for the stations ranged from 0.5 to 0.9. Table 3 shows Cronbach  $\alpha$ 's for the stations based on the systems; we can see the improvement of Cronbach  $\alpha$  for each system after

**Table 3:** Cronbach  $\alpha$  analysis for the three groups

Station subjects	Cronbach $\alpha$ if deleted		
	First	Second	Third
Cardiovascular history	0.647	0.775	0.848
Cardiovascular examination	0.613	0.798	0.844
General medicine	0.64	0.772	0.90
Pulmonary history	0.9	0.801	0.841
Infectious disease history	0.637	0.903	0.847
Nephrology examination	0.591	0.788	0.843
Nephrology history	0.627	0.782	0.847
Gastroenterology history	0.602	0.792	0.857
Gastroenterology examination	0.633	0.789	0.85
Rheumatology history	0.618	0.789	0.848
Rheumatology examination	0.582	0.784	0.846
Endocrinology history	0.626	0.90	0.90
Endocrinology examination	0.589	0.798	0.839
Hematology history	0.571	0.776	0.84
Communication skills	0.592	0.90	0.843



**Figure 1:** The OSCE scores by systems

each exam such as nephrology, rheumatology, endocrinology, hematology, and communication skill. The score ranges for each system are shown in Figure 1, which were calculated given the total possible score of 100.

The OSCE score for students was between 20 and 45.9 of 50, with mean of 37.8, median of 38.7, and skewness of -1, and standard deviation (SD) of 4.56, which indicated most of the scores were around the right side of the mean and the extreme values were on the left. The OSCE score analysis for the students in the fifth years is shown in detail in Table 4.

The reliability of the written exam was found to be 0.854 in the fifth year, which was considered very good. Our students must have at least 60% in the OSCE and 60% in the written exam to pass the course.

The detailed score analysis for the written exam is shown in Table 4; the minimum score was 15.5 and the maximum was 36 (of 40%) for the fifth-year students, with a mean of 29.5, median of 29.75, SD of 3.76, and relative SD of 12.7%.

For the validity of the exam, we compared the results of the OSCE score and the written exam score using Pearson correlation. The correlation was 0.6 for the fifth-year students, which indicated a strong correlation between the OSCE scores and the written exam scores [Figure 2].

Finally, we conducted a factor analysis (with rotated factor), which ensured the components of the OSCE stations were distended and identified the construction of the exam: for the fifth year, it ranged from 0.219 to 0.9; most of the stations were between good and very good [Table 5]. This added to the validity of the exam.

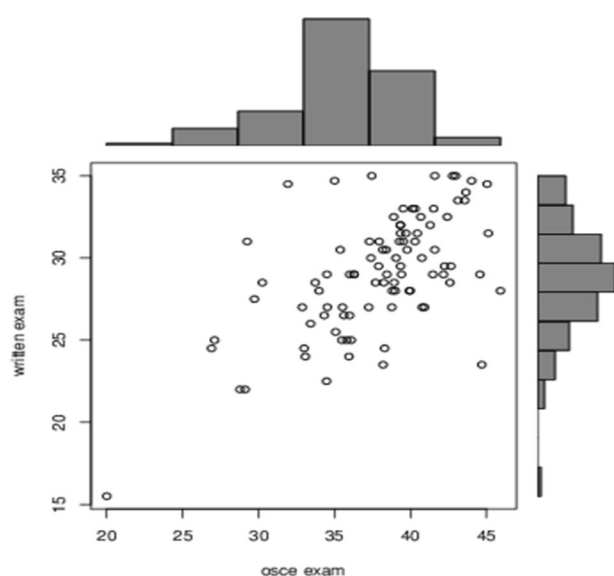
The lowest score was in cardiology and nephrology, which we noticed from the OSCE score. We interviewed the students after the exam and 95% agreed that both systems were difficult because there did not discuss the system during their

rotation secondary to: 1. 90% never present a nephrology case at all because they think it is complicated. 2. 75% never examine the mummer presented in the OSCE. 3. 45% patient refused to be examined.

**Table 4:** Analysis for the fifth-year OSCE score and written exam

Statistical parameters	Results	
	OSCE	Written exam
Minimum	20	15.5
Maximum	45.9	36
Range	25.9	20.5
Count	92	92
Mean	37.8	29.5
Median	38.7	29.75
Mode	39.35	29
Standard deviation	4.56	3.76
Variance	20.8	14.1
Mid-range	32.9	25.75
Quartiles		
Q <sub>1</sub>	35.5	27
Q <sub>2</sub>	38.7	29.7
Q <sub>3</sub>	40.7	32.5
IQR	5.2	5.5
Mean absolute deviation	3.46	2.98
RMS	38.15	29.7
Std. error of mean	0.47	0.39
Skewness	-1.00	-0.64
Kurtosis	4.67	3.73
Coefficient of variation	0.12	0.12
Relative standard deviation	12.0%	12.7%

IQR, interquartile range; RMS, root-mean-square



**Figure 2:** Pearson correlations

**Table 5:** Factor analysis for the fifth-year results

Stations	Factor
V1	0.691
V2	0.59
V3	0.691
V4	0.583
V5	0.7
V6	0.7
V7	0.655
V8	0.521
V9	0.678
V10	0.623
V11	0.723
V12	0.8
V13	0.721
V14	0.655
V15	0.68
V16	0.739
V17	0.3
V18	0.99
V19	0.219
V20	0.982

## Discussion

This is a single departmental and institutional pilot study conducted in 2014 in the Internal Medicine Department at the University of Dammam. First, the number of students who participated in the exam was a good sample, and the reliability of the stations was good as it was the first experience for the department. We used several measures, including Cronbach  $\alpha$ ,

to measure the stability of the stations, and Spearman rank correlation, and the  $R^2$  coefficient determinants to measure internal consistency. These measurements gave more power to the study result and proved the reliability of the OSCE, which was already proved by most of the published studies in many courses. There was an increase in the tools used in reliability measurement, such as the Cronbach  $\alpha$ , the Spearman rank correlation, and the  $R^2$  coefficient. These indicate improvement in both the stability and the internal consistency because of the improvements we were able to make after each exam, such as better orientation of the faculty, avoiding the exam writer errors in the station checklists, students becoming more aware about the exam design, avoiding the mistakes made by the previous examinee, and the most important is that the students were stimulated to read about all systems. In this study, we were able to assess the students in all systems, which was mentioned in their curriculum, so, the exam was fair for both students and the examiner by covering the whole subjects. In the comparison of the OSCE scores and the written scores, it was found that the results were distributed in the normal and left skewness ranges; this indicated that students were performing well and were stimulated to read about most of the topics. The small SD also supported the results. The validity of the exam was strong, and we expect achieving very strong validity in the next few years.

One of the most important findings was that there was a defect in cardiology and nephrology system, which was because of students' perception about nephrology system that must be corrected by encouraging them to present more cases. Patients' refusal is another important point that can be corrected by educating the patients about the students and supervision by faculty with preservation of patients' rights.

The results of this study are encouraging for the other clinical departments at the University of Dammam to use the OSCE in the future. The use of the OSCE could aid the internal medicine departments at other colleges improve the course curriculum and bedside teaching, especially because internal medicine is multisystem and students must be skillful and mastering all systems to be a competent physician.

In comparison to other published studies that reported the implementation of the OSCE, this is the only report that introduces this exam in an internal medicine course in details; most of the published reports focused on the reliability and validity of the exam, feedback, and gender differences, whereas a few discussed the defect of their courses and curriculum, which is one of the most important points for undergraduate students and part of a university's mission and vision. The OSCE can be considered as a teaching tool.<sup>[30]</sup> This study showed improvement in conducting the OSCE through experience, which was reflected by the improvement in the reliability indexes after each exam. This improvement occurred over a short period compared to other published reports. A final point worth noting is that although the exam occurred on different days, this did not affect the validity of the exam, a result that few studies have reported.<sup>[25]</sup>

In the future, more studies must be conducted at a variety of medical universities to improve the curriculum and course development to graduate competent physicians.

### Limitations

First, this is a single department and institute study, which involved only fifth-year medical students who agreed to the new examination format. The students in their final year did not participate due to the potential stress and lack of awareness with the approach of the exam. Second, the examiners were not the same for the period of the exam due to their commitments with clinics and inpatient services. Third, the topic of management was omitted from the exam even though it is included in the curriculum. Finally, the distribution of students was dependent on their registration in the university, which resulted in the difference in the number of students enrolled for each course.

### Conclusion

The OSCE is the gold standard for student assessment and more reliable and valid than the traditional style of exam (long/short cases). It is fair to both students and faculty. The OSCE can stimulate students to read more than the old exam style and, eventually, it will replace the old style worldwide.

Similar studies should be conducted in all clinical departments and other medical schools to further understand the strengths and weaknesses of the exam style and to identify the courses needing improvement. Such research can lead to competent physicians and future consultants.

### Acknowledgments

We would like to acknowledge the university, internal medicine department, including our chair, Dr. Waleed Albaker, who supports the idea of replacing the long/short cases exam with the OSCE, faculty members, specialists, residents, Mr. Zee Shan, and the medical students who were interested to participate in the OSCE exam.

### References

- Simon SR, Bui A, Day S, Berti D, Volkan K. The relationship between second-year medical students' OSCE scores and USMLE Step 2 scores. *J Eval Clin Pract* 2007;13(6):901–5.
- Dong T, Saguil A, Artino AR Jr, Gilliland WR, Waechter DM, Lopreaito J, et al. Relationship between OSCE scores and other typical medical school performance indicators: a 5-year cohort study *Mil Med* 2012;177(9 Suppl) 44–6.
- Dong T, Swygert KA, Durning SJ, Saguil A, Gilliland WR, Cruess D, et al. Validity evidence for medical school OSCEs: associations with USMLE® step assessments. *Teach Learn Med* 2014; 26(4):379–86.
- Hofmeister M, Lockyer J, Crutcher R. The multiple mini-interview for selection of international medical graduates into family medicine residency education. *Med Educ* 2009;43(6):573–9.
- Tombleson P, Fox RA, Dacre JA. Defining the content for the objective structured clinical examination component of the professional and linguistic assessments board examination: development of a blueprint. *Med Educ* 2000;34(7):566–72.
- Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Med Educ* 1979;13(1):41–54.
- Carraccio C, Englander R. The objective structured clinical examination: a step in the direction of competency-based evaluation. *Arch Pediatr Adolesc Med* 2000;154(7):736–41.
- Khan KZ, Gaunt K, Ramachandran S, Pushkar P. The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part II: organisation & administration. *Med Teach* 2013;35(9): e1447–63.
- Fraser R. Does observation add to the validity of the long case? [Author response] *Med Educ* 2001;35:1132–3.
- Meadow R. The structured exam has taken over. *BMJ* 1998; 317:1329.
- Flexner A. Medical Education in the United States and Canada Bethesda, MA: Science and Health Publications 1910.
- Norcini JJ. Does observation add to the validity of the long case? [Letter] *Med Educ* 2001;35:1131–3.
- Norcini JJ. The death of the long case? *BMJ* 2002;324:408–9.
- van der Vleuten CPM. Making the best of the "long case" *Lancet* 1996;347:704–5.
- Patrício MF, Julião M, Fareleira F, Carneiro AV. Is the OSCE a feasible tool to assess competencies in undergraduate medical education? *Med Teach* 2013;35(6):503–14.
- Brannick MT, Erol-Korkmaz HT, Prewett M. A systematic review of the reliability of objective structured clinical examination scores. *Med Educ* 2011;45(12):1181–9.
- Roberts C, Newble D, Jolly B, Reed M, Hampton K. Assuring the quality of high-stakes undergraduate assessments of clinical competence. *Med Teach* 2006;28:535–43.
- Downing SM. *Validity: Establishing Meaning for Assessment Data through Scientific Evidence* London: St George's Advanced Assessment Course 2010.
- Ben-David MF. Life beyond OSCE. *Med Teach* 2003;25(3):239–40.
- Hodges B. Validity and the OSCE. *Med Teach* 2003;25:250–4.
- Auewarakul C, Downing S, Praditsuan R, Jaturatamrong U. Item analysis to improve reliability for an internal medicine undergraduate OSCE. *Adv Health Sci Educ Theory Pract* 2005; 10:105–13.
- Eberhard L, Hassel A, Bäumer A, Becker J, Beck-Muotter J, Bömicke W, et al. Analysis of quality and feasibility of an objective structured clinical examination (OSCE) in preclinical dental education. *Eur J Dent Educ* 2011;15:172–8.
- Iramaneerat C, Yudkowsky R, Myford CM, Downing SM. 2008. Quality control of an OSCE using generalizability theory and many-faceted Rasch measurement. *Adv Health Sci Educ Theory Pract* 2008;13:479–93.
- Lawson D. Applying generalizability theory to high stakes objective structured clinical examinations in a naturalistic environment. *J Manipulative Physiol Ther* 2006;29:463–7.
- Schoonheim-Klein M, Muijtjens A, Habets L, Manogue M, Van der Vleuten C, Hoogstraten J, et al. On the reliability of a dental OSCE, using SEM: effect of different days. *Eur J Dent Educ* 2008;12:131–7.

26. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ* 2011;2:53–5.
27. Pell G, Fuller R, Homer M, Roberts T, International Association for Medical Education. How to measure the quality of the OSCE: a review of metrics—AMEE guide no. 49. *Med Teach* 2010; 32(10):802–11. doi: 10.3109/0142159X.2010.507716.
28. Al-Naami MY. Reliability, validity, and feasibility of the Objective Structured Clinical Examination in assessing clinical skills of final year surgical clerkship. *Saudi Med J* 2008;29(12):1802–7.
29. Raheel H, Naeem N. Assessing the Objective Structured Clinical Examination: Saudi family medicine undergraduate medical students' perceptions of the tool. *J Pak Med Assoc* 2013;63(10):1281–4.
30. Brazeau C, Boyd L, Crosson J. Changing an existing OSCE to a teaching tool: the making of a teaching OSCE. *Acad Med* 2002;77(9):932.

**How to cite this article:** Al-Osail AM, Al-Shiekh MH, Al-Said AH, Al-Osail EM, Al-Ghamdi MA, Al-hawas AM, Al-bahussain AS, Al-dajani AA. Objective Structured Clinical Examination as an assessment method for undergraduate medical students. *Int J Med Sci Public Health* 2015;4:192-198

**Source of Support:** Nil, **Conflict of Interest:** None declared.