

Evaluation of performance in different types of Objective Structured Clinical Examination (OSCE) stations among Undergraduate Medical Students

Khandaker Abu Talha^{1*}, Aung Mra¹, Farhana Selina², Tun Aung¹, Challa Venkata Rao¹, Hla Maung¹

¹Department of Surgery, Faculty of Medicine & Health Sciences,
Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia

²Department of Medicine, Faculty of Medicine & Health Sciences,
Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia

*Corresponding Author's email:katalha@yahoo.com

(Received: April 21, 2016; Accepted: December 22, 2016)

ABSTRACT

An Objective Structured Clinical Examination (OSCE) is now a well-recognized modern type of examination often used in faculties of medicine and health sciences all over the world. Though the assessment tool has been designed to assess different types of clinical skill but this is evident that student performance are not the same in different types of stations. The aim of this study was to evaluate the performances in different types of OSCE stations among undergraduate medical students. Three types of stations were set in this cross-sectional study. They were clinical reasoning, history taking and procedure performing. On the examination day all the students had attended 3 stations for procedure, 4 stations for clinical reasoning and 3 stations for history taking. The scores were collected and transferred to excel spreadsheet. Mean score of each types of modules were calculated. Statistical difference between all three means were measured by one-way ANOVA F-test. F was 7.2 and p-value was 0.001304. The result was significant at $p < 0.05$ in two-tailed hypothesis. There was a significant difference in the scores of different types of stations in OSCE. Strongest performance was observed in procedural skill and weakest in clinical reasoning. Performance was moderate in history taking skills.

Key words: Objective Structured Clinical Examination (OSCE), different station, procedural skill, clinical reasoning.

INTRODUCTION:

Several issues exist in the conventional method of assessment that will interfere with the outcome of the examination. Some of these issues are differences in experience of examiners, using different methods of instruction and using forms of assessment that are somewhat ambiguous. In 1975, to overcome these ambiguities in assessment, Harden and colleagues introduced a system of clinical assessment called Objective Structured Clinical Examination (OSCE)¹.

An Objective Structured Clinical Examination (OSCE) is now a well-recognized modern type of examination. The purpose of this assessment is to test clinical skill performance and competence in skills such as history taking, clinical examination, communication and clinical procedures².

An OSCE comprises a circuit of stations (the timing is 5–10 minutes although some use up to 15 minute), in which each candidate is examined on a one-to-one basis by one or two impartial examiner(s) where either real or simulated patients (actors or electronic patient simulators) are used. Each station has a different examiner. Candidates moving through different stations, completing all the stations within a specified amount of time. In this way, all candidates go through the same stations. It is considered to be an improvement because the stations can be standardized enabling fairer peer comparison and complex procedures can be assessed without endangering patient's health³.

The OSCE is designed to be objective and structured to assess clinical skills. Objectivity is achieved because all candidates are assessed using exactly the same stations with the same marking scheme. In an OSCE, candidates get marks for each step that they perform correctly in accordance with the standardized marking scheme provided, which therefore makes the assessment of clinical skills more objective, rather than subjective.

All stations in OSCEs are structured for a very specific task. Where simulated patients are used, detailed scripts and training are provided to ensure that the information that they give is the same to all candidates, including the emotions that the patient should use during the consultation or history taking. Careful instructions are provided to ensure that the candidate understands that he or she has to complete a very specific task. The OSCE is carefully structured to include parts from all elements of the curriculum as well as a wide range of skills.

The OSCE is designed in such a way that the candidate can apply clinical and theoretical knowledge. The theoretical knowledge is required for answering questions from the examiner at the end of the station. The questions are standardized and the candidate is only asked questions that are on the marking sheet.

Scoring in OSCEs is done by the assessor or examiner. Occasionally written stations, for example, writing a prescription chart, are used and these are marked like written examinations, again usually using a standardized marking sheet. One of the ways an OSCE is made objective is by having a detailed marking scheme and standard set of questions⁴.

The objectivity of the OSCE rests on the standardization of the task and the scoring checklist for the stations. However, some of stations are used different types of scoring system. The global rating scale score, given by expert examiners, showed higher inner-station reliability and better predictive validity than did the scores using checklist⁵.

Though the assessment tool has been designed to assess different types of clinical skill but this is evident that student performance are not the same in different types of stations. The aim of this study was to evaluate if there is any significant difference in performance in different types of OSCE stations in undergraduate medical students. Three types of stations were set in this study. They were clinical reasoning, history taking and procedure performing. Null and alternate hypotheses were set as follows:

Null hypothesis (H₀): There is no significant difference in performance in different types of OSCE stations in undergraduate medical students.

Alternate hypothesis: There is significant difference in performance in different types of OSCE stations in undergraduate medical students.

MATERIALS AND METHODS

This was a cross-sectional descriptive study. All the undergraduate students of the faculty of medicine and health sciences were the study population. All the year 5 students who had participated in the senior surgical rotation end examination were the sample. The sample size was 36. After preparing the study protocol it was submitted for the ethical committee and permission was achieved (UMS/FPSK6.9/100-6/1/95). On the examination day all the students had attended 10 OSCE stations. There were 3 stations for procedure, 4 stations for clinical reasoning and 3 stations for history taking. The scores were collected and transferred to excel spreadsheet. Mean score of each types of modules were calculated. Statistical difference between all three means were measured by one-way ANOVA F-test. P-value was calculated to accept / reject null hypothesis in 5% confidence interval.

RESULTS

The highest mean (8.61) was achieved in procedures stations followed by history taking (8.1) and clinical reasoning (7.3). The median of the marks of different types of stations also follow the same chronology. The median of procedure, history taking and clinical reasoning stations were 9, 8 and 7.5 respectively. The mode of the procedure and history taking stations were 9 when the mode of the clinical reasoning station was 8 (Figure 1).

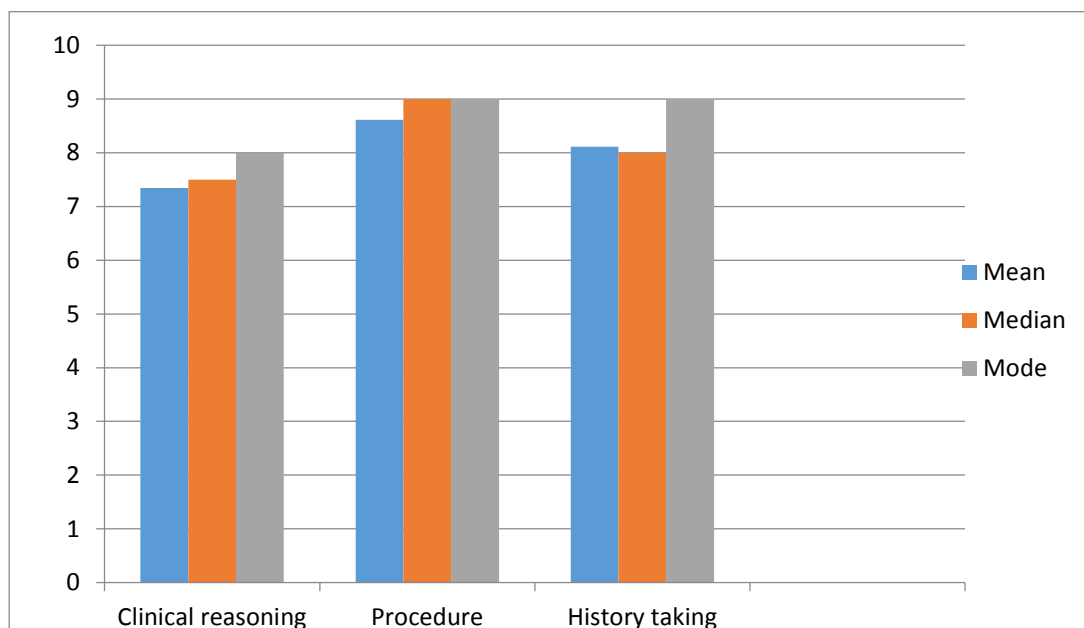


Figure 1: Mean, Median and Mode of different types of OSCE stations

Clinical reasoning stations had highest (2.31) variance followed by history taking (1.3) and procedure (0.72). Standard deviation (SD) also follows the same chronology. SD of clinical reasoning, history taking and procedure were 1.52, 1.1 and 0.82 respectively (Table 1).

Table 1: Variance and Standard Deviation (SD) of different types of OSCE stations.

	Variance	SD
Clinical reasoning	2.31	1.52
Procedure	0.72	0.85
History taking	1.3	1.1

Sum of squares (SS) was 21.25, degree of freedom (Df) was 2 and mean square (MS) was 10.6 in between-groups. On the other hand SS, Df and MS were respectively 109.69, 75 and 1.46 in within-groups. F was 7.2 and p-value was 0.001304. The result was significant at $p < 0.05$ in two-tailed hypothesis. So null hypothesis was rejected and alternate hypothesis was accepted. There was a significant difference in the scores of different types of stations in OSCE (Table 2).

Table 2: F and p-value of three different means by one-way ANOVA

Source	SS (sum of squares)	Df (degree of freedom)	MS(mean square)	F	P-value
Between-groups	21.2564	2	10.6282	7.26683	0.001304
Within-groups	109.6923	75	1.4626		
Total	130.9487	77			

DISCUSSION

Sim JH et al (2015)⁶ carried out similar study in 2015 on 6 different types of OSCE stations and evaluated the result by ANOVA test. Different types of the stations in their study were history taking, physical examination, communication skill, professionalism, procedural skill and clinical reasoning. They found students weakest in the clinical reasoning performance and strongest in the procedural skill with an F value of 2.980. In this study also strongest performance was observed in procedural skill and weakest in clinical reasoning. Performance was moderate in history taking skills.

Newble and Swanson (1988)⁷ calculated the reliability of the scoring of the examiners of same OSCE stations. Relationship was measure by Pearson correlations between examiner pairs and it was found that those varied, both within and between stations. Pearson correlation for 5 minute physical examination

stations were – 0.33, 0.45, 0.64 and 0.76; for 10 minutes physical stations were – 0.72, 0.72, 0.75, 0.82 and 0.89. In the patient education stations – 0.38, 0.39, 0.55 and 0.75; procedural skill stations – 0.48, 0.52, 0.79 and 0.91. In this study only one examiner was present for each station, so correlation between the examiner's scoring was not possible. Strongest performance was found in procedural skill with smallest standard deviation.

Barman A published his paper in 2005⁸. This was a critique paper based on the previous findings of the researchers from 1975 to 2004. This study concluded that the reliability, objectivity and validity are reasonably up to the mark though these parameters significantly depend on the construction of the stations, number of the stations and method of the scoring. The main draw-back of OSCE was mentioned as it was very much resource specific. In this study there were 10 stations with checklist for scoring. Procedures were assessed on standard manikins.

Jahan F et al. (2011)⁹ compared the assessments done by examiners and students in OSCE examination. Students were provided with self-assessing score sheet while examiners also assessed the students in the same time. There was a positive correlation in the history taking stations but significant difference in the procedural stations. In the current study student's assessment was not asked. Variance was highest in clinical reasoning and lowest in procedure stations.

Chirayu Auewaraku et al. (2005)¹⁰ performed study on the item analysis efficiency in OSCE by calculating Generalizability (G) coefficients score. Their study was on OSCE stations in undergraduate medical school. There were 7 physical examination stations, 6 history taking stations, 11 procedural skills and 1 counselling station. By item analyzing and computing, the correct item-total correlation was done for each station. Problem stations were identified and G score was calculated before and after deleting the scores of problem stations. Performance was best in the procedural skill stations and worst in the history taking stations. In this study problem stations were not detected. There was no counselling station but strongest performance was the same like the study of Chirayu Auewarakul et al¹⁰.

The performance in different OSCE stations usually always differ from station to station. This study also showed similarity with international studies. Students' performance was best in procedural skill station in most of the studies. The reliability and validity of the OSCE examination was not assessed in this study which could be the next step of the study.

CONCLUSION

In this study students showed significantly best performance in the procedural skill OSCE stations. Average score of history taking stations was less than those of the procedural stations. The average score of the clinical reasoning stations was least among others.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

REFERENCES

1. Harden RM, Stevenson M, Downie WW, Wilson GM. (1975) Assessment of clinical competence using objective structured examination. *Br Med J* 22; 1(5955):447-51.
2. Ross, M., Carroll, G., Knight, J., Chamberlain, M., Fothergill-Bourbonnais, F., and Linton, J. (1988) Using the OSCE to measure clinical skills performance in nursing. *Journal of Advanced Nursing* 13, 45-56.
3. Leichner P, Sisler GC, Harper D. (1984). A study of the reliability of the clinical oral examination in psychiatry. *Can J Psychiatry* 29(5):394-397.
4. Marliyya Zayyan. (2011). Objective Structured Clinical Examination: The Assessment of Choice. *Oman Med J* 26(4): 219–222.
5. Townsend AH, McIlvenny S, Miller CJ, Dunn EV. (2001). 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* 35(9):841-6.
6. Sim JH, Abdul Aziz YF, Mansor A, Vijayanathan A, Foong CC, Vadivelu J. (2015) Students' performance in the different clinical skills assessed in OSCE: what does it reveal? *Med Educ Online*: 18;20:26185.

7. Newble DI, Swanson DB. (1988) Psychometric characteristics of the objective structured clinical examination. *Med Educ* 22(4):325-34.
8. Barman A. (2005) Critiques on the Objective Structured Clinical Examination. *Ann Acad Med Singapore* 34(8):478-82.
9. Jahan F, Sadaf S, Bhanji S, Naeem N, Qureshi R. (2011) Clinical skills assessment: comparison of student and examiner assessment in an objective structured clinical examination. *Educ Health (Abingdon)* 24(2):421.
10. Chirayu A, Steven M. D, Rungnirand P, Uapong J. (2005) Item Analysis to Improve Reliability for an Internal Medicine Undergraduate OSCE. *Advances in Health Sciences Education* 10:105–113.