

## Objective Structured Clinical Examinations (OSCEs) as assessment tools for medical student examination

Farooq Ahmed Khan

Birmingham and Solihull Mental Health NHS Foundation Trust

Email: farooqkhandr@gmail.com

---

### Abstract

The Objective Structured Clinical Examinations (OSCEs) are one of the most common methods of assessment for medical students. OSCEs have been used in medical schools for the purposes of formative and summative assessment. In the United Kingdom (UK) the term 'assessment' is used to refer to the judgment of student's work and 'evaluation' refers to judgment of courses or course delivery and the process of making such judgments. OSCEs have psychometric principles grounded in them like norm-referenced cut-off scores predictive of later examination failures, snapshot performance at one point and clear guidance and criteria for performance outcome. There has been a change in the assessment process in medical education because of advancement of biomedical science and healthcare system though in eighties and nineties biomedical knowledge was assessed by written examinations, faculty ratings were used to judge clinical competency and national examinations were used for promotion to the next level of training<sup>[1]</sup>. In determining whether OSCEs would deliver a meaningful outcome for assessments some researchers<sup>[1]</sup> have argued that OSCEs are restrictive, non-discriminative and simplistic. A study<sup>[14]</sup> found that students undergoing similar courses score higher when assessed by OSCEs compared to standard question and answer scenario. It has been suggested that studying and reporting of institution detail, student information, role of examination, number of circuits, number of sites, number of examinations, recording and scoring system, pass/fail criteria, station duration, OSCE duration, number and detail of examiners, simulated patients and observers, validity, reliability and feasibility are significant in synthesizing the data and studies.

**Key words:** OSCE, Examination, Medical student, Assessment.

---

### Introduction

The Objective Structured Clinical Examinations (OSCEs) are one of the most common methods of assessment for medical students. OSCEs have been used in medical schools for the purposes of formative and summative assessment. There are always discussions that interpretation of these assessments must be standardized, flawless marking system must be adopted in OSCEs, any bias and gender variations in OSCEs must be avoided as much as possible, also the distinction between the OSCEs and real patient scenarios.

### Objective Structured Clinical Examination (OSCE) in medical schools

Traditionally medical school examinations were based on long case, short case discussions and viva voce as opposed to the current system of OSCEs and this change from old to new system of examinations was studied in 2003<sup>[1]</sup>. OSCEs have been introduced as assessment tools since mid-70s and have evolved since<sup>[2]</sup>. OSCEs are performance based examinations in which trainees are examined and evaluated as they go through a series of stations where they interview, examine and treat standardized patients presenting with some problems<sup>[3]</sup> which could include scenarios like 'breaking bad news', 'explaining treatment options to patients' and 'history taking from relatives and patients'. Taras<sup>[4]</sup> (2005) suggests that in the United Kingdom (UK) the term 'assessment' is used to refer to the judgment of student's work and 'evaluation' refers

to judgment of courses or course delivery and the process of making such judgments. OSCEs have psychometric principles grounded in them like norm-referenced cut-off scores predictive of later examination failures, snapshot performance at one point and clear guidance and criteria for performance outcome. Since their adoption as performance based assessment tools OSCEs have been modified and extended beyond the traditional bedside and outpatient clinical skills like being used in examination and assessment scenarios<sup>[5]</sup>. However as with many other assessment methods OSCE is not perfect and each medical school must determine how to best use them to generate competent physicians<sup>[5]</sup>. The reliability of OSCEs as a part of summative assessment was examined by a study<sup>[6]</sup> at the University of Iowa College of Medicine among year three medical students and they found that aggregated case scores from low stakes OSCEs within clerkships could reliably indicate which students were performing poorly. Medical schools can use OSCE case scores collected over a clinical year for summative evaluation<sup>[6]</sup>. Since the beginning of the OSCEs there have been new dimensions and innovations developed around the use of OSCEs; the Faculty of Health Sciences of a local University organized TOSCE (Team OSCE) in which medical students, students in nursing, occupational therapy, chaplaincy and social work participated in acting/ role playing and examining and the overall feedback at the end of 16 months was that the post station learning feedback was quite useful and 79-100% 'agree' or 'strongly agree' that TOSCE model

was feasible<sup>[7]</sup>. The TOSCE holds promise for learners at all levels for a variety of clinical scenarios where both health care content and team-based skills are necessary. The Inter-professional Education for Collaborative Patient Centred Practice (IEPCPC) raises challenges of defining the competencies necessary for teamwork, methods of how to teach and then evaluate them and this was the reason TOSCE was instigated<sup>[7]</sup>.

There has also been some work<sup>[8]</sup> done about examiner fatigue and its influence on the marking and they found no discernible pattern of systematic or unsystematic bias on part of examiner over an OSCE of two hours in a communication skills station.

### OSCE as formative or summative assessment

Townsend, McIlvenny, Miller & Dunn<sup>[9]</sup> (2001) suggested that an OSCE is useful for formative assessment, formative OSCE identifies skills learning deficiencies which can be corrected, an undergraduate general practice attachment upgrades student's skills and performance in a general practice attachment predicts student's final medical school performance. Formative assessment refers to assessment that is specifically intended to generate feedback on performance to improve and accelerate learning<sup>[10]</sup>. Summative assessments are defined as a judgment, which encapsulates all the evidence up to a given point, and this point is seen as finality at the point of the judgment<sup>[4]</sup>. As part of their medical student training they are posted in various departments of medical and surgical specialties to get wider clinical exposure. At the end of these placements the students are examined by OSCEs and case presentations as a form of formative assessment in which they will have to present cases they have seen and will be marked on them. Though intensive for medical students these assessments are not part of their final summative assessment and the pass or fail depends on their final examinations. There has been a change in the assessment process in medical education because of advancement of biomedical science and healthcare system though in eighties and nineties biomedical knowledge was assessed by written examinations, faculty ratings were used to judge clinical competency and national examinations were used for promotion to the next level of training<sup>[11]</sup>. Another study<sup>[9]</sup> concluded in their research on the general practice attachment candidates and final medical student examinations that problem solving and focused physical examination skills need to be targeted by clinical departments as there is lack of improvement in these areas. Arguments have been put forward by a study<sup>[12]</sup> suggested that in higher education, formative assessment and feedback should be used to empower students as self-regulated learners. At the end of placement as stated above medical students are assessed either on the OSCE or case presentations and they form part of their formative assessment. Taras<sup>[4]</sup> (2005), has also argued against the

separation of formative and summative assessments and suggests that they are learning steps. Taras<sup>[4]</sup> (2005) also suggests that recognition of summative assessments as central and necessary to all assessments must stop demonizing the assessment process for validation and certification and hence should be considered as stepping-stones for learning.

### Meaningful interpretation of assessments

A study<sup>[13]</sup> reported that in family medicine residency oral examinations the examiners differ significantly in their degree of severity in outcomes of the exams. Weingarten<sup>[13]</sup> et al., (2000) also suggested that candidates should be presented with a balanced panel of examiners and some degree of standardization of content must be introduced in oral examinations. In determining whether OSCEs would deliver a meaningful outcome for assessments some researchers<sup>[1]</sup> have argued that OSCEs are restrictive, non-discriminative and simplistic. A study<sup>[14]</sup> found that students undergoing similar courses score higher when assessed by OSCEs compared to standard question and answer scenario. They<sup>[14]</sup> also concluded that interactive teaching, adult education principles, opportunities for discussion, provision of feedback, and stimulation of self learning were rated more highly inclusion of OSCE scenarios. There are low correlations across stations suggesting little consistency of performance in different skills, despite the fact that reliability on the same station is high<sup>[15]</sup>. A study<sup>[16]</sup> exploring the medical students' attitudes towards communication skills using Communications Skills Attitude Scale (CSAS) found the use of these type of scales will explore the relationship between the attitudes of medical students and their demographic and education related characteristics. Ben-David<sup>[17]</sup> (2003) in an Editorial describes that life beyond OSCE will be partially determined by success in overcoming current challenges in performance test construction but Ben-David also cautioned embarking on new innovative approaches to performance assessment while still focusing on unresolved problems. They<sup>[17]</sup> suggested that there is a need first to lay the foundation before we can build more layers while discussing new approaches to assessments. It was also highlighted<sup>[17]</sup> that two very important steps were not yet fully developed in assessment procedures; one is the method of setting standards in performance assessment and the other is the clarification of the meaning of validity in performance assessment.

### The process of marking in OSCEs

The study<sup>[18]</sup> at the Leeds Medical School, UK studied the need of assessor training for the performance based examinations like OSCEs and suggested that setting up of pass mark or passing standard is contentious. They<sup>[18]</sup> also discussed that in OSCEs there are a number of different assessors and

the practical aspects of the stations can't be duplicated at all times for all the students thus raising the questions on robustness of the comparative grading mechanisms. In the above study<sup>[18]</sup> studied 207 medical students taking part in OSCEs and 108 assessors; they reported that assessors trained in OSCEs have a higher standard of passing compared to untrained assessors. They also reported that in their study if all assessors would have been trained an additional three candidates out of 207 would have failed to meet the passing criterion and this suggests the need for training the examiners and fairness in the process of OSCEs. The number of points on the scale can vary as long as there is a cohort of examinees labeled as borderline. The marking system in the OSCEs is different for different medical schools throughout the world; but there are two popular methods for marking OSCEs, the Modified Borderline Group Method (BGM) and Borderline Regression Method (BRM). A study<sup>[19]</sup> found various marking methods in depth. They emphasized that The Medical Council of Canada (MCC) and the University of Ottawa use six point scales with adjective descriptors corresponding to inferior, poor, borderline unsatisfactory, borderline satisfactory, good, and excellent. This study<sup>[19]</sup> also explains that to determine a cut score for a station, the mean checklist score for the cohort of examinees rated as borderline is calculated and then applied to all examinees. By averaging the checklist scores of the Borderline Satisfactory and Borderline Unsatisfactory groups, it is assumed that this corresponds to an examinee exactly at the pass/fail cut point between the two categories. The sum of the station cut scores becomes the cut score for the overall exam. The other method described by Wood<sup>[19]</sup> et al (2006) is Borderline Regression Method (BRM) this is very similar to the BGM but rather than selecting out a cohort of borderline examinees and calculating their mean checklist score, this method regresses all of the examinees' checklist scores onto their global ratings to produce a linear equation. By inserting the midpoint of the global rating scale corresponding to the borderline group(s) into the equation, a corresponding predicted checklist score can be determined. This predicted score becomes the cut score for the station.

### **Bias, gender influence and other variations in OSCEs**

In a study<sup>[20]</sup> it was found that in four UK medical schools among year three medical students; the authors demonstrated that pass scores awarded locally in one medical school cannot be reliably compared with another; simulated patient's performance in OSCEs will affect examiner marking; medical students in different medical schools perform differently because of different curricula and examiner training affects the marking criteria. A study<sup>[20]</sup> looked at gender role in OSCEs in Newcastle University Medical School. The author of this study examined 140 year three medical

students to see whether the gender of the student or the simulated patient will have any impact on the outcome of the examination for stations like lymphatic system and back examination. This study<sup>[21]</sup> found no statistically significant difference between simulated patient gender and student versus simulated patient gender for the back examination. But Beaini<sup>[21]</sup> (2009) found that more students failed on the lymphatic system examination of male patients which was paradoxical as the expectation was that more would fail on examination of the back but authors discussed that anatomically examination of back will not make any difference for male or female, students learnt to examine lymphatic system with breast examination getting more familiar with this aspect of females and students were more prudent with harder stations. Another study<sup>[22]</sup> looking into similar gender factors among final year medical students found that there is a significant relationship between gender and performance for some stations but correlation for multiple comparisons removed this significance. A study<sup>[23]</sup> found that in a two-day workshop for motivational interviewing the trainees achieved the same competence whether they practiced with the simulated patient or a fellow trainee. The OSCEs are quite complex and extensive in their remit and there has been an explosion of its use since its inception in main stream examinations but rarely has its validity, reliability and feasibility been tested<sup>[24]</sup>. Patricio<sup>[24]</sup> et al (2009) also suggested specific checklist to report data on validity, reliability and feasibility of OSCEs. In the checklist produced in this study<sup>[24]</sup> it has been suggested that studying and reporting of institution detail, student information, role of examination, number of circuits, number of sites, number of examinations, recording and scoring system, pass/fail criteria, station duration, OSCE duration, number and detail of examiners, simulated patients and observers, validity, reliability and feasibility are significant in synthesizing the data and studies.

### **OSCE comparison to real patient scenario**

A systematic review<sup>[25]</sup> looking into the role of patients in medical education found that patients have played a role as teacher, assessment and curriculum developers. The need of their training to be influential in the education system was also discussed by Jha<sup>[25]</sup> et al (2009). They found limited evidence for long-term effectiveness of patient involvement and ethics, psychological impact and influence on educational policy were poorly explored in the previous research<sup>[25]</sup>. Brigden & Dangerfield<sup>[26]</sup> (2008) argued that using real patients for 'practice' could be seen as unethical though they agreed that some of the clinical skills like communications, blood pressure (non-invasive, non-risky skills) could be learnt through real patients. Another study<sup>[27]</sup> (2004) suggests that simulation allows trainees to learn by their mistakes, because 'mistakes

can be allowed to proceed to their natural conclusion', this would be unlikely in clinical practice and unethical as well. In a study<sup>[1]</sup> at the University of Bristol Medical School 30 students appeared for both type of long case and OSCE examination; 20 out of 30 returned the questionnaire. Probert,<sup>[1]</sup> et al., (2003) also asked 72 consultants to fill in a questionnaire for assessment and 60 responded; traditional final examinations were inversely associated with consultant assessments and OSCEs were more consistent and showed positive association with consultant rating across the board. They<sup>[1]</sup> also highlight the fact that use of the long cases had been critical part of the assessment of clinical skills for a longtime but there are several weaknesses in its use, including reproducibility, being time consuming and inter-examiner and inter-patient variability. There were also concerns raised by a study<sup>[25]</sup> which highlighted in the systematic review regarding issues with using real patients in medical training; they were impact of reduced admission rates on learning opportunities within hospital settings shorter hospital stays and the predominance of acutely ill patients in hospitals, lack of co-operation by patients due discomfort or lack of interest, the ethical implications of involving really ill patients in medical teaching and difficulties in evaluating student performance in an uncontrolled environment.

### Conclusions

OSCEs have psychometric principles grounded in it and since their adoption as performance based assessment tools they have been used in various settings for the purposes of outcome-based assessments<sup>[5]</sup>. OSCEs are used appropriately as formative assessment but it has also been argued that they are not part of summative assessment for medical students. It had been argued by Taras in a study<sup>[4]</sup> against the separation of formative and summative assessments on the whole and suggested that they could be taken as learning steps which would reduce the demonization of summative assessments. There have also been variations in marking methods in various medical schools making it hard to compare and contrast outcomes of OSCEs. The other variations in OSCEs include timings and places of examinations and methods of examinations. OSCEs have a significant role in medical education, they have been well tested and validated and they are much better compared and contrasted against the real patients. Though there are advantages with OSCEs; some shortcomings can't be ignored that the medical students will not have the first hand experience of examining on the real patient in the examinations which would be more realistic than a standardized patient. Innovations in OSCEs have made them more validated and robust but still there are some variations in how different medical schools conduct their own OSCEs.

### References

1. Probert, C. S., Cahill, D. J., McCann, C. L. & Ben-Shlomo, Y. (2003). Traditional finals and OSCEs in predicting consultant and self-reported clinical skills of PRHOs: a pilot study. *Medical Education*;37:597-602.
2. Harden, R. M. C. G., Stevenson, M., Downie, W. W. & Wilson, G. M. (1975). Clinical competence in using objective structured examination. *British Medical Journal*, 1,447-451.
3. OSCE questions: College of Osteopathic Medicine, Ohio University (2011). Available at: <http://www.oucom.ohiou.edu/academicaffairs/predoc/osce/index.htm>
4. Taras, M. (2005). ASSESSMENT – SUMMATIVE AND FORMATIVE SOME THEORETICAL REFLECTIONS. *British Journal of Educational Studies*, Vol.53, No.4, 466-478.
5. Mavis, B. V. & Henry, R. C. (2002). Between a rock and a hard place: finding a place for the OSCE in medical education. *Medical Education*;36:408-409.
6. Bergus, G. R. & Kreiter, C. D. (2007). The reliability of summative judgments based on objective structured clinical examination cases distributed across the clinical year. *Medical Education*;41:661-666.
7. Marshall, D., Hall, P. & Taniguchi, A. (2008). Team OSCEs: evaluation methodology or educational encounter? *Medical Education*.42:1111-1146.
8. Humphris, G. M. & Kaney, S. (2001). Examiner fatigue in communication skills objective structured clinical examinations. *Medical Education*,35:444-449.
9. Townsend, A. H., McIlvenny, S., Miller, C. J. & Dunn, E. V. (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. *Medical Education*, 35:841-846.
10. Sadler, D. R. (1998). Formative assessment: revisiting the territory, *Assessment in Education*, 5(1),77-84.
11. Whitcomb, M. E, Anderson, M. B. (1999). Transformation of medical students' education: Work in progress and continuing challenges. *Academic Med*;74:1076-9.
12. Nicol, D. J. & MacFarlane, D. (2006). Formative assessment and selfregulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education* Vol. 31, No. 2, 199-218.
13. Weingarten, M. A., Polliack, M. R., Tabenkin, H. & Kahan, E. (2000). Variations among examiners in family medicine residency board oral examinations. *Medical Education*, 34:13-17.
14. Ali, J., Adam, R. U., Josa, D. et al. (1999). Comparison of performance of interns completing the old (1993) and new interactive (1997) Advanced Trauma Life Support courses. *Journal of Trauma*; 46(1):80-6.
15. Roberts, J. & Norman, G (1990). Reliability and learning from the objective structured clinical examination. *Medical Education*; 24(3):219-23.
16. Rees, C., Sheard, C. & Davies, S. (2002). The development of a scale to measure medical students' attitudes towards communication skills learning: the Communication Skills Attitude Scale (CSAS). *Medical Education*;36:141-147.
17. Ben-David, M. F. (2003). Life beyond OSCE. *Medical Teacher*, Vol. 25, No. 3, pp. 239-240.
18. Pell, G., Homer, M. S. & Roberts, T. E. (2008). Assessor training: its effects on criterion-based assessment in a medical context. *International Journal of Research & Method in Education* Vol. 31, No. 2, 143-154.

19. Wood, T. J., Humphry-Murto, S. M., & Norman, G. R. (2006). Standard Setting in a Small Scale OSCE: A Comparison of the Modified Borderline-Group Method and the Borderline Regression Method. *Advances in Health Sciences Education*, 11:115–122.
20. Chessar, A., Cameron, H., Evans, P., Cleland, J., Bourcicot, K. & Mires, G. (2009). Sources of variation in performance on a shared OSCE station across four UK medical schools, *Medical Education*, 43:526–532.
21. Beaini, A. (2009). Does student or patient gender in OSCEs really make a difference? *THE CLINICAL TEACHER*;6:127–133.
22. Wiskin, C. M. D., Allan, T. F. & Skelton, J. R. (2004). Gender as a variable in the assessment of final year degree-level communication skills, *Medical Education*;38:129–137.
23. Lane, C. L., Hood, K. & Rollnick, S. (2008). Teaching motivational interviewing: using roleplay is as effective as using simulated patients. *Medical Education* 2008;42:637–644.
24. Patricio, M., Julia, M., Farelira, F., Young, M., Norman, G. & Carneiro, A. V. (2009). A comprehensive checklist for reporting the use of OSCEs. *Medical Teacher*,31:112–124.
25. Jha, V., Quinton, N. D., Bekker, H. L. & Roberts, T. E. (2009). Strategies and interventions for the involvement of real patients in medical education: a systematic review. *Medical Education*,43:10–20.
26. Brigden, D. & Dangerfield, P. (2008). The role of simulation in medical education. *The Clinical Teacher*;5:167–170.
27. Kneebone, R. L., Scott, W., Darzi, A. & Horrocks, M. (2004). Simulation and clinical practice: strengthening the relationship. *Medical Education*,38:1095–1102.

**How to cite this article:** Khan FA. Objective Structured Clinical Examinations (OSCEs) as assessment tools for medical student examination. *Telangana Journal of Psychiatry* 2016;2(1):17-21.