

OSCE Assessment for Medical Student Clinical Skills and Performance, Are We Doing It The Right Way?

Dr. Bandar Hetaimish MD, MED, FRCS (C)¹, Dr. Hussein Elbadawi MD, CCD²

¹Assistant Professor, Department of Orthopedic Surgery, Medical College, Taibah University, Madinah, Saudi Arabia
Endocrinologist, Internal Medicine Department, Wayne State University, Detroit, Michigan, USA.

²Endocrinologist, Internal Medicine Department, Wayne State University, Detroit, Michigan, USA.
Ministry of Education, Saudi Arabia

Abstract: *The objective structured clinical examination (OSCE) is an assessment method that can be used to evaluate different learning domains such as cognitive, psychomotor and affective domains. It assesses the competence of students in a structured and systematic format that minimize subjectivity. This article reviews the literature supporting the use of OSCE in accordance with validation, reliability, feasibility, acceptability and educational impact criteria for good assessment. Moreover, it explores the major strengths and weaknesses of using the OSCE as a measurement tool when assessing medical student's clinical skills and performance. In addition, it proposes solutions to help overcome the identified weaknesses and barriers.*

Keywords: OSCE, Assessment, Clinical Skills and Performance

1. Background

The medical education field is rapidly growing, and with recent advances in the assessment and evaluation of medical students, the majority of medical schools utilize the objective structured clinical examination (OSCE) in evaluating medical students performance. The report by Harden et al. was the first to describe the OSCE structure and organization. Subsequently, many medical schools embraced this method, and it represents a major tool of assessment and evaluation¹. OSCE is a method that can be used to evaluate different learning domains. These include cognitive, psychomotor and affective domains².

Miller et al. developed a clinical assessment framework (Fig-1) that could be applied in evaluating medical students

in four major areas (knowledge, competence, performance, and action). The base of this pyramid is built by knowledge and is usually assessed by objective testing. On the other hand, the students must know how to use and analyze this knowledge, which constitute the second level of Miller's pyramid (know how). Competency based assessment is one of the methods introduced to objectively evaluate the "knows how" category in this framework. The next level in Miller's pyramid is to evaluate performance (shows how). The OSCE falls within this level, where it is a useful tool to explore the ability of the candidate to "show how" to perform certain skills in a simulated environment. The top part of the pyramid includes an assessment of what physicians do in their practice³.

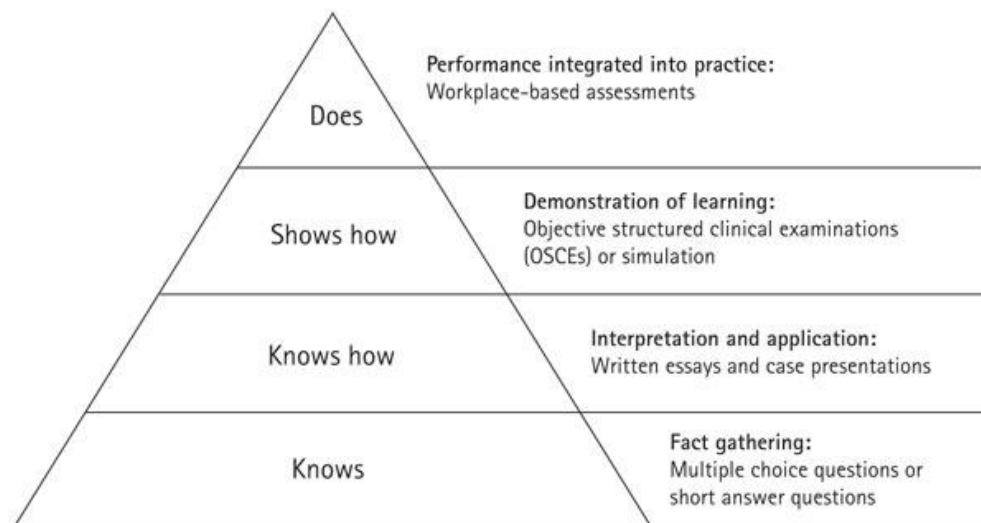


Figure 1: Miller's Assessment Pyramid (1990).³

The OSCE assesses the competence of students in a structured and systematic format that minimize subjectivity⁴.

Owing to its objectivity, this method is gaining popularity and being utilized not only in medical schools but also in the

post-graduate residency training programs⁵. Moreover, Multiple licensure examinations and medical schools incorporated OSCE in their assessment process. This includes OSCE as an assessment tool in high stake royal college examination, assessment of performance for graduation exam for medical students, certification licensure for practicing physicians and formative feedback assessment tool for medical trainee and students⁶. This method has evolved to incorporate different modalities of assessment. For instance, the introduction of standardized patients, observer ratings, written tests, physical examination, and communication skills is widely accepted^{7,8}. The ability to incorporate different modalities represents a major strength that surpasses other methods.

Norcini and colleagues outlined the criteria for good assessment that include: validity, reliability, equivalence, feasibility, educational effect, catalytic effect, and acceptability. However, it is challenging to develop an assessment tool that encompasses all the above criteria. In this seminal paper, the OSCE was classified under the first category, which implies that it is supported with an existing evidence background⁹.

2. Objectives of this paper

The objectives of this paper are:

- a) To review and summarize the literature supporting the use of OSCE in accordance with validation, reliability, feasibility, acceptability and educational impact criteria for good assessment.
- b) Identify and discuss the major strengths and weaknesses of using the OSCE as a measurement tool when assessing medical students clinical skills and performance.
- c) Propose solutions to help overcome the identified weaknesses and barriers.

3. Methods

Two reviewers (BH, HB) performed online search in PubMed engine for studies discussing OSCE in assessment of medical students clinical skills and performance.

Limitations of Traditional Assessment Methods

The traditional methods of evaluating the clinical skills and performance of medical students were based on the long case discussion at the bed side or short non-standardized patients examination. These methods of assessment failed to represent the student's ability to perform a specific task (show how category of Miller's pyramid). Moreover, these methods do not allow medical students to compensate for underperformance in a specific category¹⁰. An important prerequisite is the equality in the level of questioning that allows fair judgment when assessing medical students. The structure of questions when using traditional methods is difficult to standardize; which may violate the equality concept.

The traditional methods scoring system is tampered by huge potential for rater bias, in which the lack of controlled rating scoring system could result in variation in judgment between raters.

The content specificity is one of the major drawbacks of the traditional exams. It does not allow adequate sampling of the curriculum contents, which threaten the comprehensiveness of the exam, and limit the assessor's ability to make an integrated judgment¹¹.

The OSCE was designed to overcome the shortage in the traditional methods of assessments. It focuses on the standardization of scoring, structure and blueprinting of the exam; it overcomes the content specificity, minimizes rater and patient performance biases. One of the major goals of the OSCE is to isolate the variability in student marking to the student's variability itself avoiding any other factor that could lead to unrealistic discrimination in the student's assessment¹.

4. OSCE and Criteria of Good Assessment

1. Validation Process (Validity)

The old perspective of *validation process* have changed over time to focus more on the implication of the results of an assessment tool rather than the ability of the tool to measure what it's supposed to measure. The different levels of validation include content validity, criterion validity and construct validity. Kane and colleagues argue that all types of validity belong to the construct validity of an assessment tool¹². Harden et al. pointed out that validity of the OSCE came from its ability to measure the components of clinical competences such as history taking and physical examination performed by medical students¹.

The OSCE was associated with high validity when evaluating postgraduate internal medicine residents. Internal medicine resident's performance, plans, diagnosis and OSCE scores correlated with the level of training. Second year residents scored higher than first year residents, this observation supports construct validity of OSCE. Moreover, the result of the OSCE was compared to the result of an internationally accepted test the American board of internal medicine; and again was associated with high concurrent validity⁵. The content validity of the OSCE is well documented in the literature. This level of validation process came through the blueprinting and building up of solid framework to consolidate the structure and exam contents^{1,13}.

Auewarakulet al. identified multiple sources to validate the OSCE. The evidence of validity was derived from the exam content, internal structure, response process and interaction with different factors through the OSCE process. In this paper, the OSCE exam was associated with high validity in assessing medical student during their clinical sessions. These observations support validation process of the OSCE in the assessment of medical student clinical skills and performance¹⁴. The results of another study by Martin and colleagues showed that OSCE can be associated with high predictive validity. In this study, poor performance of the third year medical students was predictive of poor performance on their fourth and fifth year clinical exams¹⁵.

The concurrent and construct validity are more evident when using global rating scale rather than checklist in OSCE for experts compared to novices¹⁶.

2. Reliability

Reliability is the degree of consistency and reproducibility of a measurement tool in discriminating between the subjects of interest. This could be achieved utilizing time factor, test format or sampling process¹⁷. There are many factors that can enhance the reliability of the OSCE. These include increasing the number of stations, standardizing the scoring system rubrics, standardizing patients' performance in the exam, and using trained assessors during the OSCE⁶. Other factors that can also reduce the reliability of the OSCE including: candidates and examiners fatigue, anxiety and fear from the examination method, and reduced candidates' attention span. Even after controlling for these factors the OSCE reliability remains widely variable 0.41 to 0.88¹⁸.

In a systematic review of 39 studies by Brannick et al. the reliability of the OSCE using the alpha value was 0.78 (95% CI 0.73–0.82). Moreover, the reliability of clinical skills assessment by OSCE was higher compared to communication skills assessment¹⁹. Bergus and colleagues used the generalisability and decision theories to show that the cumulative scores of multiple OSCEs can determine students with poor performance²⁰.

Given the aforementioned limitations, we propose the following solutions to help overcome these barriers and eventually improve the reliability. To help reduce candidates' and examiners' fatigue, incorporating short break-stations during the exam may allow students to rest and formulate their thoughts. Moreover, it may enhance the rater's ability to focus on the scoring parameters. Another proposed solution is to divide the candidates and raters into multiple groups that undergo the OSCE on different times or days, this will help reduce rater's drift and consequently improve the reliability. However, the downside is feasibility of embracing this approach.

Moreover, preparing OSCE courses that aim to familiarize students with the process may reduce students' anxiety, and increase raters' knowledge about the OSCE. Future studies are required in order to affirm the value of implementing these solutions.

3. Feasibility & acceptability

Feasibility is defined as the practicality of using the OSCE as an assessment tool. While acceptability represents the degree of usefulness of the OSCE results to the stakeholders⁶. Although studies have shown that OSCE is a feasible and acceptable assessment tool; few barriers are worth discussing. OSCE is more expensive and requires longer duration than traditional methods of assessment²¹. Moreover, the ability to recruit standardized patients, need for training of examiners, and the availability of instruments and tools are important obstacles to feasibility¹⁸. The inherent flexibility of the OSCE is one of the key elements that enhance feasibility. This is apparent by variable number of candidates, cases and time required for the evaluation²².

Multiple methods are available to enhance the acceptability of the OSCE to medical students and faculties. For instance, the planning and organization process are one of the crucial steps for the success of OSCE. The set up duration for an OSCE may take up to 6 months. This period includes

different stages to develop the OSCE structure. In the paper by Harden et al. four stages are described. First, Advance planning of the OSCE. Second stage involves early preparation on the day of the exam. Third stage includes activities during the day of the examination. Final stage is the feedback and scoring that takes place after the exam¹. The development of OSCE requires leadership, manpower, assignment of the roles, and team assembly. Setting up the exam schedule and blueprinting is a crucial step to enhance the learning outcomes and clarify the contents of the assessment process²³.

Harmony between OSCE stations and domains that need to be assessed is vital for the success of the OSCE. In addition, the scoring sheets rubric need to be prepared and piloted before the OSCE. The team leader has to assemble the OSCE organizational committee before the exam. This will facilitate assigning responsibilities and allows brainstorming within committee members. Preparatory workshops are a valuable opportunity for the organizational committee and the faculty to clarify any ambiguous areas and to generate constructive solutions^{24,25}.

In our opinion, piloting the OSCE in low-stake small groups could be of great advantage. Piloting may facilitate the transition from traditional exam methods to OSCE, eventually enhancing its acceptability. In addition, it helps to highlight potential difficulties that could emerge in future exams. It will also allow estimation of cost, preparatory time and the number of personnel required for the upcoming exam.

4. Educational impact and catalytic effect

The OSCE provides positive impact on the learning environment and student's education. Students report that OSCE scenarios highly resemble real life cases. It is the OSCE's proximity to the reality that impacts future learning of the candidates⁶. Complete dependence on checklist scoring system in OSCE may decrease the educational impact, as students focus more on filling the checklist to pass the exam³. In our opinion, having a mixed evaluation method will help to overcome this negative effect of using checklists. The proposed method that I would recommend is to combine checklist-scoring sheet with a global rating scale during the assessment process. This combination will increase the candidates' awareness on the importance of deep understanding of major domains, rather than memorizing the components of the checklist. Moreover, this method expands the level of assessment to accommodate novices and experts candidates' level of thinking. The process of problem solving largely differs between experts and novices. Experts have more focused and efficient gathering of information, while novices have prolonged and detailed scheme of thinking. In a study by Hodges and colleagues the performance of experts on the checklist part of the OSCE was worse than medical clerks. However, when using the global rating scales, experts were more efficient in problem solving than medical clerks¹⁶.

The educational impact of OSCE on medical students' clinical skills can be enhanced through applying multiple small group and feedback augmented OSCE²⁶. The catalytic effect of OSCE was highlighted in a nine years follow up

that shows improvement in students' performance, faculty teaching and curriculum outcomes²⁷. A recent systematic review suggested that OSCE has a positive effect on steering students' learning and faculty teaching²².

5. Strengths and Weaknesses of the OSCE

To summarize the earlier discussion, the OSCE has many advantages that include: multiple sampling of different medical knowledge and skills, controlling the difficulty of the exam cases according to the skill-level of the student, and clearly setting the knowledge, skills, and attitudes to be assessed²⁸. The OSCE is a patient-focused exam that gives attention to student patient interaction. It reminds medical students with their responsibilities as future physicians²⁹. Moreover, OSCE allows the evaluation of large number of medical students in a short time. The OSCE is associated with high reliability, validation process, acceptability and educational impact on medical students²¹.

Despite all the advantages of the OSCE, there are some limitations that are worth further discussion. OSCE is a time consuming process, associated with high cost, and requires manpower³⁰. Possible solutions include: assembling OSCE committee with independent academic funding. This will increase the financial resources and ensure appropriate financial planning. More efforts to train standardized patients and examiners will ensure high quality of the OSCE and will provide manpower for future exams. Eventually, the process will be less time consuming and more efficient.

In high stake exams, focusing on using OSCE purely as a summative assessment tool without formative feedback may burden its educational benefit²⁶. Our proposed solution is to increase the awareness of the exam providers and emphasize the importance of giving summarized feedback.

The performance of the students in the OSCE may differ from their performance in the workplace even on identical tasks. This is because the OSCE is a simulated environment that aims to mimic the original workplace, however, it is not identical³¹. Using high fidelity simulation is a good solution to this problem. Although it is expensive, the value of high fidelity simulation need to be tested in future studies.

Furthermore, we suggest collecting feedbacks from the candidates after each OSCE. These post OSCE reflections will identify exam gaps and students difficulties. It will help exploring further solutions to overcome any future obstacles and will enhance the validity of the OSCE. The feedbacks can be written (such as questioner or comments) or verbal (such as small group discussion).

Another potential weakness is missing student's response by the assessor during the exam. The rapid sequence of this assessment can explain why this could happen. We suggest running videotape recording in each OSCE station to keep track of the exam process without disruption. Moreover, assignment of an audio-video technician who is responsible for collection of the recordings for future review will help in resolving these issues.

Additionally, The OSCE has a limited role in the assessment of the knowledge part of medical students. The combination of different assessment tools such as multiple choice questions or short answer questions can empower the OSCE to overcome this limitation. Harden and Gleeson raised up issues of OSCE compartmentalization of clinical knowledge and skills²⁸. The evaluation of the sum of different domains will not always be equal to evaluation of the whole. In our vision, students' skills assessment through multiple areas will be more resourceful than implementing a holistic approach. We suggest integrating long OSCE stations where some procedures will get evaluated in detailed fashion. While, complex procedures divided into smaller domains and assessed in shorter OSCE stations. This combined approach will help reducing the compartmentalization effect.

Finally, to answer the paper's question, after reviewing the literature we believe that the OSCE is a powerful assessment tool for evaluating medical students clinical skills and performance. Although we are in the right track, we believe that there is more room for improvement. The proposed solutions in this paper may help improving this assessment method. We are in need of future studies that accommodate these alternatives to define the effect on the OSCE improvement.

6. Disclosure

No part of the article was presented in conference proceedings

References

- [1] Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. *British medical journal* 1975;1:447-51.
- [2] Bloom BS, Krathwohl DR, Masia BB. *Taxonomy of educational objectives: The classification of educational goals handbook 1: Cognitive domain* [S.l.]; New York: 1974 Longman.
- [3] Miller GE. The assessment of clinical skills/competence/performance. *Academic medicine : journal of the Association of American Medical Colleges* 1990;65:S63-7.
- [4] Harden RM. What is an OSCE? *Medical teacher* 1988;10:19-22.
- [5] Petrusa ER, Blackwell TA, Ainsworth MA. Reliability and validity of an objective structured clinical examination for assessing the clinical performance of residents. *Archives of internal medicine* 1990;150:573-7.
- [6] Khan KZ, Ramachandran S, Gaunt K, Pushkar P. The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part I: an historical and theoretical perspective. *Medical teacher* 2013;35:e1437-46.
- [7] Howley LD. Performance assessment in medical education: where we've been and where we're going. *Evaluation & the health professions* 2004;27:285-303.
- [8] Swanson DB, van der Vleuten CP. Assessment of clinical skills with standardized patients: state of the art

- revisited. *Teaching and learning in medicine* 2013;25 Suppl 1:S17-25.
- [9] Norcini J, Anderson B, Bollela V, et al. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. *Medical teacher* 2011;33:206-14.
- [10] Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *Lancet* 2001;357:945-9.
- [11] Norcini JJ. The death of the long case? *BMJ* 2002;324:408-9.
- [12] Kane MT. Current Concerns in Validity Theory. *Journal of Educational Measurement* 2001;38:38.
- [13] Newble DI, Hoare J, Elmslie RG. The validity and reliability of a new examination of the clinical competence of medical students. *Medical education* 1981;15:46-52.
- [14] Auewarakul C, Downing SM, Jaturatamrong U, Praditsuwan R. Sources of validity evidence for an internal medicine student evaluation system: an evaluative study of assessment methods. *Medical education* 2005;39:276-83.
- [15] Martin IG, Jolly B. Predictive validity and estimated cut score of an objective structured clinical examination (OSCE) used as an assessment of clinical skills at the end of the first clinical year. *Medical education* 2002;36:418-25.
- [16] Hodges B, Regehr G, McNaughton N, Tiberius R, Hanson M. OSCE checklists do not capture increasing levels of expertise. *Academic medicine : journal of the Association of American Medical Colleges* 1999;74:1129-34.
- [17] Buckendahl CW, Plake BS. *Handbook of test development*. Mahwah, NJ: Lawrence Erlbaum Associates, 2006:725-38.
- [18] Turner JL, Dankoski ME. Objective structured clinical exams: a critical review. *Family medicine* 2008;40:574-8.
- [19] Brannick MT, Erol-Korkmaz HT, Prewett M. A systematic review of the reliability of objective structured clinical examination scores. *Medical education* 2011;45:1181-9.
- [20] Bergus GR, Kreiter CD. The reliability of summative judgements based on objective structured clinical examination cases distributed across the clinical year. *Medical education* 2007;41:661-6.
- [21] Barman A. Critiques on the Objective Structured Clinical Examination. *Annals of the Academy of Medicine, Singapore* 2005;34:478-82.
- [22] Patricio MF, Juliao M, Fareleira F, Carneiro AV. Is the OSCE a feasible tool to assess competencies in undergraduate medical education? *Medical teacher* 2013;35:503-14.
- [23] Vargas AL, Boulet JR, Errichetti A, van Zanten M, Lopez MJ, Reta AM. Developing performance-based medical school assessment programs in resource-limited environments. *Medical teacher* 2007;29:192-8.
- [24] Cusimano MD, Cohen R, Tucker W, Murnaghan J, Kodama R, Reznick R. A comparative analysis of the costs of administration of an OSCE (objective structured clinical examination). *Academic medicine : journal of the Association of American Medical Colleges* 1994;69:571-6.
- [25] Whelan GP. Educational Commission for Foreign Medical Graduates: clinical skills assessment prototype. *Medical teacher* 1999;21:156-60.
- [26] Mavis BE, Henry RC, Ogle KS, Hoppe RB. The emperor's new clothes: the OSCE reassessed. *Academic medicine : journal of the Association of American Medical Colleges* 1996;71:447-53.
- [27] Duerson MC, Romrell LJ, Stevens CB. Impacting faculty teaching and student performance: nine years' experience with the Objective Structured Clinical Examination. *Teaching and learning in medicine* 2000;12:176-82.
- [28] Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Medical education* 1979;13:41-54.
- [29] Barrows HS. An overview of the uses of standardized patients for teaching and evaluating clinical skills. *AAMC. Academic medicine : journal of the Association of American Medical Colleges* 1993;68:443-51; discussion 51-3.
- [30] Vallevand A, Violato C. A predictive and construct validity study of a high-stakes objective clinical examination for assessing the clinical competence of international medical graduates. *Teaching and learning in medicine* 2012;24:168-76.
- [31] Khan K, Ramachandran S. Conceptual framework for performance assessment: competency, competence and performance in the context of assessments in healthcare-deciphering the terminology. *Medical teacher* 2012;34:920-8.