

Enhancing Consistency and Competence: Standardizing OSCE Assessments within MD Programs

Gilchrist, C., Bismilla, Z.

Introduction

Objective structured clinical examinations (OSCEs) are a longstanding cornerstone in the assessment of clinical competencies in medical education (Harden & Gleeson, 1979). These assessments play a crucial role in evaluating the practical and interpersonal skills of medical students, ensuring they are prepared for real-world clinical practice (Swanson & van der Vleuten, 2013). Within one large Canadian MD Program, a review of OSCE assessments was found to have inconsistencies that slowly developed over years of programmatic divergence. Part of what cultivated this divergence was that OSCE assessments were administered under different portfolios for Pre-clerkship (Years 1 and 2) and Clerkship (Years 3 and 4) years. Over time this segregation has allowed variations in station design, operational procedures and assessment criteria to occur. These discrepancies hinder the ability of OSCEs to accurately evaluate student progress and uniformly assess across different stages of training (Holmboe *et al.*, 2010; Epstein, 2007). This commentary reflects on the process of OSCE standardization that has occurred in our program and provides recommendations for other medical schools to consider when reviewing their own clinical assessments.

Temerty Faculty of Medicine, University of Toronto, Toronto, Ontario Canada

Corresponding author: Dr. Chris Gilchrist
Email: Christopher.Gilchrist@THP.ca

OSCEs Prior to Standardization

In Canada, MD programs are currently offered at 18 different Universities across 8 provinces. Program pre-requisites vary from school to school however all schools require completion of a previous University degree (e.g. Bachelor of Science, Arts, Engineering etc.) Programs are typically four years in duration (or 3 years without breaks). Most follow a common structure in which the first half focuses on pre-clinical education ("pre-clerkship"), typically covering foundational sciences. This is followed by a clinical phase where students rotate through core clinical specialties and participate in patient care in hospital and community settings ("clerkship"). The University of Toronto's 4 year MD Program is among the largest in Canada, training over 1,000 students across a network of affiliated teaching hospitals.

Before standardization, the OSCE structure showed inconsistencies across the MD program. Pre-clerkship OSCEs featured 10-minute stations without post-encounter probes (PEPs) and used a combination of checklists and global rating scales (GRSs). Conversely, the clerkship OSCEs had 8-minute stations with 2-minutes for PEPs and exclusively used GRSs. Additionally, each year had distinct GRSs which led to further confusion and increased cognitive load for examiners, reducing the reliability and validity of the assessments (Epstein, 2007).

Students struggled with these differences, feeling unprepared due to the yearly variations in OSCE design and assessment criteria. The lack of consistency forced them to adapt with a new OSCE approach annually, potentially



© SEAJME. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

South-East Asian Journal of Medical Education
Vol.19, no.1, 2025

impacting their performance. Examiners faced similar challenges, needing separate faculty development for each year's version of the OSCE which further complicated obtaining accurate student assessments. Ultimately, this divergence led to reduced validity and reliability of one of the MD program's key assessments and marker of graduation readiness.

We developed a new model for OSCEs within the MD Program which has led to improved student and faculty experience, as well as increased assessment validity and reliability of our OSCEs. Based on experience, we propose 5 recommendations for MD programs developing or reviewing OSCE assessments. (Table 1).

Table 1: OSCE Standardization Recommendations

Recommendation	Description
Recommendation 1	Standardize the OSCE leadership
Recommendation 2	Develop a unified OSCE framework across the MD program
Recommendation 3	Integrate entrustable professional activities (EPAs)
Recommendation 4	Standardize faculty development for OSCE examiners
Recommendation 5	Engage Standardize Patient Program leadership in OSCE development

Standardize the OSCE Leadership

The need for standardization of the OSCE assessments within an MD program is clear. There are notable benefit to students, faculty and the program. However, before beginning to standardize the OSCE assessments, it is important to consider whether the leadership itself should be standardized. Although there is always a role for committee consultation to ensure that broader programmatic goals are being considered for a given assessment, having one leader responsible for the entire OSCE alignment process has been essential to our success. This ensured that a cohesive strategy could be developed and that considerations impacting all years of study can uniformly be addressed throughout the process. As such, it is recommended that before the structural design of OSCEs is reassessed, the responsibility of that process should fall to a single leader for all OSCEs within an MD program.

Develop a unified OSCE framework across the MD program

This recommendation comprises two main components. We define the "framework" of an OSCE as consisting of both the logistical structure of an OSCE (i.e. station timing,

formatting of materials, use of signals, etc.) and the assessment tools of an OSCE (what the examiners use to assess student performance). First, we recommend that the structure of OSCEs should be consistent across all years. Such consistency allows students to become familiar with the examination style from Year 1, reducing cognitive load and improving the ability of the OSCE to accurately capture performance by their final year (Epstein, 2007). Examiners similarly benefit from understanding the consistent expectations of the OSCEs, leading to more valid assessments of student performance (Swanson and Vleuten, 2013).

Second, the assessment tools used in OSCEs, such as checklists, PEPs, and GRSs should be standardized. Overly detailed checklists should be minimized (Pugh *et al.*, 2006) to focus on key clinical features, avoiding box-ticking and promoting holistic assessment (Epstein, 2007; Pugh *et al.*, 2006). Standardized PEPs should progressively evaluate higher-order skills throughout the program, ensuring a coherent framework for the development of clinical reasoning skills (Swanson & van der Vleuten, 2013; Ten Cate *et al.*, 2015). Unified GRSs should be applied consistently across all years, ensuring that difficulty is scaled by the complexity of the

clinical scenarios rather than by changing assessment scale criteria. This approach helps both students and examiners understand and meet consistent expectations, enhancing the overall reliability and validity of OSCEs (Frank *et al.*, 2015; Ten Cate & Scheele, 2007).

Integrate entrustable professional activities

Previously, the sequential challenge of an OSCE arose mainly from the inherent complexity of each year's courses. Standardization has afforded the opportunity to ensure OSCEs build in difficulty through intentional design and allow for assessment of progressively attained clinical skills across all program years. Entrustable Professional Activities (EPAs) for MD training in Canada have recently been developed by the Association of Faculties of Medicine of Canada structure, the multiple raters and varied case scenarios can further enhance the reliability of OSCE scores (Suneja, 2024).

The use of EPAs to help set the clinical challenge of an OSCE aligns well with standardized GRSs. Instead of using different GRSs (or any other assessment criteria change) to scale difficulty at higher years, students should instead face an increased challenge based on the scenario encountered (as determined by an EPA and their associated contextual variation). For instance, in Year 1, a student may take a basic history from a pleasant patient (*EPA1 – Obtain a history and perform a physical examination adapted to the patient's clinical situation*) (AFMC, 2016). While in Year 3, a more complex history may be taken from an acutely ill patient (*EPA1 and EPA8 – Recognize a patient requiring urgent or emergent care, provide initial management and seek help*) (AFMC, 2016). In both cases, communication is graded consistently via the same GRS, but task complexity varies based on EPAs. Thus it will be more challenging for the Year 3 student to meet standard on the communication GRS because the scenario is inherently more difficult to interact with and not because the standard has changed for the Year 3 OSCE. However, the process of integrating EPAs into the curriculum requires aligning learning

(AFMC, 2016) (Table 2). EPAs are key clinical tasks that trainees must perform independently by graduation, and they bridge theoretical knowledge with practical application, making them ideal for structuring OSCEs (Ten Cate, *et al.*, 2015). EPAs should be used to identify expected clinical skills at each level of training and guide assessment task complexity (Ten Cate *et al.*, 2015). Integrating EPAs into the assessment ensures that OSCEs reflect the expectations for learners to accomplish progressively complex clinical tasks as they advance throughout medical school (Ten Cate & Scheele, 2007; Holmboe *et al.*, 2010). By aligning OSCEs with EPAs, we can provide a structured framework that mirrors real-world competencies required for medical practice (Touchie & Ten Cate, 2016). Through the EPA implementation within an OSCE assessment

objectives with specific EPAs and educating both students and faculty. This is no small task, but the effort can significantly benefit the standardization and utility of the OSCE as an assessment tool.

Standardize faculty development for OSCE examiners

This recommendation emphasizes the crucial role of examiners in delivering robust clinical assessments. In our large program, any given OSCE may require up to 100 examiners and the previous differences in each year's OSCE framework necessitated that examiners understand multiple sets of OSCE structures and assessment tools. By aligning OSCE frameworks and assessment tools across all years, this challenge has been mitigated. Standardized faculty development can train examiners in a unified manner, reducing variability in scoring and enhancing the overall quality of assessments (Holmboe *et al.*, 2010; Epstein, 2007). The Office of Faculty Development at our institution has extensive experience and expertise in developing and delivering effective faculty development sessions to a wide range of faculty. Enlisting the help of such a specialized unit when embarking on faculty development of this scale is essential.

Table 2: Association of Faculties of Medicine of Canada EPAs

EPA	Description
EPA 1	Obtain a history and perform a physical examination adapted to the patient's clinical situation
EPA 2	Formulate and justify a prioritized differential diagnosis
EPA 3	Formulate an initial plan of investigation based on the diagnostic hypotheses
EPA 4	Interpret and communicate results of common diagnostic and screening tests
EPA 5	Formulate, communicate and implement management plans
EPA 6	Present oral and written reports that document a clinical encounter
EPA 7	Provide and receive the handover in transitions of care
EPA 8	Recognize a patient requiring urgent or emergent care, provide initial management and seek help
EPA 9	Communicate in difficult situations
EPA 10	Contribute to a culture of safety and improvement
EPA 11	Perform general procedures of a physician
EPA 12	Educate patients on disease management, health promotion and preventive medicine

To achieve standardization, faculty development should focus on two major issues: the tools of the OSCE assessment and EPAs. Examiners need to understand how to apply checklists of key features and appropriately grade PEPs. Perhaps most critically, examiners must be trained to consistently apply the criteria for each GRS level across different contexts and student interactions. Calibration exercises, where examiners rate sample performances and discuss evaluations, can help illustrate how GRSs work and improve consistency (Holmboe *et al.*, 2010; Epstein, 2007). Addressing EPAs is likely more challenging, as EPAs are a relatively new concept (AFMC, 2016; Ten Cate, 2013). Foundational work will be necessary, but most of the creation and alignment of EPAs will occur at the leadership level during OSCE station design. This effort will culminate in reduced cognitive load for examiners during exams. When examiners can clearly understand the logistics of an OSCE and standardized assessment tools are employed uniformly, they can focus on

accurately assessing the student in front of them. This approach should lead to more accurate assessments of student performance, which is the primary goal of any OSCE.

Engage Standardize Patient Program leadership in OSCE development

Engaging the Standardized Patient (SP) Program from the outset in OSCE standardization is essential for ensuring consistency and authenticity in student assessments, particularly in communication as well as to ensure diverse patient representation. Early involvement allows for standardized role portrayals as they relate to communication style (i.e. ensuring uniform responses to open-ended questions), which enhances the reliability of communication skill evaluations. Incorporating the SP Program early also affords the opportunity to increase representation of diverse patient demographics within OSCEs. This step is much easier to accomplish from the beginning of case design and ensures that we are

preparing our students for culturally competent care. Furthermore, early SP Program involvement ensures that specific aspects of cases, such as portraying a patient with aphasia, are confirmed to be possible and depicted uniformly across the SP population. Given that SP standardization is essential to a valid assessment across an entire cohort of students, their early involvement can ensure that role portrayal will be able to accurately achieve the educational goal.

Conclusion

Standardizing OSCE assessments within our large MD program has been essential to further enhance the validity and reliability of our clinical evaluations. Consolidating OSCE administration under a single portfolio eliminates programmatic discrepancies, enabling uniform assessment of student progress. A unified OSCE framework provides consistent structure and assessment tools, helping students and examiners develop proficiency in the assessment. Integrating EPAs into OSCEs represents an effective means to progressively increase the complexity of clinical tasks, preparing students for real-world practice. Standardizing faculty development equips examiners with the skills to apply assessment tools consistently and accurately. SP Program involvement ensure that patient characteristics essential to the clinical case can be accurately portrayed in scenarios. By adopting this standardized approach, we can ensure that our assessments are fair, comprehensive, and reflective of the competencies required for effective medical practice. This will ultimately lead to better-prepared graduates who are equipped with the skills and knowledge necessary to excel in their clinical careers and provide high-quality patient care.

References

Association of Faculties of Medicine of Canada (2016) *AFMC entrustable professional activities for the transition from medical school to residency* [Internet]. Ottawa (ON): AFMC. Available from: https://md.utoronto.ca/sites/default/files/afmc_en_trustable_professional_activities_en.pdf [Accessed 12 August 2024]

- Epstein, R.M. (2007) 'Assessment in medical education', *New England Journal of Medicine*, 356(4), pp. 387-396. doi:10.1056/NEJMr054784.
- Frank, J.R., Snell, L. & Sherbino, J. (2010) *The CanMEDS 2015 Physician Competency Framework*. Ottawa (ON): Royal College of Physicians and Surgeons of Canada. Available from: https://canmeds.royalcollege.ca/uploads/en/framework/CanMEDS%202015%20Framework_EN_Reduced.pdf [Accessed 12 August 2024].
- Harden, R.M. & Gleeson, F.A. (1979) 'Assessment of clinical competence using an objective structured clinical examination (OSCE)', *Medical Education*, 13(1), pp. 41-54. doi:10.1111/j.1365-2923.1979.tb00918.x.
- Holmboe, E.S., Sherbino, J., Long, D.M., Swing, S.R. & Frank, J.R. (2010) 'The role of assessment in competency-based medical education', *Medical Teacher*, 32(8), pp. 676-682. doi:10.3109/0142159X.2010.500704.
- Pugh, D., Halman, S., Desjardins, I., Humphrey-Murto, S. & Wood, T.J. (2016) 'Done or almost done? Improving OSCE checklists to better capture performance in progress tests', *Teaching and Learning in Medicine*, 28(4), pp. 406-414. Available from: <http://myaccess.library.utoronto.ca/login?url=https://www.proquest.com/2Fscholarly-journals%2Fdone-almost-improving-osce-checklists-better%2Fdocview%2F1835367671%2Fse-2%3Faccountid%3D14771>.
- Suneja, M., Hanrahan, K.D., Kreiter, C. & Rowat, J. (2024) 'Psychometric properties of entrustable professional activity-based objective structured clinical examinations during transition from undergraduate to graduate medical education: a generalizability study', *Academic Medicine* [published ahead of print]. doi:10.1097/ACM.00000000000005719.
- Swanson, D.B. & van der Vleuten, C.P. (2013) 'Assessment of clinical skills with standardized patients: state of the art revisited', *Teaching and Learning in Medicine*, 25(Suppl 1). doi:10.1080/10401334.2013.842916.
- Ten Cate, O. (2013) 'Nuts and bolts of entrustable professional activities', *Journal of Graduate Medical Education*, 5(1), pp. 157-158. doi:10.4300/JGME-D-12-00380.1.
- Ten Cate, O. & Scheelche, F. (2007) 'Competency-based postgraduate training: can we bridge the gap between theory and clinical practice?', *Academic Medicine*, 82(6), pp. 542-547. doi:10.1097/ACM.0b013e31805559c7.
- Ten Cate, O., Hart, D., Ankel, F., Busari, J., Englander, R., Glasgow, N., et al. (2015)

'Entrustment decision making in clinical training',
Academic Medicine, 91(2), pp. 191-198.
doi:10.1097/ACM.0000000000001044.

Touchie, C. & Ten Cate, O. (2016) 'The promise, perils, problems, and progress of competency-based medical education', *Medical Education*, 50(1), pp. 93-100. doi:10.1111/medu.12839.