Looking into the future of clinical skills assessment in undergraduate medical education during the COVID-19 Pandemic

Neha Bapatla[1], Stephanie Pearson [1], Sydney Stillman[1], Lauren Fine [1], Kyle Bauckman[1], Vijay Rajput[1]

Corresponding author: Dr Vijay Rajput vrajput@nova.edu
Institution: 1. Nova Southeastern University, Dr. Kiran C. Patel College of Allopathic Medicine
Categories: Assessment, Clinical Skills, Undergraduate/Graduate

Received: 03/12/2020
Published: 11/06/2021

Abstract

In light of the COVID-19 pandemic in 2020, United States Medical Licensure Examinations (USMLE) announced its momentary cancellation of its Step 2 Clinical Skills (CS) Examination. This suspension brought to attention the need to evaluate the current methods of clinical skills assessment. Objectively, this period in medical education marks a time for change and improvement. Although this may seem radical, medical education has been continuously changing over the past few decades. The utilization of long case, short case, and viva voce examinations for clinical skills assessment morphed into using the Objective Structured Clinical Examination (OSCE) and Step 2 CS. While OSCEs and Step 2 CS are currently mainstay assessment methods in medical education, the new challenges that COVID-19 has imposed requires medical educators to improve these methods in order to maintain social distancing guidelines. Special consideration should be made to incorporating modalities such as video conferencing, artificial intelligence, virtual reality, and workplace based assessments. The momentary suspension of medical school activities was clearly unexpected, but it is vital that medical educators continue to improve clinical skills assessment in conjunction with the present times.

Keywords: OSCE; clinical skills; COVID-19; Assessment

Considering the present state of medical education assessment

Clinical skills assessment in medical education has continuously transformed to fit the needs of its time. George Miller developed the hierarchal pyramid of physician behaviors, which has served as an influential model for medical education in the United States. Each level of the pyramid serves to emphasize a specific method in the assessment of medical students (Dauphinee, 1995). At the base of the pyramid is "knows," which reflects the knowledge required of a physician (Dauphinee, 1995). As the pyramid ascends it describes "knows how" as the
appropriate actions taken when being presented with a specific circumstance in an exam, "shows how" as the ability to perform in a particular clinical circumstance, and "does" as a physician's behavior in the real-life clinical setting (Dauphinee, 1995). For example, performance-based testing is based on a candidate’s ability to "show how" (Dauphinee, 1995). Prior to the 1970s, clinical testing within medical education focused on "knows" and "knows how" via the administration of multiple choice questions and patient management problems (Dauphinee, 1995). However, the advent of the OSCE by R.M. Harden in 1979 ushered in a new era in medical education that focused on enhancing clinical skills at the bedside in conjunction with medical knowledge gained in lecture halls (Harden et al., 1975).

Over the following decades, OSCEs have served as the primary method of clinical skills assessment at the medical school level. Since 1979, its format has constantly evolved to address its value as a reliable and valid assessment method. The number of clinical encounters and time allotted for each station has served as two important variables that affect the psychometric properties of this examination. However, despite constantly changing the OSCE format, the range of reliability and validity scores seemed to be too broad. Thus, there was a need for a standardized assessment of clinical skills that could improve the psychometric properties of the OSCE. In the early 2000s, the focus on developing clinical skills in future physicians gained traction nationwide and resulted in the development of Step 2 Clinical Skills (CS) by the USMLE as a solution to standardize testing of clinical skills.

However, due to the public health risks posed by the Coronavirus pandemic, in May 2020 the USMLE board announced its suspension of Step 2 CS for the following twelve to eighteen months. This brief pause allows us to revisit the current method of evaluation of clinical skills and consider how this may change going forward, with special considerations made to creating a "COVID-free" testing environment as an adjunct to the evolving use of technology in modern medicine.

**A brief history of clinical skills assessment**

Prior to the 1970s, medical education was greatly dependent upon assessment of clinical knowledge rather than its application at the bedside (Epstein, 2007). Short case, long case, and viva voce (oral) examinations were commonly used methods to test clinical skills in candidates during this period. Short case assessments required candidates to perform a focused examination on five to six patients while long case assessments allocated thirty to forty-five minutes to allow for the candidate to perform a history and complete physical examination, followed by questions regarding clinical management (Khan et al., 2013). Oral examinations consisted of a time period for candidates to assimilate the given clinical material followed by ten to fifteen minutes of questioning by examiners (Khan et al., 2013). The psychometric properties of these examinations are questionable at best. The viva voce assessments have poor content validity, higher inter-rater variability, and inconsistency in marking, essentially rendering this assessment method as highly unreliable (Tabish, 2008). Due to the relatively unstructured nature of questioning by examiners in each situation, there was an incredible need for the development of an assessment method that would improve the reliability and validity of the previous examinations.

The introduction of OSCEs in the late 1970s served to remedy this problem. Simply defined, OSCEs consist of multiple standardized patient (SP) stations that assess medical student interactions with patient-related medical issues. Student performance is dictated by clinical reasoning skills and problem solving abilities of these clinical scenarios (Courtelle et al., 2008). Harden's first conducted OSCE in 1972 was a 100-minute examination consisting of eighteen testing stations with two rest stations, with each station lasting 4.5 minutes (Khan et al., 2013). The composition of OSCEs has since varied with the number of SP stations and length of time at each station. Importantly, research has consistently demonstrated the need for multiple SP stations for each defined clinical problem (Stillman et al., 1982) due to findings suggesting highly variable reliability scores between 0.20 and 0.95.
(Harasym, Mohtadi and Henningsmoen, 1997). However, other studies have suggested that station length plays a minimal role in reliability (Schuwirth and van der Vleuten, 2003), unless extreme changes in length are implemented. Consequently, research efforts have been focused on identifying factors that increase or decrease reliability (Doig et al., 2000), such as test content, design, and implementation factors (Turner and Dankoski, 2008). For example, the use of checklists in OSCEs was initially implemented in order to minimize inter-rater unreliability; nonetheless, this was refuted by various studies suggesting that the use of global rating scales offers very little influence on reliability (Schuwirth and van der Vleuten, 2003). Although OSCEs were implemented with the consideration of improving reliability and validity of long and short case examinations, there are clear complications that require modern solutions.

The development and controversy of USMLE Step 2 Clinical Skills

Prior to 2004, clinical skills assessments at the national level were not administered to US medical graduates. Such an assessment, the clinical competence assessment (CCA), was only administered to graduates of foreign medical schools (Sutnick et al., 1994). In 1992, the Educational Commission for Foreign Medical Graduates (ECFMG) conducted a series of pilot projects to assess clinical competency of foreign medical graduates (FMG). The CCA consisted of integrated clinical encounters with ten standardized patients, which provided profiles of clinical competencies including data gathering, interviewing and interpersonal skills, diagnosis and management skills, interpretation of laboratory and diagnostic procedures, written communication, and spoken-English proficiency (Sutnick et al., 1994). These scores proved useful to residency directors, and they supplemented these results with scores on written examinations (Sutnick et al., 1994).

Through collaborative efforts of the ECFMG and National Board of Medical Examiners (NBME), the USMLE Step 2 CS was developed to administer an examination to US medical seniors testing competency in clinical skills through a multiple-station, standardized patient modality (Hawkins, 2005). The development of Step 2 CS was largely influenced by the need for promoting public safety through ensuring US trained physicians were appropriately prepared for the clinical setting, similarly to the already established CCA for FMGs. Step 2 accomplishes this through testing of vital physician competencies such as taking medical history, performing physical examinations, effective communication with patients, accurate documentation of findings, and identifying appropriate initial diagnostic studies (Hawkins, 2005). Data interpretation scores provided by this examination proved to provides useful information that predicts clinical performance of physicians in supervised practice (Cuddy et al., 2016).

In 2016, medical students from Massachusetts initiated a movement to end Step 2 CS, which quickly gained traction and the support from the Michigan State Medical Society, Massachusetts Medical Society, and the American Medical Association (AMA) Student section (Elder, 2018). Arguments for cessation of Step 2 CS cite expense, limited accessibility, and questionable psychometric properties (Kashaf, 2017). Indeed, the examination is administered in five major US cities, which adds to the already steep price required to register for it. Additionally, opponents argue that the pass/fail nature of the examination does not provide any value to medical assessment; rather, quantitative results may better assist residency directors in targeting clinical strengths and weaknesses (Nguyen, 2018). Conversely, a survey of residency program directors indicated the high utility of Step 2 CS in screening residency applicants for professionalism, communication skills, and translation of knowledge into clinical skills in practice (Paniagua et al., 2018). Furthermore, there are concerns that without Step 2 CS it would be difficult to ensure a minimum entry standard of clinical competence in international medical graduates (IMGs) (Elder, 2018).

In spite of the surrounding controversy, in May 2020 Step 2 CS was suspended for twelve to eighteen months in accordance with CDC recommendations surrounding the COVID-19 pandemic. With a war raging in the wards across the globe, medical schools have also had to adapt to the new normal and find creative ways to continue
education for students. Although there are hopes that eventually we may contain this novel virus and return to normalcy, it is prudent to consider a new era of medical education that not only incorporates a "COVID-free" environment but may also solve issues of cost and standardization.

Accomplishing a “COVID-free” testing environment

The Centers for Disease Control and Prevention policy suggested social distancing guidelines with the arrival of COVID-19 in early 2020. As a result, there was a major upheaval in medical education. Navigating social distancing in conjunction with assessing clinical skills have proven difficult yet not impossible. From a logistical standpoint, OSCEs may be administered in environments that ensure strict infection control and personal hygiene, no large group gatherings, and social distancing of individuals (Boursicot et al., 2020). Despite the potential for these control measures to mitigate the spread of COVID-19 among examiners, students, and SPs alike, the USMLE Step 2 CS chose to take a more cautious route with its suspension. While COVID-19 may be contained in the near future, it is imperative that we consider altering the methods of assessing clinical skills in order to avoid complete suspension of medical education in similar future scenarios.

Video conferencing has been a readily adapted modality during the pandemic that may be useful as a permanent method for assessment. Using software such as Zoom to conduct a web-OSCE requires access to a reliable Internet connection and personal devices with built-in audio and video capabilities (Major et al., 2020). With the proper technical support and a trained pool of SPs clinical skills such as accurate history taking, communication, and critical reasoning can be assessed virtually (Major et al., 2020). Utilization of web-OSCEs may further serve a purpose in training medical students in telemedicine (Lara et al., 2020), which has grown increasingly important in modern clinical care. Additionally, with the use of video conferencing, faculty may observe students from any geographic location (Cantone et al., 2019), which allows for lower out of pocket costs for the student and easier scheduling of examiners and examinees.

Unfortunately, video conferencing is blatantly unable to assess the performance of an adequate physical examination. Perhaps artificial intelligence (AI) may be a better-suited modality for this. With the use of AI, evaluation can be more objective, fast, and cost-effective (Masters, 2019). Additionally, AI can provide a more extensive individualized feedback of medical student performance (Masters, 2019). Virtual reality (VR) may also be useful in implementation of a COVID-free environment. VR applications allow students to have 360-degree views of real or simulated places. Software could be developed to simulate patient care scenarios and allow students to perform history taking and physical examinations while concurrently mitigating the spread of infection. VR has shown to be an interactive and engaging educational tool that supports knowledge retention and skills acquisition (Sultan et al., 2019).

Recently, the use of workplace-based assessments (WPBA) has gained traction in graduate medical education. With consideration of Miller's pyramid of clinical competence, WPBA targets "does" and enables collecting information of physician performance in the everyday clinical setting through commonly used tools such as direct observation of procedural skills (DOPS), Mini-Clinical Evaluation Exercise (mini-CEX), and case-based discussions (Liu, 2012). WPBA is especially superior to OSCEs because it targets limitations such as the deconstruction of the doctor-patient encounter in favor of performing isolated aspects of the clinical encounter (Liu, 2012). In this aspect, WPBA are considered more authentic than OSCEs when applying a resident’s ability to perform in a real-life clinical setting. It may be advantageous to apply WPBAs to medical student education, especially considering how many institutions have had to halt clinical skills assessment with the COVID-19 pandemic. If precautions were taken to ensure the safety of medical students in the hospital setting in the event of another pandemic, trained physicians could continue clinical skill assessment in students.
The evolution of clinical skills assessment has mirrored the needs of its time. The development of OSCEs in the 1970s was a reflection of the necessity in administering a standardized clinical exam to medical students in order to facilitate the importance in clinical skills from medical knowledge. The implementation of Step 2 CS in 2004 further built on the foundation provided by OSCEs to assess medical students fairly at a national level while also providing residency program directors appropriate information to screen applicants equally from US and international medical schools. Currently, further studies are needed to develop an entirely COVID-free clinical assessment using modalities such as videoconferencing, AI, and VR in order to adapt to the ever-changing needs of medical education. However, necessity may find itself leading to the invention of new, less expensive, safer, and valid teaching and assessment tools.

Take Home Messages

- The suspension of Step 2 CS due to COVID-19 requires evaluation of current methodology of clinical skills assessment in order to prevent future suspension of medical school activities in the face of similar events.
- OSCEs served as a solution to the low reliability and validity of short case, long case, and vive voce assessments.
- Step 2 CS standardized examination of clinical competencies across medical schools.
- Incorporation of AI, VR, videoconferencing, and WPBA should be considered for assessment of clinical skills at the undergraduate medical school level in order to establish a "COVID-free" environment.

Notes On Contributors

Neha Bapatla is a MS-III in the charter class at Dr. Kiran C. Patel College of Allopathic Medicine at Nova Southeastern University. She completed her BS in Biology at the University of Florida and MS in Medical Sciences at Boston University. ORCiD: https://orcid.org/0000-0003-3382-6688

Stephanie Pearson is a MS-III in the charter class at the Dr. Kiran C. Patel College of Allopathic Medicine at Nova Southeastern University. She completed her BA in Biology and Chemistry at Cornell University.

Sydney Stillman is a MS-III in the charter class at the Dr. Kiran C. Patel College of Allopathic Medicine at Nova Southeastern University. She completed her BS in Biomedical Sciences with a minor in Spanish at University of Central Florida and MS in Biomedical Sciences at Tufts University.

Dr. Lauren Fine is an Assistant Professor of Medical Education and a Founding Faculty Member at the Dr. Kiran C. Patel College of Allopathic Medicine at Nova Southeastern University. Her areas of interest are Ethics, Humanism, Communication and Clinical Skills. She also enjoys writing within a variety of genres.

Dr. Kyle Bauckman is an Assistant Professor at Nova Southeastern University, Dr. Kiran C. Patel College of Allopathic Medicine, Davie, FL. ORCiD: https://orcid.org/0000-0003-4046-2392

Dr. Vijay Rajput is Professor and Chair, Department of Medical Education, Nova Southeastern University, Dr. Kiran C. Patel College of Allopathic Medicine, Davie, FL. His area of interest is humanism, Ethics, Curriculum innovations. ORCiD: https://orcid.org/0000-0001-5203-5342
Acknowledgements

None.

Bibliography/References

Boursicot, K., Kemp, S., Ong, T., Wijaya, L., et al. (2020) ‘Conducting a high-stakes OSCE in a COVID-19 environment’, MedEdPublish, 9(1), pp. 54. https://doi.org/10.15694/mep.2020.000054.1.

Cantone, R. E., Palmer, R., Dodson, L. G. and Biagioli, F. E. (2019) ‘Insomnia Telemedicine OSCE (TeleOSCE): A Simulated Standardized Patient Video-Visit Case for Clerkship Students’, MedEdPORTAL: The Journal of Teaching and Learning Resources, 15, pp. 10867. https://doi.org/10.15766/mep_2374-8265.10867.

Courteille, O., Bergin, R., Stockeld, D., Ponzer, S., et al. (2008) ‘The use of a virtual patient case in an OSCE-based exam - A pilot study’, Medical Teacher, 30(3), pp. 66-76. https://doi.org/10.1080/01421590801910216.

Cuddy, M. M., Winward, M. L., Johnston, M. M., Lipner, R.S., et al. (2016) ‘Evaluating Validity Evidence for USMLE Step 2 Clinical Skills Data Gathering and Data Interpretation Scores’, Academic Medicine, 91(1), pp. 133-139. https://doi.org/10.1097/acm.0000000000000908.

Dauphinee, W. D. (1995) ‘Assessing clinical performance: Where do we stand and what might we expect?’, Journal of the American Medical Association, 274(9), pp. 741-743. https://doi.org/10.1001/jama.1995.03530090073025.

Doig, C. J., Harasym, P. H., Fick, G. H. and Baumber, J. S. (2000) ‘The effects of examiner background, station organization, and time of exam on OSCE scores assessing undergraduate medical students’ physical examination skills’, Academic Medicine, 75(10 SUPPL.), pp. 96-98. https://doi.org/10.1097/00001888-200010001-00031.

Ecker, D. J., Milan, F. B., Cassese, T., Farnan, J. M., et al. (2018) ‘Step Up—Not On—The Step 2 Clinical Skills Exam’, Academic Medicine, 93(5), pp. 693-698. https://doi.org/10.1097/acm.0000000000001874.

Elder, A. (2018) ‘The Future of the USMLE Step 2 Clinical Skills Exam’, Academic Medicine: Journal of the Association of American Medical Colleges, 93(11), pp. 1601. https://doi.org/10.1097/ACM.0000000000002408.

Epstein, R. M. (2007) ‘Assessment in medical education’, New England Journal of Medicine, 356(4), pp. 387–396. https://doi.org/10.1056/NEJMra054784.

Harasym, P. H., Mohtadi, N. G., Henningsmoen, H. (1997) ‘The Use of Critical Stations to Determine Clinical Competency in a "High Stakes" OSCE.’ In: Scherpbier A.J.J.A., van der Vleuten C.P.M., Rethans J.J., van der Steeg A.F.W. (eds) Advances in Medical Education, pp. 661-664.

Harden, R. M., Stevenson, M., Downie, W. W. and Wilson, G. M. (1975) ‘Assessment of Clinical Competence using Objective Structured Examination’, British Medical Journal, 1(5955), pp. 447-451. https://doi.org/10.1136/bmj.1.5955.447.

Hawkins, R. E. (2005) ‘The introduction of clinical skills assessment into the United States medical licensing
examination (USMLE): a description of USMLE step 2 clinical skills (CS), *Journal of Medical Regulation*, 91(3), pp. 22–25. https://doi.org/10.30770/2572-1852-91.3.22.

Kashaf, M. S. (2017) ‘Clinical Skills in the Age of Google’, *Academic Medicine*, 92(6), pp. 734. https://doi.org/10.1097/ACM.0000000000001707.

Khan, K. Z., Ramachandran, S., Gaunt, K. and Pushkar, P. (2013) ‘The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part I: An historical and theoretical perspective’, *Medical Teacher*, 35(9), pp. 1437-1446. https://doi.org/10.3109/0142159X.2013.818634.

Lara, S., Foster, C. W., Hawks, M. and Montgomery, M. (2020) ‘Remote Assessment of Clinical Skills During COVID-19: A Virtual, High-Stakes, Summative Pediatric Objective Structured Clinical Examination’, *Academic Pediatrics*, 20(6), pp. 760-761. https://doi.org/10.1016/j.acap.2020.05.029.

Liu, C. (2012) ‘An introduction to workplace-based assessments’, *Gastroenterology and Hepatology from Bed to Bench*, 5(1), pp. 24-28. https://journals.sbmu.ac.ir/ghfbb/index.php/ghfbb/article/view/222.

Major, S., Sawan, L., Vogsen, J. and Jabre, M. (2020) ‘COVID-19 pandemic prompts the development of a Web-OSCE using Zoom teleconferencing to resume medical students’ clinical skills training at Weill Cornell Medicine-Qatar’, *BJM Simulation and Technology Enhanced Learning*, bmjstel-2020-000629. https://doi.org/10.1136/bmjstel-2020-000629.

Masters, K. (2019) ‘Artificial intelligence in medical education’, *Medical Teacher*, 41(9), pp. 976-980. https://doi.org/10.1080/0142159X.2019.1595557.

Nguyen, D. R. (2018) ‘It's Time to Invest in, Not Divest, the USMLE Step 2 Clinical Skills Test’, *Academic Medicine: Journal of the Association of American Medical Colleges*, 93(11), pp.1600. https://doi.org/10.1097/ACM.0000000000002418.

Paniagua, M., Salt, J., Swygert, K. and Barone, M. A. (2018) ‘Perceived Utility of the USMLE Step 2 Clinical Skills Examination from a GME Perspective’, *Journal of Medical Regulation*, 104(2), pp. 51-57. https://doi.org/10.30770/2572-1852-104.2.51.

Schuwirth, L. W. T. and Van Der Vleuten, C. P. M. (2003) ‘The use of clinical simulations in assessment’, *Medical Education*, 37(1 SUPPL), pp. 65-71. https://doi.org/10.1046/j.1365-2923.37.s1.8.x.

Stillman, P. P., Rutala, P., Nicholson, G., Sabers, D., et al. (1982) ‘Measurement of clinical competence of residents using patient instructors’, *Annu Conf Res Med Educ*, 21, pp. 111–116.

Sultan, L., Abuznadah, W., Al-Jifree, H., Khan, M. A., et al. (2019) ‘An experimental study on usefulness of virtual reality 360° in undergraduate medical education’, *Advances in Medical Education and Practice*, 10, pp. 907-916. https://doi.org/10.2147/AMEP.S219344.

Sutnick, A. I., Stillman, P. L., Norcini, J. J., Friedman, M., et al. (1994) ‘Pilot study of the use of the ECFMG clinical competence assessment to provide profiles of clinical competencies of graduates of foreign medical schools for residency directors’, *Academic Medicine*, 69(1), pp. 65-67. https://doi.org/10.1097/00001888-199401000-00019.
Tabish, S. A. (2008) ‘Assessment Methods in Medical Education’, *International Journal of Health Sciences*, 2(2), pp. 3-7. PMID: 21475483; PMCID: PMC3068728.

Turner, J. L. and Dankoski, M. (2008) ‘Objective structured clinical exams: a critical review’, *Fam Med*, 40(8), pp. 574–78. PMID: 18988044.

**Appendices**

None.

**Declarations**

*The author has declared that there are no conflicts of interest.*

*This has been published under Creative Commons "CC BY 4.0" ([https://creativecommons.org/licenses/by-sa/4.0/](https://creativecommons.org/licenses/by-sa/4.0/))*

**Ethics Statement**

Ethical approval was not required for this opinion piece because it is not reporting research findings.

**External Funding**

This article has not had any External Funding

---

MedEdPublish: rapid, post-publication, peer-reviewed articles on healthcare professions’ education. For more information please visit [www.mededpublish.org](http://www.mededpublish.org) or contact [mededpublish@dundee.ac.uk](mailto:mededpublish@dundee.ac.uk).