Discontinuation of the USMLE Step 2 Clinical Skills Examination: Studying the Past to Define the Future

Lisa D. Howley, PhD, and Deborah L. Engle, EdD, MS

Abstract

The United States Medical Licensing Examination (USMLE) Step 2 Clinical Skills (CS) was discontinued in January 2021, marking a significant milestone in assessment of clinical skills. In this commentary, the authors trace the history of the Step 2 CS exam—beginning with its early roots in the 1960s up to its discontinuation in 2021. In this new era, the medical education community is replete with opportunities for advancing methodology and content associated with clinical skills assessment. The authors propose 3 main lessons gleaned from this rich history and modern evolution, which are aimed at defining a future that includes creative collaboration toward development of comprehensive, equitable, student-focused, and patient-centered clinical performance assessment.

First, as it has done throughout history, the medical education community should continue to shift to more complex and student-driven approaches of assessment, that is, assessments that provide an unstructured environment, are realistic with respect to the natural conditions, and do not limit students to lists of options or force them to take a certain path of reasoning. Third, medical educators should continue to rethink the role of assessment and ensure that all assessments, regardless of stakes or type, provide sufficient feedback for the student to identify areas of strength and weakness.

On January 26, 2021, the Federation of State Medical Boards (FSMB) and the National Board of Medical Examiners (NBME) announced that the United States Medical Licensing Exam (USMLE) Step 2 Clinical Skills (CS) would be discontinued. This decision marked a significant milestone in the assessment of clinical skills in the United States and one that has historical as well as practical significance for medical students, educators, researchers, and others. The decision to suspend the licensing performance examination was not without controversy, and although a detailed exploration of these issues is beyond the scope of this commentary, we believe the historical context of this significant decision warrants a closer look. After all, Confucius reminds us to "study the past, if you would define the future," and several key lessons from history may better inform our path forward in this new era of clinical performance assessment.

The USMLE includes 3 steps and, until recently, 4 examinations toward medical licensure. It is intended to assess a physician’s ability to “apply knowledge, concepts, and principles, and to demonstrate fundamental patient-centered skills, that are important in health and disease and that constitute the basis of safe and effective patient care.” Each Step toward licensure is completed at a developmentally appropriate time in a physician’s training, and Step 2—parts CK (Clinical Knowledge) and the former CS—is typically completed at the end of the third year of medical school training. As medical educators who have been deeply engaged in clinical performance assessment for decades, we share several historical milestones and lessons as we consider a new era and the path forward for comprehensive, equitable, student-focused, and patient-centered clinical performance assessment.

1960s–1970s: Era of Origination

The modern-day performance assessment dates back to 1963 when Dr. Howard Barrows, a neurologist and educator at the University of Southern California, struggled while teaching his clerks clinical skills at the bedside. To solve this challenge, he recruited and trained a person to act like a patient—presenting numerous physical signs and symptoms and varying his or her affect—while the medical students individually conducted a history and physical examination of this “programmed” patient. What would eventually be termed the standardized patient (SP) was born. “[I was] learning things about those students I would have never found otherwise,” Dr. Barrows commented about his experiences using SPs with students. This pioneering work, which was originally met with ridicule and laughter, launched a new era of clinical skills teaching and assessment across North America and paved the way for more formats based on human simulation. The SP innovation would grow to become one of the most pervasive teaching methodologies in medical education with benefits including flexibility, standardization, and safety.

1980s–1990s: Era of Psychometrics

The 2 decades that followed resulted in adaptations and advances in the performance assessment of clinical skills across medical schools with a growing number of SP programs being developed to support the educational program. Various models of formative and
summative performance assessment were developed, including objective structured clinical exams, clinical skills assessments, and more. These timed examinations required medical students to interact with multiple SPs while receiving detailed oral and written feedback from faculty, and often the trained SPs themselves, on their ability to perform a variety of clinical skills, such as history taking, physical examination, patient education, communication, and clinical reasoning.

In 1984, the Association of American Medical Colleges (AAMC) released a report with conclusions and recommendations from its Panel on the General Professional Education of the Physician and College Preparation for Medicine. One recommendation was that appropriate evaluation methods be used, including those that assess more than cognitive skills such as memorization, recognition, and recall: "A new emphasis on active, independent problem solving will be undermined if evaluation methods are not congruent with this approach to general professional education." In the late 1980s, with the support of the Josiah Macy Jr. Foundation and endorsement by the AAMC, the Educational Commission for Foreign Medical Graduates (ECFMG), and the NBME, Southern Illinois University School of Medicine undertook a collaborative dissemination project to stimulate the adoption of SP-based examinations across medical schools throughout the United States. In 1992, the AAMC hosted the Consensus Conference on the Use of Standardized Patients in the Teaching and Evaluation of Clinical Skills, and a special interest group of the AAMC’s affinity group on educational affairs was formed to advance SP-based teaching and assessment (later becoming the independent international Association of Standardized Patient Educators). The psychometric properties of these complex standard assessment methods became a central focus of educational researchers and resulted in over 13,000 publications between 1980 and 2000, with 87% of those in the 1990s, including L.D.H.’s doctoral dissertation.5

During this time, the ECFMG, which was responsible for assessing the readiness of internationally trained medical graduates to enter U.S. residency or fellowship programs, struggled to effectively assess the clinical skills of international graduates. In 1998, building from the many lessons and considerable research on the use of SPs for performance assessment, the ECFMG successfully launched a large-scale clinical skills assessment intended to assess a graduate’s ability to communicate with patients and health professionals in English. The first national clinical performance examination in the profession was established.

**Early 2000s: Era of Licensure**

Due in part to concerns that all medical licensing examinations tested knowledge with a single format of multiple-choice questions and that international graduates were required to pass a broader assessment of clinical performance than U.S. graduates, the FSMB and the NBME tested and launched the USMLE Step 2 CS exam in 2004. This new licensing exam assessed students’ abilities to take a history, perform a physical exam, and communicate findings. The ECFMG clinical skills assessment was replaced by the new USMLE Step 2, and for the first time, all international graduates and U.S.-trained medical students would complete a clinical skills performance assessment. The modern Step 2 CS exam consisted of 3 components: history taking, physical examination, and a written patient note. Each examinee was required to encounter 12 SPs and had 15 minutes to take each patient’s history and, if relevant, perform a focused physical exam. Following each encounter, the examinees had 10 additional minutes to write a patient note about their findings and plan.6

This expanded assessment model for medical licensure in the United States did not come without controversy. The 12-hour Step 2 CS examination had to be completed at 1 of 6 testing centers in the United States and cost medical students approximately $1,500 to complete (in addition to the expenses associated with travel and time away from their training program), which was approximately 60% higher than the other USMLE licensing exams. The pass–fail examination had a high pass rate of approximately 95% and did not provide specific performance feedback to students about their clinical skills. In addition, due to the increased need for standardization across sites and its reliance on SP raters, the clinical encounter was reduced to discrete observable behaviors documented via a checklist and a patient note. The encounters or stations included in the examination were limited to initial visits with single patients, and access to resources for informing clinical decisions was limited. In response, many local SP programs that had been designing and researching new models for clinical skills assessment changed their approach to more closely align with the Step 2 CS model.

This new licensure exam also resulted in the increased development of local SP-based clinical performance assessments, which were intended, in part, to prepare students for Step 2 CS. These local methods became part of the larger assessment system of the school that developed them. Although beyond the scope of this brief commentary, we would be remiss not to mention the broader movement toward competency-based medical education (CBME), which is an outcomes-based approach to the design, implementation, and evaluation of educational programs and the assessment of learners. As medical schools began to adopt a CBME approach, they also broadened their assessment systems to include more criterion-based, formative, and workplace-based methods and to conduct assessments more frequently. Additionally, this shift underlined the value of group process in making judgments about trainees. The need to more frequently and effectively assess medical students’ clinical skills was only heightened within the context of CBME.

**2020s and Beyond: Era of Creative Collaboration**

In May 2020, due to the COVID-19 pandemic, the FSMB and the NBME suspended Step 2 CS for 12 to 18 months with the intent of reinstating it. They explained they would bring back “a modified Step 2 CS exam that was appreciably better than the prior assessment,” but in January 2021, they announced their decision to discontinue the Step and to take the opportunity to “focus on working with our colleagues in medical education and at the state medical boards to determine innovative ways to assess clinical skills.” As we begin to define the future, we propose several lessons from the history and evolution of clinical performance assessment.

First, as we have done throughout history, the medical education community should
continue to innovate, collaborate, and improve upon our methods of clinical skills assessment. SP-based modalities remain a powerful method for patient-centered assessment while also protecting our learners and patients from harm. We should work to provide performance-based formative and longitudinal methods that simulate clinical encounters that are representative of diverse patients, family members, teams, and varied settings. The discontinuation of Step 2 CS may allow for greater creativity and innovation at medical schools to explore assessments that measure the complex clinical skills of its students, including communicating with diverse patients and teams and providing in-person and virtual care that is equitable and safe. Furthermore, these methods should also be considered in context with the school’s broader assessment system, including the increasing use of workplace-based methods. Multiple assessment formats, across different settings and by numerous assessors, will better inform longitudinal evaluation of students’ competence.

Second, medical educators should continue to shift to more complex and student-driven approaches for assessment. We define student-driven assessments as those that provide an unstructured environment, are realistic with respect to the natural conditions, and do not limit students to lists of options or force them to take a certain path of reasoning. The complexity of clinical skills should not be reduced to a purely discrete checklist of behaviors. We should instead integrate into our assessment methods our growing understanding of how students develop clinical reasoning, make decisions, make errors, and develop competency. We should shift from overreliance on dyad (student–SP) encounters, for example, to multiple-person simulations and modify the method of measurement from checklists to measures more capable of assessing advanced cognitive skills. Ultimately, these less reductionist approaches may lead to more cognitively advanced measures of clinical competence.

Third, medical educators should continue to rethink the role of assessment and ensure that all assessments, regardless of stakes or type, provide sufficient feedback for the student to identify areas of strength and weakness. In addition, we should work to ensure that our methods are equitable, accessible for all students, and free of bias. In other words, we should continue to shift from the assessment of learning toward the assessment for learning and equity. Similarly, we should work toward identifying greater efficiencies when designing and administering SP-based assessments, including multi-institutional and regional collaborations for virtual and in-person examinations. The medical education community, including medical schools, their affiliates, and licensing and accrediting bodies, has a shared obligation to the public to ensure clinical skills are appropriately, fairly, and uniformly assessed. Clinical skills assessment should continue to expand locally and should remain within the national licensing pathway. Ultimately whether the assessment is locally, nationally, or jointly designed and administered, our broad community should work collaboratively, take lessons from our shared past, and design our new future of clinical skills assessment.

The past year has been transformative as COVID-19 has disrupted nearly all practices in medical education. Substantial changes to the U.S. national licensing examination will have many ripple effects, including opportunities for advancing what and how we assess the clinical skills of our students within and across our medical schools. Although we cannot know what the future will hold, we predict that through greater creativity and collaboration, we will have defined a future that includes comprehensive, equitable, student-focused, and patient-centered assessment of clinical skills.

Funding/Support: None reported.

Other disclosures: None reported.

Ethical approval: Reported as not applicable.

L.D. Howley is senior director of strategic initiatives and partnership, Association of American Medical Colleges, Washington, DC, and adjunct associate professor, University of North Carolina School of Medicine, Chapel Hill, North Carolina.

D.L. Engle is assistant dean of assessment and evaluation and associate professor, Duke University School of Medicine, Durham, North Carolina.

References
6 United States Medical Licensing Examination. Step 2 CS Content Description and General Information. Federation of State Medical Boards and the National Board of Medical Examiners; 2015.