Evolution of Clinical Skills Assessment in the USMLE: Looking to the Future After Step 2 CS Discontinuation
Peter J. Katsufrakis, MD, MBA, and Humayun J. Chaudhry, DO, MS

Abstract
The COVID-19 pandemic interrupted administration of the United States Medical Licensing Examination (USMLE) Step 2 Clinical Skills (CS) exam in March 2020 due to public health concerns. As the scope and magnitude of the pandemic became clearer, the initial plans by the USMLE program’s sponsoring organizations (NBME and Federation of State Medical Boards) to resume Step 2 CS in the short-term shifted to long-range plans to relaunch an exam that could harness technology and reduce infection risk. Insights about ongoing changes in undergraduate and graduate medical education and practice environments, coupled with challenges in delivering a transformed examination during a pandemic, led to the January 2021 decision to permanently discontinue Step 2 CS. Despite this, the USMLE program considers assessment of clinical skills to be critically important. The authors believe this decision will facilitate important advances in assessing clinical skills.

Factors contributing to the decision included concerns about achieving desired goals within desired time frames; a review of enhancements to clinical skills training and assessment that have occurred since the launch of Step 2 CS in 2004; an opportunity to address safety and health concerns, including those related to examinee stress and wellness during a pandemic; a review of advances in the education, training, practice, and delivery of medicine; and a commitment to pursuing innovative assessments of clinical skills. USMLE program staff continue to seek input from varied stakeholders to shape and prioritize technological and methodological enhancements to guide development of clinical skills assessment. The USMLE program's continued exploration of constructs and methods by which communication skills, clinical reasoning, and physical examination may be better assessed within the remaining components of the exam provides opportunities for examinees, educators, regulators, the public, and other stakeholders to provide input.

On January 26, 2021, the NBME and the Federation of State Medical Boards (FSMB) announced that the United States Medical Licensing Examination (USMLE) program had discontinued its work to modify and relaunch the Step 2 Clinical Skills (CS) examination. This decision followed the initial, temporary cessation of Step 2 CS announced in March 2020, as infection with SARS-CoV-2 (the virus that causes COVID-19) began spreading throughout the United States, and the subsequent May 2020 extension of the suspension for 12–18 months. As events unfolded and the COVID-19 pandemic caused alarming surges of infections, hospitalizations, and deaths around the United States, our plans to resume a version of Step 2 CS that substantially reduced the risk of virus transmission evolved to permanent discontinuation of the exam.

This shift in plans involved careful consideration of multiple factors intrinsic and external to the USMLE program. By no means, however, is this decision intended to minimize the importance of assessing clinical skills in the licensure exam. Evaluating clinical skills within the USMLE program and throughout the duration of medical education and training is critically important. As we will discuss in this commentary, our goals are to reflect evolution of both educational and clinical practice as we evolve clinical skills assessment in the USMLE.

USMLE Governance and the Step 2 CS Decision
As cosponsors of the USMLE program, the FSMB and NBME have ultimate responsibility for program oversight of the entire exam sequence. The NBME and Educational Commission for Foreign Medical Graduates (ECFMG) collaboratively administered the Step 2 CS component. Many USMLE policy decisions—such as determination of exam content, cutoffs for a passing score, and related operational matters—are delegated to various oversight committees comprising volunteers from medical school faculties, state medical boards, and the public. The decision to discontinue Step 2 CS was made by governance and staff leadership of the FSMB and NBME in consultation with governance and staff leadership of the ECFMG.

As part of the USMLE program's continuous efforts to enhance assessment, research over the past several years has been focused on longer-term transformation of clinical skills assessment—by using avatars, multimedia, and other enhancements in Step 2 CS to improve not only the nature of the assessment but also the experience of the examinee. The pandemic shutdown and the drive to launch a revised exam provided an unforeseen opportunity to commit additional resources to this work. When the 12- to 18-month hiatus of Step 2 CS was announced in May 2020, the USMLE program refocused on shorter-term revisions and relaunching an exam that was appreciably enhanced compared with the exam that was suspended.
in March 2020. We envisioned some combination of changes that harnessed technology, reduced or eliminated exam-associated COVID-19 risk, reduced or eliminated the need for examinee travel, and reflected changes in medical education arising since the exam was first launched in 2004.

During the rest of 2020, USMLE staff worked to analyze current elements of medical practice, engaged with educators working in medical school and residency program settings, solicited input from stakeholder groups, and explored various technology solutions to enable relaunch of Step 2 CS. While good progress had been made, the FSMB and NBME jointly determined in January 2021 that it was not feasible to relaunch an exam that was appropriately enhanced within our targeted time frame.

Factors Contributing to the Step 2 CS Decision

No single factor led to this determination. The decision to discontinue Step 2 CS involved a holistic review of USMLE program goals, progress made toward relaunching the exam, input collected from varied stakeholders, and analysis of the education and practice environments. We briefly summarize, below, some of the salient considerations.

The rapid spread of COVID-19 throughout the world, with risks across all age groups and demographics, dramatically changed our plans for Step 2 CS. The exam by design required close physical proximity and physical contact between examinees and exam staff, especially staff acting as standardized patients. The safety and health of examinees, including stress and anxiety caused by the uncertainty of the pandemic, were of paramount concern, as were the safety and welfare of standardized patients and other staff. We initially sought to modify the exam to simulate a virtual telehealth platform where neither examinees nor staff would need to come into contact with others, to reduce or eliminate potential COVID-19 risks associated with the exam. We also strove to develop a model for remote administration that could reduce or eliminate the travel-associated costs for examinees. These exam design characteristics would have retained assessment of noncognitive domains involving information gathering, manner of interaction, and communicating findings to patients and colleagues, but not assessment of physical examination skills.

Balanced against the benefits of a remotely administered exam were the challenges associated with technology, security, equity, and exam logistics, as well as the obvious limitations to assessing physical examination skills via a virtual performance assessment. While we had the benefit of learning from the cautionary experiences of other regulatory organizations attempting remote exam administration and from the successful NBME launch of school-based remote proctoring of exams in medical schools, we found that these experiences did not adequately generalize to a large-scale national licensing exam specific to the USMLE program’s needs. Ultimately, we did not identify a remote administration solution that would provide sufficient value.

In addition, meaningful advances in medical education have occurred since Step 2 CS was launched in 2004, when it represented a significant advance as a licensing tool by requiring examinees to demonstrate skills beyond traditional knowledge and content areas. These developments include the establishment of objective structured clinical examination (OSCE) labs at U.S. MD-granting and DO-granting medical schools and the adoption of a competency framework—including entrustable professional activities, Accreditation Council for Graduate Medical Education (ACGME)/American Board of Medical Specialties core competencies, and graduate medical education (GME) milestones—to assess learners’ progress along the continuum of medical education.2 Independent, third-party verification and support for a national standard, however, remain cornerstones of a robust system of medical licensure in the United States. State medical boards rely on the USMLE program,3 and their licensing decisions are predicated on an education and training system that includes successful completion of a program of undergraduate medical education (UME); completion of some or all of a program of GME; independent and separate accreditation of UME and GME programs; and, for international medical graduates, certification of physician applicants’ qualifications by the ECFMG. In the context of this robust system of assessment, certification, and regulation, the USMLE leadership weighed the incremental additional value of relaunching a modified Step 2 CS.

The growth in U.S. medical schools’ clinical skills training and assessment programs, both in prevalence and sophistication, has arguably improved the preparation of newly minted physicians entering residency training. As the systems for training and assessing clinical skills have evolved since the launch of Step 2 CS,4 the practice of medicine has also evolved in ways not reflected in the exam. Resources such as online reference materials, artificial intelligence decision aids, and other technology-enabled tools have changed how medicine is practiced and care is delivered. While telehealth was growing in acceptance and utilization before the COVID-19 pandemic, the year 2020 saw a dramatic increase in its use by physicians and patients. The significant staff resources required to relaunch a Step 2 CS exam modified for the pandemic, we agreed, could instead be devoted to advancing the assessment of clinical skills in a more transformative fashion.

Looking to the Future

State medical boards remain committed to ensuring that physicians provide safe and effective patient care. The USMLE program will continue to serve that mission and deploy assessments that meet the known and emerging requirements of medical licensure. We are soliciting input from key stakeholders, including members and representatives of state medical boards and individuals involved directly and indirectly in the education and training of future physicians—medical school faculty, residency program directors and faculty, medical students, examinees, practitioners, and members of the public. In seeking stakeholder guidance, we are striving to identify how the evolution of medical education, training, and practice should best be reflected in USMLE assessments. One line of inquiry seeks to understand where the greatest needs for clinical skills assessment exist. The information we gather will help prioritize subsequent research and development. Another line of inquiry seeks innovations...
In clinical skills assessment that are applicable to the context of medical licensure.

Based on the feedback we have already received, we are likely to increase emphasis on subjects particularly important to medical practice (e.g., clinical reasoning) and areas in which state medical boards identify deficiencies (e.g., communication). The initial enhancements to clinical skills assessment in USMLE exams will augment current formats and integrate them into the existing exam structure. To the extent possible, content revisions identified through stakeholder input may be reflected in multiple-choice questions in all 3 Step exams and/or computer case simulations in the Step 3 exam. Ongoing research, however, will seek to push beyond the limitations of these formats, incorporating advances in audio and visual media, patient avatars, natural language processing, artificial intelligence, and other combinations of assessment science and technology that extend the frontiers of clinical skills assessment. It is likely that technological advances will result in the development of new item types.

We believe that the process outlined above—diverse stakeholder input combined with research and development that targets promising technological innovations—will create the optimal developmental path for clinical skills assessment in the USMLE program. There will almost certainly be tradeoffs among various exam and design characteristics as we strive to assess physical examination and communication skills; inform determinations of readiness for entry into supervised practice and subsequent unsupervised practice; assess clinical reasoning; reflect contemporary medical practice; ensure equity and access; and optimize the examinee experience.

In deciding to discontinue Step 2 CS, we recognized the potential for some learners and educators to wrongly conclude that clinical skills assessment is no longer being valued. The USMLE program remains committed to assessment of clinical skills. As such, educators and administrators should resist relaxation of their standards and continue to prioritize the time and resources devoted to clinical skills training and assessment. All parties in the education and regulatory systems will need to continue to work independently and collaboratively across the continuum of medical education in the public interest. Working independently, medical educators must continue to address the broad array of knowledge and skills necessary to practice medicine regardless of whether they are reflected in USMLE exams, and the USMLE program must continue to serve its function of validating examinees’ readiness to practice. We are hopeful that our recent experience with the Invitational Conference on USMLE Scoring and the current work of the Coalition for Physician Accountability’s UME-GME Review Committee provide models for effective, coordinated collaboration among multiple stakeholders to guide systemic improvements to medical education, assessment, and regulation. Our short-term and long-range plans are to enhance the assessment of examinees’ clinical skills, particularly in the area of physician–patient communication, as we transform all Steps of the USMLE program to meet the evolving needs of state medical boards and the public we collectively serve.

Acknowledgments: The authors wish to thank Michael Barone, MD, Chris Feddock, MD, Eric Fish, JD, Linda Gadsby, JD, Michael Jodoin, PhD, David Johnson, MA, Joe Knickrehm, Alex Mechaber, MD, and Pam Treves for their review of the manuscript and comments.

Funding/Support: None reported.

Other disclosures: The authors are the presidents and CEOs of the organizations that sponsor the USMLE program.

Ethical approval: Reported as not applicable.

References