Author: Kogan Jennifer R. MD; Hauer Karen E. MD, PhD; Holmboe Eric S. MD

Title: The Dissolution of the Step 2 Clinical Skills Examination and the Duty of Medical Educators to Step Up the Effectiveness of Clinical Skills Assessment

DOI: 10.1097/ACM.0000000000004216
Academic Medicine

DOI: 10.1097/ACM.0000000000004216

The Dissolution of the Step 2 Clinical Skills Examination and the Duty of Medical Educators to Step Up the Effectiveness of Clinical Skills Assessment

Jennifer R. Kogan, MD, Karen E. Hauer, MD, PhD, and Eric S. Holmboe, MD

J.R. Kogan is associate dean, Student Success and Professional Development, and professor of medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania; ORCID: https://orcid.org/0000-0001-8426-9506.

K.E. Hauer is associate dean, Competency Assessment and Professional Standards, and professor of medicine, University of California, San Francisco, School of Medicine, San Francisco, California; ORCID: https://orcid.org/0000-0002-8812-4045.

E.S. Holmboe is chief, Research, Milestones Development and Evaluation, Accreditation Council for Graduate Medical Education, Chicago, Illinois; ORCID: https://orcid.org/0000-0003-0108-6021.

Correspondence should be addressed to: Jennifer R. Kogan, Jordan Medical Education Center, 3400 Civic Center Boulevard, Building 421, Philadelphia, PA 19104; telephone: (215) 662-2250; email: koganj@pennmedicine.upenn.edu; Twitter: @KarenHauer4.

Funding/Support: Dr. Kogan receives research support from the Accreditation Council for Graduate Medical Education (ACGME) and, as co-director of the SEEF (Strategic Educators Enhancement Fund) Fellowship, from the National Board of Medical Educators (NBME). Dr. Holmboe receives royalties from Elsevier.
Other disclosures: Dr. Hauer is a member of the Undergraduate Medical Education to Graduate Medical Education Review Committee (UGRC) convened by the Coalition for Physician Accountability.

Ethical approval: Reported as not applicable

Disclaimer: The views expressed are those of the authors and not necessarily their institutions.
Abstract

In this Invited Commentary, the authors explore the implications of the dissolution of the Step 2 Clinical Skills Examination (Step 2 CS) for medical student clinical skills assessment. The authors describe the need for medical educators (at both the undergraduate and graduate level) to work collaboratively to improve medical student clinical skills assessment in order to assure the public that medical school graduates have the requisite skills to begin residency training. The authors outline 6 specific recommendations for how to capitalize on the discontinuation of Step 2 CS to improve clinical skills assessment: (1) defining national, end-of-clerkship and transition-to-residency standards for required clinical skills and for levels of competence; (2) creating a national resource for standardized patient, augmented reality, and virtual reality assessments; (3) improving workplace-based assessment through local collaborations and national resources; (4) improving learner engagement in and coproduction of assessments; (5) requiring, as a new standard for accreditation, medical schools to establish and maintain competency committees; and (6) establishing a national registry of assessment data for research and evaluation. Together, these actions will help the medical education community earn the public’s trust by enhancing the rigor of assessment to ensure the mastery of skills that are essential to providing safe, high-quality care for patients.
In May 2020, in the midst of the COVID-19 pandemic, the Federation of State Medical Boards (FSMB) and the National Board of Medical Examiners (NBME), which cosponsor the United States Medical Licensing Examination (USMLE), temporarily suspended the Step 2 Clinical Skills Examination (Step 2 CS). The original intent was to relaunch an improved version of the exam; however, on January 26, 2021, after “reviewing current and anticipated progress with the exam and in consideration of the rapidly evolving medical education, practice and technology landscapes,” the FSMB and NBME announced the discontinuation of Step 2 CS. Their focus shifted to partnering with the medical education community to identify innovative ways to assess clinical skills while ensuring that clinical reasoning and communication skills would continue to be assessed through other exams in the USMLE sequence.

In this Invited Commentary, we review the background of Step 2 CS and consider the implications of its discontinuation for clinical skills assessment in undergraduate medical education (UME). We advocate for medical school educators, working with the graduate medical education (GME) community, to assume an increased role in and accountability for assuring the public that medical school graduates have the requisite clinical skills to begin residency training.

**Background: A Brief History of Step 2 CS**

Step 2 CS, which had strong support from the public as a means of fulfilling medicine’s obligation to society, promised to standardize clinical skills assessment and, thereby, ensure minimum competence for physician licensure and guarantee physicians’ readiness for residency.

A clinical skills assessment was initially required for international medical graduates seeking residency training in the United States, and then, in 2005, clinical skills assessment became a licensure requirement for students graduating from U.S. MD-granting medical schools as well. Step 2 CS assessed examinees’ ability to apply both clinical science knowledge and the skills foundational to providing safe, effective patient-centered care under supervision. The exam
comprised 10 to 12 clinical encounters portrayed by highly trained standardized patients (SPs). Collectively, the encounters were designed to assess medical students’ skills in history-taking, physical examination, patient communication, and written documentation. Scoring addressed three areas: (1) the integrated clinical encounter measuring data-gathering skills, clinical reasoning, and note-writing; (2) patient-centered communication skills; and (3) spoken English proficiency. The exam was graded pass/fail; examinees were required to pass all three components to achieve a passing score.

The introduction of Step 2 CS catalyzed changes in medical school curricula and assessment. Step 2 CS helped transform the objectives, content, and emphases of Introduction to Clinical Medicine courses.\(^4,5\) Medical schools implemented and/or enhanced not only clinical skills training but also clerkship and end-of-clerkship-year comprehensive SP exams—in part to prepare students for the national licensing exam.\(^6\) Passing Step 2 CS ultimately became a graduation requirement at many medical schools.

**After the Dissolution of Step 2 CS: A Vacancy in Clinical Skills Assessment**

Medical educators and students should not infer that the discontinuation of Step 2 CS means that the skills assessed on the exam are unimportant or that enhanced clinical skills assessment at the local level is not warranted. Despite its limitations (e.g., cost, limited testing sites, validity concerns, pass/fail scoring), Step 2 CS provided a national standard for minimal competence in the application of clinical science knowledge and in the ability to communicate effectively with patients. Additionally, Step 2 CS provided residency educators some assurance of students’ preparedness to begin residency training.\(^7\) The discontinuation of Step 2 CS, along with the transition of both Step 1 and, increasingly, core clerkships to pass/fail grading, eliminates assessments or specific assessment score data upon which residency program directors have traditionally relied in the process of selecting residents. While program directors used data from
these assessments to make decisions, whether the data provided the type of information program directors needed about students’ achievement of core clinical skills and their ability to competently interact with patients, is not certain.

Now, more than ever, it is incumbent upon medical schools to ensure that graduates have the core patient care and communication skills necessary to provide high-quality care for patients under supervision. Multiple studies have shown the important relationship between clinical skills and the provision of safe, high-quality patient care.\(^8\)–\(^10\) Importantly, these skills cannot be entirely assessed using proxy information, such as giving an oral presentation, answering questions on rounds, or selecting the correct answer on a multiple-choice examination. Being able to correctly identify an appropriate communication technique on a written exam is an especially inadequate proxy for being able to communicate effectively with patients and colleagues in a clinical setting.

Given the importance of clinical skills in residency training, it is imperative that medical schools continue to develop their assessments of medical students’ attainment of these skills. Most medical schools use SP examinations; however, different, localized expert groups create substantially different checklists depending on their personal preferences or other factors irrelevant to the construct being measured\(^11\)—a reality that makes the loss of the national standard (i.e., Step 2 CS) especially problematic. Furthermore, workplace-based assessment (WBA), through which medical students are directly observed interacting with patients, remains suboptimal in quantity and quality. As such, both SP examinations and WBA during medical school require urgent attention if medical schools are to show that they have prepared their students effectively for residency training.

Medical educators from across the UME-GME continuum must together shoulder the responsibility for clinical skills assessment. They must clearly articulate the core skills that medical students must demonstrate through observation and assessment. Educators must also
define and share a mental model of what level of performance is required to deem a medical school graduate competent to provide safe, effective, patient-centered care under supervision. National standards for the skills and skill levels to be attained in UME do not currently exist and have been only partially defined by each specialty via the GME competency milestones. While the discontinuation of Step 2 CS has currently focused the spotlight on UME, substantial reform in assessing clinical skills is also urgently needed in GME as well. As such, increased collaboration between UME and GME going forward is essential.\textsuperscript{12}

**Recommendations to Enhance Clinical Skills Assessment**

We believe the dissolution of Step 2 CS should catalyze medical schools to enhance their assessment of medical students’ clinical skills and other competencies. We believe schools should implement programmatic assessment, which entails multiple assessment methods, both formative and summative.\textsuperscript{13} This comprehensive approach extends beyond just giving SP-based clinical skills exams to also fulfilling the obligation of medical education to serve patients’ needs. A shift to programmatic assessment necessitates the presence of skilled assessors working within a learning environment that cultivates medical student engagement through a mastery-oriented learning culture.\textsuperscript{14} Additionally, to interpret assessment data, medical schools need competency committees to evaluate student performance in core competency domains using national, criterion-referenced performance standards. That is, these committees must determine whether each individual student has achieved expected competence rather than comparing students’ performances to those of their peers or ranking students within a class. Such assessments require not only substantial time and investment of resources but also a substantial shift in the approach to education and assessment, which many schools may be unable to accomplish independently. Therefore, below we outline 6 specific suggestions for how the medical education community might work together to realize this vision. Our recommendations
for how to capitalize on the discontinuation of Step 2 CS to improve clinical skills assessment
are as follows: (1) defining national, end-of-clerkship and transition-to-residency standards for
required clinical skills and for levels of competence; (2) creating a national resource for SP,
augmented reality, and virtual reality assessments; (3) improving WBA through local
collaborations and national resources; (4) improving learner engagement in and coproduction of
assessments; (5) requiring, as a new standard for accreditation, medical schools to establish and
maintain competency committees; and (6) establishing a national registry of assessment data for
research and evaluation.

Recommendation 1: Define national, end-of-clerkship and transition-to-residency
standards for clinical skills and for levels of competence

Going forward, we believe U.S. medical schools, together with other key stakeholders, should
come to a consensus on what clinical skills are necessary by the end of the core clerkship year
and which specialty-specific clinical skills are needed by graduation to ensure an effective
specialty-specific transition to residency. The Association of American Medical Colleges
(AAMC) Core Entrustable Professional Activities for Entering Residency could define core
clerkship year (generalist) national competence standards.15 A national group of specialty-
specific residency educators could define (and redefine, as needed) the skills and levels of
competence necessary to begin specialty-specific residency training. Aligning skills required
upon graduation with residency competency milestones would highlight the role of the fourth
year of medical school as a period for transitioning to residency.16 Every medical school would
be required to have a summative pass/fail assessment at the end of the core clerkship year,
followed by a specialty-focused assessment to inform learner handovers prior to the transition to
residency. That is, end-of-medical-school assessments would provide learners and residency
program directors with key information about each graduate’s strengths and areas requiring improvement.

National standards of competence would be required for these assessments to assure the public of graduates’ readiness to advance. Historically, medical schools have enjoyed independence in standard setting and determining pass/fail cutoffs. However, locally defined and potentially idiosyncratic criteria, especially regarding what constitutes insufficient competence (i.e., failure), contribute to inadequate accountability. As a result, the entire medical education system is not fulfilling its responsibility to patients and the public. National standards could be combined with each school’s assessment of the skills that address the institution’s unique curricular emphasis and mission. Transition to a nationally based standard could be mandated through new accreditation standard requirements. National standards of competence in core knowledge, skills, and attitudes would also benefit residency program directors, especially if those standards are assessed and reported similarly across medical schools.

**Recommendation 2: Create a national resource for SP, augmented reality, and virtual reality assessments**

In addition to national standardized tests to assess medical student competence, we advocate a national, and potentially international, electronic repository of formative and summative clinical skills assessments that include SPs. The national repository would provide resources for case development; SP and other trainer resources; and checklists, scoring strategies, and rubrics grounded in criterion-referenced assessments based on what is required for starting interns. The repository would include assessments that focus on achievement of core skills required by the end of the clerkship year, as well as assessments that focus on transition to residency and specialty-focused skills. These materials could be used for a high-stakes summative assessment or, better yet, ongoing, formative progress testing.
Development of this repository would require collaboration among UME clinical skills educators, UME clerkship directors, and GME educators. We believe that the USMLE’s extensive experience with SP exams could be shared with medical schools to strengthen SP assessments at the local level. A similar repository of augmented reality and virtual reality resources should also be developed, perhaps in conjunction with a national organization such as the NBME.

A national repository, supported financially by all medical schools, would enable medical school leaders to prioritize assessing students rather than duplicating efforts in creating, pilot-testing and, collecting validity evidence for simulated cases. As new skills or competencies are identified, national experts could collaborate to create new cases. Ongoing work to define ownership of and access to the repository, to develop and refine the assessment materials, and to gather evidence of validity would be needed.

**Recommendation 3: Improve WBA through local collaborations and national resources**

The discontinuation of Step 2 CS heightens the need for medical schools to ensure a longitudinal, developmental approach to WBA; medical schools must facilitate direct observation of students’ clinical skills and provide students with feedback about their work in clinical settings. Despite the value of SP exams and their use as a powerful tool to assess how a trainee executes a skill, SP assessments are hampered by high costs, limited availability, and reduced authenticity. Given the diversity and range of actual patients in various clinical contexts, simulations cannot fully capture medical students’ performance during authentic encounters with real patients. Realizing the value of WBA requires evidence-based strategies for faculty development and learner engagement; that is, faculty need programs to promote their observation of skills, and faculty and students alike must learn to initiate and participate in feedback dialogues.\(^{17}\)
Medical school deans and department chairs must prioritize and commit necessary resources to faculty assessment of medical students’ clinical skills through WBA. Institutional leaders must ensure that the clinical learning environment is structured to allow time both for frequent WBA and for the development of the longitudinal learner-teacher relationships necessary for effective direct observation and feedback. Medical schools should consider developing education champions and core groups of clinician-educator faculty to serve as local experts on clinical skills teaching and assessment. Since it is unlikely that all medical schools have the resources, expertise, and bandwidth to develop and implement assessor training programs, national resources and regional or national collaborations could be leveraged to develop and provide assessor training. Such evidence-based training, available to all schools, would help faculty assessors consistently apply the standard of what medical students must be able to do to provide safe, effective, patient-centered care under supervision. Training is essential; without it faculty are likely to use normative judgements (comparing one student to another) or idiosyncratic standards.\textsuperscript{18} Training in observation and in applying national minimal standards should be accompanied by training in providing high-quality feedback through recommended, evidence-based approaches.\textsuperscript{19} Institutions must support faculty with protected time to undertake initial training and maintain WBA assessment skills.

**Recommendation 4: Improve learner engagement in and coproduction of assessments**

Medical schools should embrace the philosophy and techniques of coproduction, through which educators partner with their students; outcomes are optimized when the expertise of teachers and learners drives educational experiences and assessments.\textsuperscript{20} Coproduction entails training students to work actively and collaboratively with their supervisors and patients to identify opportunities for WBA, to request observations, to share self-assessments of their clinical performance, and to engage in feedback conversations. Several competency-based education pilots have
demonstrated the power of coproducing assessments. For example, students at the University of Minnesota School of Medicine (UM SOM), as a part of the Educating Pediatricians across the Continuum (EPAC) pilot, have been empowered to seek assessment on a regular basis. The result has been a dramatic increase in the number of assessments completed per EPAC student compared with the number completed for students in the traditional UM SOM program.\textsuperscript{21} Coproduction requires being transparent with students about the purpose of assessments and about expected skills and performance levels. By sharing expectations, medical educators help students become self-regulated learners who can set meaningful clinical skills learning goals, seek out observation, and assess their own performances using criterion-referenced, evidence-based, agreed-upon standards. Feedback from faculty should guide students, help them refine their learning goals, and become master adaptive learners.\textsuperscript{22} WBA applications for mobile devices can facilitate efficient, real-time data capture and promote students’ own self-assessment. Some apps are programmed with evidence-based frameworks, which allow students to compare their performance on both core and specialty-specific skills to criterion-based standards.\textsuperscript{23,24} If such apps were implemented correctly and disseminated widely, medical educators could leverage the catalytic effect of ongoing, workplace-based clinical skills assessment to drive ongoing learning.

**Recommendation 5: Require, as a new standard for accreditation, medical schools to establish and maintain competency committees**

Medical schools should create competency committees to assess students longitudinally, as has been done in GME. Systematic, ongoing program-wide assessment can help address the limitations associated with high-stakes SP exams by synthesizing multiple sources of assessment data (e.g., simulations, WBA in actual clinical settings, written exams). Medical school competency committees would be the parties responsible for this synthesizing of multiple forms
of assessment data from different assessment types. National resources, applying lessons learned from GME, could support this endeavor. Competency committee members would be trained using national guidelines that define standards of competence (and could incorporate school-based measures used to assess competencies specific to the local curriculum). Medical school competency committees—integral to systematic, ongoing, program-wide assessment—would contribute to the longitudinal development of students by helping them identify their strengths and weaknesses and by providing individualized learning plans. These committees would simultaneously supply medical schools with valuable information about gaps and strengths in their learning programs, which would, in turn, supporting curricular enhancement.

**Recommendation 6: Establish a national registry of assessment data for research and evaluation**

The aforementioned national collaborations (e.g., to create a standard for competence, to develop SP assessments) could promote the creation of a national registry of deidentified assessment data, which could be leveraged for research. For example, if an SP exam were required for graduation and if schools could agree on a set of minimal standards for passing, a national registry of experience and results could be used for continuous quality improvement and validity research. With the growth of WBA and assessment apps to capture performance information, the data resulting from multi-school studies could drive improvements not just in WBA, but in the sharing of evidence-based standards for assessing core clinical skills “in vivo.” A powerful example of the potential value and benefits of a registry of longitudinal national data is the information on milestones now being used for validity research, for continuous quality improvement of curriculum and assessment, and for the ongoing refinement of the competency framework in GME.\(^{25,26}\) Individual medical schools could use the data to better understand their
own curricula and to drive continuous quality improvement. A national registry could also support research focused on the UME-to-GME transition.

Conclusions

Step 2 CS is gone. The dissolution of the exam creates the risk that some may view the skills assessed by Step 2 CS as too difficult or costly to measure or, even worse, not important enough to measure. This conclusion does tremendous disserve to patients and jeopardizes the trust the public has placed in medical education. Instead, let us medical educators take this opportunity to enhance clinical skills assessment in medical schools by creating a coordinated approach to direct observation and feedback, by determining and applying agreed-upon standards for achievement and minimal competence, and by using systematic, ongoing program-wide assessment. Let us enhance clinical skills assessment in a way that ensures that we are assessing the right skills, in the right way, using the right standard: safe, high-quality, patient-centered care. This sea change provides an opportunity to earn the public’s trust by advancing the rigor of assessment within medical schools, assuring that as institutions and educators we cultivate the mastery of skills that are essential to providing the highest quality of care for patients.
References


