USMLE Step-1 is Going to Pass/Fail, Now What Do We Do?

Andrea N. Belovich1 · Ingrid Bahner2 · Giulia Bonaminio3 · Anthony Brenneman6 · William S. Brooks5 · Cassie Chinn6 · Nehad El-Sawi7 · Michele Haight8 · Sandra B. Haudek9 · Uzoma Ikonne10 · Robert J. McAuley11 · Douglas McKell12 · Rebecca Rowe13 · Tracey A. H. Taylor11 · Richard C. Vari14

Accepted: 21 May 2021 / Published online: 4 June 2021 © International Association of Medical Science Educators 2021

Abstract
The Winter 2021 Webinar Audio Series (WAS) of the International Association of Medical Science Educators (IAMSE), titled, “USMLE Step-1 is Going to Pass/Fail, Now what do we do?” was broadcast live to audiences at academic institutions worldwide in five weekly webinars from January 7, 2021, to February 4, 2021. Recognized experts from various stakeholder groups discussed the impact of the decision to score the United States Medical Licensing Examination (USMLE) Step 1 exam Pass/Fail (P/F). The speakers identified challenges to their respective programs and explored creative ways to address potential consequences. Sessions included the perspectives of allopathic and osteopathic residency program directors, basic science faculty, undergraduate medical education curriculum designers, clinical educators, and programs for international medical students matriculating to the United States. On February 25, 2021, a bonus session provided cutting-edge updates from a co-chair of the Coalition for Physician Accountability Undergraduate Medical Education (UME) to Graduate Medical Education (GME) Review Committee (UGRC).

USMLE Transition to Pass/Fail: Implications for Resident Candidate Assessment and Selection

Presenters: Mark L. Jordan and Justin La of the University of California, Irvine, CA, USA

In the opening webinar of the series, Dr. Jordan explained that the National Board of Medical Examiners (NBME) created the USMLE Step exams to inform medical licensure by determining minimum competence for practicing physicians, not to assess an applicant’s qualifications for admittance to residency programs. In contrast to its original purpose, the USMLE Step 1 exam has become integral to the residency application process. Its 3-digit scoring system has provided residency program directors with a reliable source of “hard data” by which applications could be screened and filtered. This objective metric was deemed useful compared to parameters that often lack uniformity between schools, such as grades, class rank, and letters of recommendation. For students, the numerical score of the USMLE Step 1 exam offered a “level playing field,” allowing individuals to distinguish themselves from their peers, regardless of their...
school’s ranking or prestige. Despite the advantages of a numerical scoring system, there is a growing consensus that high-stake exams no longer best serve holistic evaluation of resident candidates. Dr. Jordan highlighted the problem of “USMLE mania,” where achieving a high USMLE Step 1 score has become the predominant goal of preclinical medical education. This contributes to increased student anxiety, parallel curricula, and increased pressure on medical schools to “teach to the test.” Ultimately, the Invitational Conference on USMLE Scoring (InCUS) was convened in March 2019, and recommended that the NBME score the USMLE Step 1 exam using a P/F paradigm.

Dr. La next discussed the USMLE Step 1 P/F transition from the perspective of current residents. He hypothesized that input from current residents contributes significantly to the residency interview process, helping to build successful teams and environments that facilitate learning and sound patient care. Current residents can assess the “fitness” of applicants for residency as well as their “overall fit” amongst the cohort by using non-USMLE components of their residency applications [1]. As mixed data exists regarding the ability of the USMLE Step 1 to predict resident success, Dr. La presented comprehensive literature reviews demonstrating that other selection criteria are heterogeneously studied, with publications differing even in definition and assessment of resident success [2, 3]. In light of these data, Dr. La concluded that there is no single predictive success factor in residency.

Dr. Jordan continued the webinar by addressing the implications of losing the 3-digit USMLE Step 1 score as an objective screening tool for residency candidates. Program directors now face an increased burden in reviewing applications, raising concerns that applicants could receive less in-depth reviews. This change leaves applicants with uncertainty about distinguishing themselves for competitive specialties, particularly if applying from newer, international, or less well-known medical schools. Dr. Jordan then presented the advantages and disadvantages of other evaluation metrics, ending with the question of whether residency programs will place greater emphasis on the USMLE Step 2 Clinical Knowledge (CK) exam. He also noted that many in-person tools for residency candidate assessment have not been feasible due to the SARS-CoV-2 pandemic, such as away rotations, which were relied upon to observe qualities and factors that electronic applications cannot effectively evaluate. This change introduces greater uncertainty regarding how candidates are considered for interviews and assessed for residency.

In conclusion, Dr. Jordan outlined potential solutions: (1) standardized candidate assessment tools based on specialty or nontraditional assessments (e.g., Jefferson Empathy Scale, Grit Scale); (2) standardized Medical Student Performance Evaluations (MSPEs), rotation evaluations, grades, recommendation letters, and transcripts; (3) increased medical school transparency on applicant strengths, professionalism, and performance; (4) program director-driven development of mission-based holistic criteria for application review; (5) Electronic Residency Application Service (ERAS) modification to permit extraction of information for holistic review; and (6) early application cycle to demonstrate program interest.

Pass/Fail USMLE Step 1: Implications for a Foundational Sciences Department

Presenter: Douglas Gould of the Oakland University William Beaumont School of Medicine, Auburn Hills, MI, USA.

In the second installment of this series, Dr. Gould discussed how P/F scoring of the USMLE Step 1 might broadly impact biomedical/foundational sciences departments. While USMLE Step 1 scoring will change, its content focus of foundational sciences will not. Dr. Gould shared the National Resident Matching Program’s 2018 Program Director survey results, presenting metrics that could replace USMLE Step 1 scores in screening applications. According to the survey, the top seven criteria used to rank residency applications were (1) USMLE Step 1 score, (2) Letters of Reference, (3) Medical Student Performance Evaluation (MSPE), (4) USMLE Step 2 CK score, (5) Personal Statement, (6) Grades in required clerkships, and (7) Failed USMLE attempts [4]. Dr. Gould then shared results from an audience poll question, revealing that audience members considered USMLE Step 2 CK scores, grades in required clerkships, and letters of reference most likely to replace USMLE Step 1 scores to screen residency applications. This analysis emphasized the unpredictability of which factors may become most important going forward and underscored the need for medical schools to be flexible. Dr. Gould stressed the potential for preclinical curricula to transition to P/F as well, possibly reorienting towards USMLE Step 2 CK preparation.

While Dr. Gould anticipates a de-emphasis on basic science education going forward, holistic review of residency applications reveals an opportunity for biomedical and foundational sciences to promote strong applications by providing students with more research activities. Dr. Gould showcased the Oakland University William Beaumont School of Medicine’s required longitudinal, 4-year research program called “Embark.” In this program, students are paired with mentors to receive training across multiple research design and implementation domains. Dr. Gould noted that residency programs could use quantifiable student research activities as a metric for offering residency interviews, which will favor the development of research programs.

Dr. Gould concluded with optimism, stating that many of the changes medical education is currently experiencing have been ongoing, although recent events (such as the
SARS-CoV-2 pandemic) have catalyzed and accelerated them. He credited the OUWB’s Foundational Medical Studies Department’s collaborative nature, the presence of faculty already devoted to medical education, and a robust faculty development program with the quick and successful transition to online learning and asynchronous education. Although the number of traditionally structured foundational sciences departments has decreased over the last two decades, overall numbers of basic/foundational science departments have increased 33% during the previous 5 years, with 21% of medical school faculty holding a PhD as of 2018 [5]. As continuous change to medical education is inevitable, Dr. Gould advocated for increasing and developing foundational science medical education faculty who are invested in innovating effective teaching in the face of unforeseen events, such as COVID-19 and the USMLE Step 1 P/F transition.

USMLE Step 1 P/F: A UME Curriculum Dean’s Perspective

Presenter: Jonathan Amiel of the Columbia University Vagelos College of Physicians and Surgeons, New York City, NY, USA.

In this session, Dr. Amiel discussed how the switch to USMLE Step 1 P/F grading could impact a broad range of areas in undergraduate medical education (UME). In terms of the mission, objective, values, and ethics of educational programs and accreditation, he expressed optimism that pre-clerkship faculty may take a more holistic approach to the knowledge and skills students acquire in pre-clerkship curricula. Rather than “teaching to the test,” institutions now have an opportunity to work with the National Board of Medical Examiners (NBME) from a competency-based perspective. While he did not anticipate a significant impact on accreditation in general, Dr. Amiel noted that the NBME might increase the USMLE Step 1 passing threshold, introducing higher failure rates that could affect accreditation. Logistically speaking, institutions will still need to ensure they provide adequate support to learners who have difficulty with standardized tests, including increased support for the USMLE Step 2 CK exam, which may replace USMLE Step 1 as a tool for residency application ranking.

Dr. Amiel predicted that the most significant architectural impact on UME curricula would be changing dedicated study time for the USMLE Step exams. If schools are going to increase USMLE Step 2 CK preparation, dedicated study time for USMLE Step 1 or elective rotation time could be reduced. This shift could result in individualized study tracks or parallel curricula for different graduating classes, resulting in confusing messaging for students. Concerning the potential changes in assessment methods, Dr. Amiel anticipated that faculty development programs would provide new training for faculty. He emphasized this as an opportunity to reframe “assessment of learning” as “assessment for learning” and help faculty broaden their thinking about testing medical knowledge early in the curriculum. He discussed the potential impact of the USMLE Step 1 P/F shift on resource management, as the process of making changes to curricula is resource intensive. Students may also be required to take USMLE Step 2 CK earlier to obtain scores needed for residency applications, which would necessitate UME curricula preparing students earlier. Schools that schedule the USMLE Step 1 exam later in students’ training (e.g., Columbia University) will need to consider shifting both the USMLE Step 1 and Step 2 CK exams earlier. Specialties may also begin to develop and administer their own exams to rank applicants, which would further encourage changes to UME curricula to help students be prepared and competitive for their desired specialties.

Overall, Dr. Amiel suggested that scoring the USMLE Step 1 as P/F will be positive for medical education. However, stakeholders must be careful to mitigate any potential negative, unintended consequences. He stressed the importance of basing initiatives in education theory and best practices, promoting competency-based medical education and cooperation with the NBME and USMLE programs. He recommended that students should be able to demonstrate their competency in a basic body of knowledge and clinical reasoning skills before moving to the clinic, yet still have experience with high-stake assessments. Finally, Dr. Amiel cautioned that any new methods of programmatic assessment should be mindful of holistic review processes and how generating data about learner performance may impact residency applications. In particular, institutions must be careful to employ anti-racist pedagogy and assessment approaches. He concluded with the suggestion that the USMLE Step 1 P/F shift may result in an abundance of opportunities for conducting scholarly activity.

USMLE Step 1 Going Pass-Fail: Are We Just Kicking the Can Down the Road?

Presenters: Bruce Morgenstern of the Roseman University College of Medicine, Henderson, NV, USA; Brenda Roman of the Wright State University Boonshoft School of Medicine, Dayton, OH, USA.

As the immediate past president and current president of Alliance for Clinical Affairs, respectively, Drs. Morgenstern and Roman discussed endemic issues within the current residency application process that will likely remain after the USMLE Step 1 scoring changes to P/F. Dr. Morgenstern began by contrasting the USMLE’s purpose of informing licensure by assessing physicians’ medical knowledge, values, skills, and attitudes with the off-label use of the exams to rank residency applications. He challenged the ability of multiple-choice questions to assess abstract qualities such
as values, skills, and attitudes, noting a similar dilemma for medical education: the ultimate goal of medical education is to produce “good” physicians, but the field struggles to articulate how and when an individual is a “good” physician. In response, competency-based medicine is emerging, but many areas commonly used to measure competence, such as professional identity development, professionalism, values, and attitudes, are difficult to assess, especially with the intense focus on USMLE Step 1.

Dr. Morgenstern acknowledged that although USMLE Step 1 scores do not correlate with other measures of overall success in residency, the practice of developing interviewee lists using USMLE Step 1 scores is generally accepted. Further contributing to the reliance on ranking tools is the lack of staffing for residency program directors to support holistic application review. Program directors will feel pressured to select applicants likely to pass specialty boards. They will seek mechanisms to rank applicants meaningfully since subjective metrics often lack consistency and undergo “modifier inflation.” The use of modifier inflation leads to a general lack of trust between UME and GME. He proposed that the current residency application process itself may not be fair, stating that ERAS contributes to the overwhelming number of applicants. The National Residency Match Program algorithm is proprietary, so it is impossible to verify whether/how it favors applicants or programs. Going forward, how can program directors screen applicants and conduct holistic reviews? Dr. Morgenstern emphasized that it is incumbent upon UME to develop tools that are more intentional in their design and purpose to assess “good” residents compared to the minimum competency licensure exams previously coopted for the match process.

Dr. Roman then continued the webinar by further exploring problems that will remain inherent to the residency match process despite the USMLE Step 1 P/F transition, such as the issue of student stress. Student stress will probably shift from USMLE Step 1 to USMLE Step 2 CK if program directors focus more on USMLE Step 2 CK as an application screening tool. Clerkship years may be negatively affected, detracting from patient care and clinical learning. Medical schools and programs will still be evaluated by match rate, with the job performance of Student Affairs Deans tied to successful match rates. Program directors may still negatively evaluate applicants with well-explained gaps in their education if they required more time to complete medical school, and students and programs will still try to “game the system” to their advantage. Despite these issues, Dr. Roman shared that almost all students successfully match in a residency program.

As the webinar drew to a close, Dr. Roman reminded the audience that medical schools could improve the residency application process and trust between UME and GME by developing better ways to identify and address professionalism issues. This may include involving basic science faculty early on in medical curricula to help detect concerning patterns of unprofessional behavior. She left the audience with a provocative proposal intended to spark thought and discussion. With a disclaimer that she did not intend the audience to consider the idea in its current form seriously, Dr. Roman suggested that programs should define an optimal fit for residents in their programs and applicants should define their preferred residency characteristics. The matching algorithm then should prioritize compatibility between applicant and program. Dr. Roman also suggested that as a field, medicine should ask media publications to stop ranking schools and hospitals, as this creates false realities that may drive students to apply to too many residencies. Ultimately, UME and GME must work together more closely to avoid simply moving the stress from USMLE Step 1 to USMLE Step 2 CK, and to involve both clinical and basic science faculty in helping program directors assess applicants.

**USMLE Step 1: Osteopathic and International Perspectives**

Presenters: John (Jack) Boulet of the Educational Commission for Foreign Medical Graduates (ECFMG) (retired), Philadelphia, PA, USA; Robert Cain of the American Association of Colleges of Osteopathic Medicine (AACOM), Washington D.C., USA.

Dr. Boulet began the webinar by presenting the likely impact of the recent changes to the USMLE Step Series exams on international medical graduates (IMGs) applying to United States (U.S.) residency programs. For IMGs, the USMLE Step exams carry additional weight beyond their use in the residency application screening process as they are also necessary for Educational Commission for Foreign Medical Graduates (ECFMG) Certification. The Accreditation Councils for Graduate Medical Education (ACGME) will not allow their programs to admit IMGs without ECFMG certification. Previously, ECFMG Certification required applicants to successfully complete the USMLE Step 1 exam, USMLE Step 2 CK exam, and the now-suspended USMLE Step 2 Clinical Skills (CS) exam. The ECFMG now allows applicants to demonstrate clinical skills through alternative avenues, including having a license to practice medicine in another country or passing a standardized clinical skills exam for medical licensure. Additionally, students may meet this requirement by graduating from a medical school that meets specific accreditation, loan-program, or partnership requirements. Finally, IMGs must also pass the Occupational English Test in medicine to assess the communication requirements formerly tested by USMLE Step 2 CS.

Dr. Boulet next shared that, on average, IMGs’ match rate with U.S. residencies is approximately 60% (compared to >90% for medical graduates trained in the U.S.), and
thus, they tend to submit a larger number of applications. This increases competition for interviews. As IMGs can no longer depend upon strong numerical USMLE Step 1 scores to distinguish themselves, they must now reconsider how to stand out from their peers. Dr. Boulet suggested that using the USMLE Step 2 CK exam to screen residency applicants may disadvantage certain IMGs, as their schools may not offer early clinical experiences or be able to add them. Schools with a small minority of graduates applying to U.S. residencies may not restructure their curriculum, and U.S. residency program directors may rely on the international reputation of medical schools. Given these concerns, IMGs’ motivation to secure U.S. residency programs could decrease. Dr. Boulet concluded by emphasizing the need to reevaluate the residency selection process, supporting efforts by the Coalition for Physician Accountability, the ECFMG, and the National Board of Osteopathic Medical Examiners (NBOME) to make the UME to GME transition more efficient.

Dr. Cain continued the webinar by sharing the overall impact of USMLE Step 1 P/F scoring on osteopathic medicine. Despite the osteopathic community having its own independent licensing body, the NBOME, many residency programs require doctor of osteopathic medicine (DO) students to take the USMLE Step 1 to be considered for residency interviews and/or admission. This requirement is associated with high financial, personal, relational, and opportunity costs to students, and osteopathic medical schools feel the impact of the “USMLE Step 1 mania” on their preclinical curricula. These additional costs raise ethical questions regarding the fairness of requiring students to take two licensure exams, particularly considering the USMLE Step 1 transition to P/F scoring. When the NBME scored USMLE Step 1 numerically, DO students experienced the benefit of distinguishing themselves amongst other DO, MD, and IMG applicants. However, this benefit disappears with P/F scoring, particularly since the Comprehensive Osteopathic Medical Licensing Examination (COMLEX) Level 1 and the USMLE Step 1 test the same material and the COMLEX also fulfills medical licensure requirements.

Next, Dr. Cain presented the advantages and disadvantages of USMLE Step 1 P/F scoring for osteopathic medical students. Decreased stress and cost for students are advantageous, along with schools experiencing less pressure to “teach to the test.” Reduced pressure to score highly on an assessment designed for allopathic medical education aligns with and supports the development of the osteopathic professional identity. Disadvantages include concerns regarding reduced transparency of the residency application screening process and whether program directors will rely on school reputation or their familiarity with the individual DO UME programs. Finally, a greater portion of DO students than MD students pursue clinical clerkship training in non-hospital-based settings (i.e., ambulatory and physical rehabilitation training), which disadvantages them from scoring as highly on the USMLE Step 2 CK exam. For DO students, this raises the concern that the USMLE exams have been misused as tools for residency placement, particularly since there are differences in performance compared to allopathic students.

In contrast, the COMLEX exams are designed to test the osteopathic-specific components at each stage of student development and are embedded throughout the 4-year UME experience. To summarize the osteopathic community’s general sentiment, Dr. Cain stated that the change of the USMLE Step 1 scoring to P/F could harm DO students when there should be no impact. This outcome highlights a great need for continuing education around the use of the COMLEX exam and understanding different tools for determining residency readiness.

Coalition for Physician Accountability
UME-GME Review Committee

Presenter: Elise Lovell of the University of Illinois at Chicago, Chicago, IL, USA.

During the sixth and final webinar of this series, Dr. Lovell, Co-Chair of the Coalition for Physician Accountability UME-GME Review Committee (UGRC), provided a front-line update of the UGRC’s work to review the UME to GME transition. Dr. Lovell first introduced the Coalition for Physician Accountability as a collaboration between national organizations responsible for the education, assessment, accreditation, certification, and licensure processes ranging from UME to practicing physicians. Member organizations include the Association of American Medical Colleges (AAMC), the American Association of Colleges of Osteopathic Medicine (AACOM), the American Board of Medical Specialties (ABMS), the Accreditation Council for Continuing Medical Education (ACCM), the Accreditation Council for Graduate Medical Education (ACGME), the American Medical Association (AMA), the American Osteopathic Association (AOA), the Council of Medical Specialty Societies (CMS), the Educational Commission for Foreign Medical Graduates (ECFMG), the Federation of State Medical Boards (FSMB), the Liaison Committee of Medical Education (LCME), the National Board of Medical Examiners (NBME), and the National Board of Osteopathic Medical Examiners (NBOME) as well as members of the public. The mission of the coalition is “to advance health care and promote professional accountability by improving the quality, efficiency, and continuity of the education, training, and assessment of physicians” [6].

To provide the audience with context for the development of the coalition’s UGRC, Dr. Lovell revisited the recommendations resulting from the Invitational Conference on
UME Scoring (InCUS) in March 2019: (1) consider score reporting changes, e.g., pass/fail, composite score, and categorical/tiered scoring for USMLE exams; (2) research how USMLE is (or is not) related to performance in residency and practice; (3) continue work to address group (racial/ demographic) differences in USMLE exam scores; and (4) undertake a comprehensive review of the UME to GME transition process. While the first recommendation regarding the USMLE Step 1 P/F scoring has received significant attention, the USMLE Step 1 numerical score was only one factor of many contributing to the broader problems surrounding the UME to GME transition. After briefly discussing progress made towards meeting the first three recommendations, Dr. Lovell focused on reviewing the current UME-GME transition system.

The UME to GME review process was anticipated to involve eight significant areas of focus: (1) improving the application process, (2) reducing the number of applications submitted by applicants, (3) improving transparency of the UME to GME transition process, (4) identifying potential adjustments to the Match, (5) improving program directors’ ability to more holistically review applicants, (6) improving the trust/ transparency of medical school-based assessments, (7) reviewing the role of standardized testing in UME to GME transition, and (8) developing assessments for other essential measures beyond knowledge. InCUS identified the Coalition for Physician Accountability as the organization best suited to address this review. In September of 2019, the coalition formed a planning committee to identify the UGRC membership and determine its key questions, deliverables, milestones, timeline, and budget.

Membership of the UGRC comprises representatives from each of the coalition member organizations, medical students, resident physicians, members of the public, and support staff. Along with Dr. Lovell, Dr. George Mejicano was selected to co-chair the UGRC. Areas of focus for the UGRC include fit of applicants for residency programs based on abilities and interests, an increase of trust between medical schools and residency programs, and mitigation of reliance on licensing exams during the application process. Also, the UGRC’s focus encompasses ensuring fairness for all types of applicants (allopathic, osteopathic, and international medical graduates) while considering cost and the potential for systemic bias and diversity issues throughout the transition from UME to GME. The coalition has requested that the UGRC also review and consider specialty-specific competencies, including when and how specialty-specific preparatory training for residency should occur and where responsibility should rest in resolving any gaps in residency readiness.

Once established, the UGRC created workgroups to address the charges from the coalition planning committee and added overall considerations of diversity, equity, inclusion, fairness, student wellbeing, and the public good. The UGRC asked the workgroups to plan for continual modification and support the formation of an iterative review process of the UME-GME transition going forward. Dr. Lovell concluded by inviting the audience to follow UGRC meeting updates on http://physicianaccountability.org/, providing an expected timeline for delivering the UGRC’s recommendations. These draft recommendations will be reviewed by the UGRC in March of 2021 and made available for public review in April of 2021. The UGRC expects to deliver its final recommendations to the Coalition for Physician Accountability in May of 2021.

We wish to thank each of our presenters for their contributions to this series!

References

1. Nemani VM, Park C, Nawabi DH. What makes a “great resident”: the resident perspective. Curr Rev Musculoskelet Med. 2014;7(2):164–7.
2. Yang A, Gilani C, Saadat S, Murphy L, Toohey S, Boysen-Osborn M. Which Applicant Factors Predict Success in Emergency Medicine Training Programs? A Scoping Review. AEM Educ Train. 2020;4(3):191–201.
3. Bowe SN, Laury AM, Gray ST. Associations between otolaryngology applicant characteristics and future performance in residency or practice: a systematic review. Otolaryngol Head Neck Surg. 2017;156(6):1011–7.
4. National Resident Matching Program. Program Director Survey Results. 2018.
5. Association of American Medical Colleges. 2018 U.S. Medical School Faculty Report. https://www.aamc.org/data-reports/faculty-institutions/interactive-data/2018-us-medical-school-faculty
6. Mission, Coalition for Physician Accountability. https://physicianaccountability.org/about/

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.