

*Summary Report and Preliminary Recommendations from the
Invitational Conference on USMLE Scoring (InCUS), March 11-12, 2019*

Executive Summary

The United States Medical Licensing Examination® (USMLE®) serves as the primary assessment tool to help inform licensing decisions by state medical boards. In addition to state medical boards' use, others have developed uses of USMLE scores over the past quarter century. Examples include score use for learner assessment and program evaluation within undergraduate medical education (UME) and, over time, the increased use of USMLE scores as a screening tool for selection into graduate medical education (GME). The latter has recently drawn increasing scrutiny and criticism.

Multiple factors are contributing to the reliance on the USMLE score for residency screening and selection, including the increasing UME shift to pass/fail grading systems, and the total number of GME applicants outpacing slots available through the National Resident Matching Program. This latter factor closely relates to the current trend of applicants submitting growing numbers of residency applications, per applicant, year after year, and the administrative challenge facing program directors having to sort through what could amount to two orders of magnitude more applications than available program positions. This has led to a current environment characterized by medical students' efforts to maximize their USMLE scores (particularly Step 1) at the potential expense of focusing on other educational/curricular offerings. There are also growing concerns over the role of high stakes examinations and their effect on student well-being.

The Association of American Medical Colleges (AAMC), American Medical Association (AMA), the Educational Commission for Foreign Medical Graduates (ECFMG), the Federation of State Medical Boards (FSMB) and the National Board of Medical Examiners (NBME) convened a multi-stakeholder Invitational Conference on USMLE Scoring (InCUS) in March 2019 to explore these issues and consider recommendations specific to USMLE score reporting and the broader system of transition from UME to GME. A general consensus emerged: The current UME-GME transition system is flawed and not meeting the needs of various stakeholders; unilateral changes to USMLE alone will not "fix" the overall system; and changes, both systemic and specific to USMLE, must be identified and implemented on a reasonable timeline.

Several guiding principles emerged from InCUS discussions. These included seeking changes that will improve examinee and physician well-being; improving the UME-GME transition while limiting negative consequences to any one stakeholder group; allowing time to prepare for, and mitigate, any possible disruptive changes while acting with a sense of urgency; maintaining a licensure system in which both US and international graduates are rigorously assessed; and further promoting holistic review of residency applicants.

Recommendations specific to USMLE:

- 1) *Reduce the adverse impact of the current overemphasis on USMLE performance in residency screening and selection through consideration of changes such as pass/fail scoring.*
- 2) *Accelerate research on the correlation of USMLE performance to measures of residency performance and clinical practice.*
- 3) *Minimize racial demographic differences that exist in USMLE performance.*

Recommendations to the UME-GME transition system:

- 1) *Convene a cross-organizational panel to create solutions for the assessment and transition challenges from UME to GME, targeting an approved proposal, including scope/timelines by end of calendar year 2019.*

Background

The United States Medical Licensing Examination (USMLE) —co-sponsored by the Federation of State Medical Boards (FSMB) and the National Board of Medical Examiners (NBME)—is a three-part, four-examination series of assessments providing state medical boards with information regarding a medical licensure candidate’s knowledge and skills to practice medicine in the United States and its territories.

Typically, among students enrolled in a medical education curriculum and continuing into residency, USMLE Step 1 is taken after foundational medical school curricula; USMLE Step 2 Clinical Knowledge (CK) and Clinical Skills (CS) are taken prior to a residency program applicant submitting a rank order list to the National Resident Matching Program (NRMP) system; and USMLE Step 3 is taken during residency. International medical students and graduates (IMGs) often demonstrate a different timing pattern in their interaction with USMLE. Since USMLE’s inception in 1992, scores on its computer-based exams (Steps 1, 2 CK, 3) have been reported on a numeric (3-digit) score scale, relative to a minimum passing score. This score is set by USMLE “national faculty” committees that consist of representatives from state medical boards, undergraduate medical education (UME), graduate medical education (GME) and the public. The standardized patient-based Step 2 CS (added to USMLE in 2004) is the only examination in the series currently reported as Pass/Fail. (1) A candidate for physician licensure in the United States must score at or above a minimum passing score on all components of USMLE to be considered eligible for state medical licensure. USMLE performance in the context of a state licensing decision is the primary purpose of the USMLE. There exists a separate path to state licensure for osteopathic physicians.

Problem Identification

Over the USMLE’s more than 25-year history, other uses of USMLE numeric scores have emerged. Prior NBME exams were subject to similar secondary uses. In the early 1990s a series of articles authored by members of the academic medical community and NBME staff were published. (2, 3, 4, 5) These provided detail on secondary uses of USMLE scores and cautioned against the use of USMLE as a sole measure of competency, given the test’s principal design for

maximum precision at the pass/fail point. Said another way, the exams were developed as medical licensure examinations and not as academic achievement exams. Current secondary uses are similar, and relate to medical schools' use of USMLE scores in promotion and advancement of their students, as well as the use of scores for curriculum evaluation or as part of institutional self-study and accreditation reviews. USMLE scores are also increasingly used by residency program directors for screening and/or selection of residency applicants. (6) The USMLE program did not select its score scale or the design of its numeric score reporting with these secondary uses as a major consideration. (7)

Despite known limitations of using USMLE scores for residency screening and selection, program directors and some other stakeholder groups have expressed a need for nationally standardized assessments of knowledge, skills and behaviors, in part because the current UME to GME transition system lacks standardized assessments other than the USMLE. The USMLE's standardization is one of the reasons residency program directors view the USMLE as a trusted assessment for screening both domestic and international medical school applicants and graduates. (8) This may be particularly true given the following trends impacting the UME-GME transition:

- Since 2010, the number of matriculants to M.D. and D.O.-granting medical schools in the United States has increased 23.3%, from 24,093 to 29,710 individuals. (9,10)
- The number of applicants for residency positions has risen every year for more than a decade, reaching a high of 38,376 applicants in 2019, competing for 32,194 PGY-1 positions.
 - There remain, however, more positions available in the Match than US senior medical students. In 2019, 18,925 active applicants were US seniors from M.D.-granting schools. There were 6,001 D.O. candidates who submitted rank order lists. (11)
- The average number of residency programs to which Electronic Residency Application Service (ERAS) applicants apply continues to rise — from 78 applications per applicant in 2014, to 90 in 2018: an increase of 15%. (12) This has occurred despite available tools from the AAMC aimed at helping applicants identify an “optimal” number of applications, which typically ranges from 15-55 applications per applicant. (13)
- Meanwhile, M.D. and D.O.-granting medical schools in the United State have continued to move from tiered grading systems to pass/fail grading or other broad categorical grading. This is occurring at all levels of the curriculum, from pre-clerkship settings to 4th year. (14, 15, 16, 17, 18) Changes have been made for many justifiable reasons, including efforts to enhance a competency-based approach to medical education, as well as concerns for grading bias, student well-being and limited reliability of available assessments. Nevertheless, this has impacted program directors' ability to have school-level comparison data.
- International medical graduates, U.S. citizen and non-U.S. citizen, make up a significant proportion of matched applicants, and a critical pipeline of the U.S. physician workforce, particularly in some specialties such as internal medicine. (11) Many of these physicians view the standardized assessment which the USMLE provides, a means to compete fairly for a residency position.

As a result of the above factors and other influences, students continue to focus on maximizing USMLE scores, given their awareness of the use of scores as a residency screening tool. Among some specialties, there is also a growing trend among program directors to screen applicants based on Step 2 CK scores. (6) This has led to significant impact on medical schools and medical students, including:

- Potentially negative impact of preparation for high-stakes testing on student well-being. (19, 20)
- Concern from schools that students' focus on USMLE numeric scores limits their ability to enact meaningful curricular change. (21, 22)
- Concern that students pay less attention to the development of competencies that are either not assessed, or less emphasized, on USMLE. (21, 22)
- Concern from students that a mismatch exists between curricular content at their medical school and what is assessed on USMLE. This has been referred to as the "parallel curriculum." (23, 24)

Historical Aspects of USMLE Score Reporting

The practice of numeric score reporting has been considered formally by the USMLE program on three prior occasions. At the program's inception in the early 1990s, there was a decision to report scores on a three-digit score scale. At that time, there was acknowledgment among USMLE's sponsors that scores might be used for reasons other than licensure, despite the USMLE's caution that scores not be used for purposes beyond licensure (2). In 1998, NBME undertook a comprehensive survey of stakeholders' views on USMLE score reporting. Results were interpreted as showing a trend toward favoring continued numeric score reporting. There were differences among respondents, however. For example, among responding examinees, there was a desire that USMLE scores be reported in numeric fashion, but reported to schools and residency directors only in Pass-Fail form. Ultimately, the analysis led USMLE governance responsible for scoring policy to decide that there was "no basis ...for changing" existing scoring policy. (25) Lastly, during the comprehensive review of the USMLE program in 2007-08, a change in the score reporting system was considered but not enacted. The Committee to Evaluate the USMLE Program (CEUP) recognized the varied opinions and uses of scores, but did not provide a recommendation on any specific changes to score reporting, stemming in large part from the extent of other significant programmatic and technical changes that USMLE was simultaneously undertaking, and the unclear impact that such changes might have on scoring. The CEUP did note that other uses of examination results may be recognized provided they did not compromise the primary purpose of USMLE. (26)

Today, among individuals as well as stakeholder organizations – including the USMLE co-sponsors (NBME and FSMB) -- there is broad recognition of the current uses and challenges regarding USMLE numeric score reporting. Many feel that the focus on USMLE numeric scores represents a symptom of larger existing problems in the system of transition from undergraduate medical education to graduate medical education.

Invitational Conference on USMLE Scoring (InCUS)

Conference Origins

A national conversation has reemerged on USMLE score reporting, focusing in particular on the use of USMLE Step 1 and Step 2 CK scores for residency screening and selection. This national discussion gained clarity at a meeting convened by the American Medical Association for their Accelerating Change in Medical Education consortium schools, hosted in April 2018 by the Warren Alpert Medical School of Brown University, Providence, RI.

After the conclusion of this meeting, the chief executive officers (CEOs) of five stakeholder organizations agreed to co-sponsor a subsequent invitational conference on USMLE score reporting. Staff responsible for InCUS began planning as quickly as feasible under the sponsorship of the American Medical Association (AMA), the Association of American Medical Colleges (AAMC), the Educational Commission for Foreign Medical Graduates (ECFMG), and the USMLE parent organizations, the Federation of State Medical Boards and the National Board of Medical Examiners.

The expressed goal of the InCUS conference was to collaboratively review the USMLE program's practice of numeric score reporting within the context of its primary use of licensure, and to discuss any secondary uses and the broader regulatory and educational environments in which USMLE exists. The deliverables were to be recommendations pertaining to the USMLE program as well as recommendations to the broader environment of the UME and GME transition. It was felt that the group of five co-convening organizations could identify and assemble the stakeholder voices necessary to carefully examine USMLE score reporting in the context of physician licensure as well as the UME to GME transition.

InCUS Planning Committee

The CEOs in turn designated a five-person planning committee with representatives drawn from each of the co-sponsoring organizations. The charge to the planning committee called for them to develop, organize and execute an invitational conference to facilitate open dialogue and exchange of ideas on the complex, interconnected environment within which USMLE numeric scoring plays a role in the transition of physicians into GME.

The planning taskforce began work through the summer and fall of 2018 to develop a framework for the conference. The taskforce identified necessary stakeholder perspectives required at the conference, desired outcomes (envisioned broadly as guiding principles; short-and long-term recommendations), timelines, etc. This framework also delineated the decision-makers specific to certain areas (e.g., USMLE program governance and boards of FSMB and NBME for USMLE-specific matters). Ultimately, this framework became the basis for the materials presented at <https://www.usmle.org/inCus/>.

Pre-conference Preparation

InCUS targeted key stakeholder groups for representation: medical educators drawn from both the UME and GME communities; representatives from state medical boards; medical students and residents representing examinees who have taken USMLE (both U.S. and international); numerous relevant groups or organizations (e.g., National Resident Matching Program, American Board of Medical Specialties, etc.); and members of the public.

The co-sponsors envisioned a conference designed to facilitate open dialogue and mutual exchange of challenges and perspectives across multiple stakeholder groups. Understandably, the number of individuals interested in participating in the conference greatly exceeded what would be optimal for a conference structured to maximize personal interaction and small group discussion. Thus, the co-sponsors structured the conference as an invitational event with approximately 45 invited guests, representing a broad view of stakeholder opinions on this issue. Along with key staff members from the five organizations, the total attendance was approximately 65 individuals. InCUS took place at NBME headquarters in Philadelphia to reduce meeting costs. Attendees received no honoraria; monetary reimbursement was limited to hotel and travel-related expenses. Background information on InCUS as well as a list of conference attendees is available at: <https://www.usmle.org/InCUS/>.

In recognition of the significant interest in InCUS, and given the attendance limitations, the planning committee developed a mechanism for gathering pre-conference input. Commentary, including position statements if such existed, was solicited from over 200 professional organizations, societies and state medical boards in advance of InCUS. All comments were then collated by the planning committee and shared as pre-read materials for conference attendees. To facilitate informed and productive dialogue, staff conducted a literature review focusing on research related to USMLE performance and outcomes. Most studies pertained to how scores on USMLE Step 1 and Step 2 CK relate to specialty board certification, measures of clinical performance, or state board disciplinary action. Articles were selected for merit regardless of whether or not they shined a favorable light on the USMLE or the program's existing scoring policy. Summaries were created for studies published since 2008, and represented research from diverse institutions employing varied relevant datasets and methodological techniques. The summaries provided an overview of the current state of available empirical knowledge about the pros and cons associated with contemporary uses of USMLE Step 1 and Step 2 CK scores, with the intent of fostering discussion and an empirically grounded set of recommendations emerging from the InCUS. A comprehensive overview of any validity argument for USMLE goes beyond the intended scope of this InCUS conference report. The USMLE literature search, however, is available at <https://www.usmle.org/inCus/#additional>

In addition, a sampling of relevant Commentaries and Letters to the Editor related to USMLE Step 1 and Step 2 CK scores, and their interpretations, was assembled to provide examples of the types of conversations that are playing out in the medical literature and at medical and educational meetings.

Conference Description: InCUS, March 11-12, 2019

The co-sponsoring organizations secured the services of David Baum and Associates, a conference facilitator with extensive national and international experience in dealing with challenging, sometimes contentious, and complex issues. The two-day conference was structured to move attendees through three phases of activities designed to *Connect, Inform, and Advise*.

Attendees participated with a shared agreement that (1) all ideas were valid; (2) listening deserved top priority; (3) positive intent would always be assumed; and (4) brevity in spoken comments was the best mechanism to keep the dialogue moving. In addition, attendees were asked to share personal experiences/observations rather than attempting to represent the collective outlook of a larger group of stakeholders. This approach was adopted as it was felt that the necessary stakeholder groups were present at InCUS, and that such an approach would allow for a mindset among participants which embraced change and possibilities.

Connect

The diverse stakeholder communities represented at InCUS meant that attendees would likely be engaging with individuals not personally known to them. Accordingly, the initial focus centered upon connecting this diverse group of individuals to each other in order to create a degree of familiarity and comfort in sharing personal viewpoints and/or experiences specific to the USMLE program, medical education and medical regulation. The conference attendees and support staff moved through a series of exercises intended to establish personal relationships as a necessary first step toward trust and shared open dialogue. This exercise highlighted the personal experiences and strengths of the participants, providing context and value for each individual's comments during the meeting.

Inform

This brief phase was designed to ground all conference participants in a common informational base for discussion. Attendees were provided with a brief history of medical licensing examinations in the United States, as well as information on current assessment trends within medical education and current data pertaining to the UME to GME transition.

Advise

This final phase of the conference was viewed by the co-sponsors as critical to the outcome of InCUS. Every conference participant, including learners and trainees, was given an opportunity to respond to each of the following six questions, designed to explore USMLE numeric scores within the current UME and GME environments. Each participant was given five minutes to answer each of these questions. The answers were detailed and provided deep personal insight. The subsequent oral "report-out" focused on summarizing responses to these questions. Naturally, given the personal perspective that these questions sought, responses varied quite a bit. The major themes resulting from each question's responses, however, grounded all participants in areas of potential consensus as well as areas where consensus was lacking. The full extent of the responses to these questions goes beyond the scope of this document. The sample responses below reflect frequency of similar themes as well as importance (described as "amplitude" at the InCUS meeting).

The Six Questions

Q1: What is at the heart of your challenge?

- USMLE has a well-defined purpose, but the score has been co-opted for other uses. USMLE performance is uniformly available (particularly Step 1) and standardized; therefore program directors use it. A great deal of other information is available on residency program applicants, but wide variations in availability and comparability make it less trusted by program directors.
- Any changes made to USMLE scores and related systems must ultimately serve the public good.
- There is a myth that a single number is useful to define overall competence and to select residents. A much larger context than Step 1 scores alone must be considered.
- This is about the broader ecology influencing the UME-GME transition. Change management across the system will be needed.
- Residency program directors are overwhelmed with the numbers of applications vs. the number of available residency positions. Program directors have very limited ways to compare applicants.
- The health and wellness of students and physicians in training is at stake. The current system of UME-GME transition, of which USMLE scores are a part, is harming people, and we have a responsibility to address this.

Q2: When considering the UME-GME transition system as a whole, what is no longer working, needs to be changed, needs to be maintained, and why? What is your tolerance for disruption of the current system?

- Overall, the current process does not allow for the assessment of a medical student's performance and where a student stands within their class.
- Many factors contribute to our current dilemma, and USMLE has become the focus for the larger issues in the UME-GME transition, for issues such as lack of transparency and trust in assessments.
- There was uniform agreement that the exam itself is a valid measure of some competencies, but that it does not predict all aspects of physician practice.
- Tolerance for change varies based on stakeholder views.

Q3: What about the current numeric score reporting process do you value and why? What about the current numeric score reporting process do you not value, and why?

Value

- Provides a point of comparison for all candidates (US and IMG), including across US schools with varied curricula
- Reliable, valid, and psychometrically sound

Not-Value

- Negatively impacts diversity based on known group differences in performance
- Demeans, devalues, or hijacks the pre-clinical curriculum
- The score can become associated with self-worth and can adversely and disproportionately affect career choice or career progression

Q4: Is it reasonable for the GME community to seek a common denominator encompassing all applicants (both US and IMG)? If so, how would you describe the ideal common denominator? How would you describe an acceptable common denominator?

- Views varied on necessity of common assessment for residency screening and selection, based on stakeholder group.
- An ideal common denominator would include more holistic measures of assessment.
- An ideal common denominator would not demonstrate group differences and would allow all candidates to have access to the same resources for preparation (time, money, etc.).
- Views varied a great deal on acceptability of USMLE as a common denominator based on stakeholder group.

Q5: As we seek compromise and possibility, what are the top two considerations that should drive our pathway forward?

- Our improvement plans should be careful, deliberate, and well-paced, and should make clear the problems they are trying to solve.
- The proposed solutions should do no harm and should be fair to all examinees.

Q6: What is the best advice you have for the decision makers as they go forward?

- Always do what is best for patients.
- Address the real issue: It is not the fact that the exam is a scored exam, it is how the score is used.
- Define the end goal: Is this controversy about exam scores only, or is the issue related to the UME-GME transition?
- It is hard to just do one thing – we should take on the bigger issue rather than the smaller issues.

The final activity at InCUS utilized a “village fair” format allowing participants to directly provide advice, recommendations and suggested ideas.

Ten stations, each representing a potential path forward, were available for participants to provide input. Stations 1-6 focused on gathering ideas, options and recommendations either specific to USMLE or to the GME selection process. Stations 7-10 solicited ideas for additional input and suggestions for communications specific to InCUS. Every station received significant input from attendees. Bullet points capturing discussions at each of these stations are provided below.

Station 1: Shift all or part of USMLE to pass/fail

- Changing Step 1 to pass/fail and maintaining a score on Step 2 CK.
- Maintaining Step 1 (i.e., an assessment of the scientific foundations of medicine) as part of physician licensure, given the unique education of physicians and the relevance of science to clinical practice.
- Conversations at this station led to suggestions regarding exam design changes to USMLE for Step 1, Step 2, Step 3, and/or combining some of these steps.

Station 2: Explore other scoring options

- Categorical scoring (e.g., quartiles, quintiles, etc.; binary with cut-off for examinees at risk of board certification failure).
- Reporting scores for the constructs measured on Step 2 CS – making Step 2 CS a scored examination.
- Composite score across USMLE “Decision Point 1” – including scored performance information from Step 1, 2 CK and Step 2 CS.
- Including data from additional sources outside of USMLE (schools, GME, etc.) into USMLE.

Station 3: Options related to score release/disclosure, including timing of score reporting, how scores are shared in GME selection process

- Changes in timing of score release – and to whom scores are released
- Making numeric reporting optional

Station 4: Explore opportunities to improve existing tools for GME screening and selection

Attendees categorized existing tools to include:

- USMLE score/performance
- Medical Student Performance Evaluation (MSPE)
- Letters of recommendation/ Standardized Letter of Evaluation (SLOE)
- Assessment of research, community service experience, etc.
- Perceived geographic or programmatic preferences of applicants
- Medical school reputation
- Audition/visiting rotation assessments
- Applicant personal statement
- Demographic characteristics

Suggestions for new or changed tools included:

- Development of a peer review process for the MSPE to level set and to limit conflicts of interest and promote adherence to MSPE guidelines
- Promotion of a peer review process for the electronic SLOE and similar letters of evaluation
- Continued development of newer modalities such as AAMC Standardized Video Interview and Situational Judgment Tests
- Building additional sorting / analytic tools into the residency application system. Such tools could give program directors the ability to sort applicants on overall/holistic profiles consisting of many measures, not simply single measures, with the intent of finding applicants most suitable to their programs (based on program’s objectives/strengths etc.).

Station 5: Explore development of new tools for GME selection

- Developing nationally comparable assessments for the competencies that USMLE currently does not capture.
- Valid and predictive measures in areas that we most care about are needed: situational judgment, self-awareness, emotional intelligence, how students identify knowledge gaps, seek information and solve problems.

- A measure or process is needed that supports diversity and differentiation across all competencies – which would allow students to bring forward their best attributes.
- Students should be assessed in ways that reflect the work that residents/physicians must do.

Station 6: Changes that could be made to residency selection process

- A matching process that allows better alignment with program attributes and student attributes should be developed.
- Limit number of residency applications per applicant.
- Improve the quality/reliability of information coming from schools and programs alike.
- Multistage resident matching process – first phase, second phase – distributed over time.
- Selection or pre-matching based on “fit” or visiting rotations.
- More explicit rules for interview invitations/acceptances, etc.

Station 7: What further input, if any, should be sought?

Discussions focused on areas such as what makes a good doctor, how we will assess any changes to the residency screening and selection system, what will we measure and how will we know if it is successful. Suggestions included:

- Making sure all stakeholder groups are adequately listened to.;
- Learning from experiences with registries/licensure in international settings (Canada, Australia, etc.) and other professions.
- Better understanding of what program directors want/expect – i.e., the mismatch between an educational model and a workforce training model.
- More information from program directors to schools and applicants.

Station 8: What would you wish us to say publicly about this conference?

Communication should be swift and clear to drive transparency, assuage anxiety (particularly of examinees), and continue the momentum around the discussion and potential for change created by the conference. Additional comments included:

- Emphasize that the issues are complex, and changes cannot be made overnight.
- Acknowledge that this problem is a systems issue, and “ownership” needs to be shared and communicated.
- Communicate that:
 - Any decision(s) will be in support of the training and identification of good doctors in the service of the health of the public;
 - The InCUS conference was about inclusivity and many voices were heard in a congenial and collaborative forum. The plan for public comment is an important aspect of inclusivity in this process.

Station 9: What do you want communicated to the USMLE parent organizations?

The most overarching message was the importance of conveying the passion and intensity of the dialogue at the InCUS conference to the USMLE parents. There was

expressed concern for the “unsustainable” nature of the current system for transition to GME. FSMB and NBME need to try to clarify:

- What is the issue?
- Who has ownership of it?
- Who should lead broader systemic discussions?
- Who makes the decision(s)?

Additional comments included:

- Think of USMLE as a “lever” that can facilitate the start of broader, positive systemic change.

Station 10: What information should be communicated to specific stakeholders?

- Nothing happens unless someone takes the first step.
- Being the first organization to take the first step is challenging and may upset some stakeholders, but some group has to take the first step.
- The status quo is not acceptable.
- Students/examinees need to know and hear that learners were included in InCUS conversations and were engaged in the process. There is a need to continue to engage stakeholders in this discussion. Student wellness was an oft-discussed topic at InCUS and is being taken seriously.

Reflections on the Invitational Conference

Given the at-times contentious nature of previous discussions of USMLE score reporting, the varied opinions, and the lack of consensus among many individual stakeholder groups, InCUS had the potential to be unproductive, particularly given the brief time the invited guests and staff had together. Quite the opposite occurred. Conference feedback from participants noted the collaborative and collegial nature of the dialogue and commitment to problem solving. The positive tone left conference attendees feeling “hopeful,” “confident,” “optimistic.” There was agreement that “everyone owns the problems” that exist in the UME-GME transition, and that we must all work together to address them.

The conference represented the first genuine effort toward assembling all important points of view together. Many prior discussions of USMLE numeric score reporting, or larger issues in the UME-GME transition, occurred within less diverse stakeholder settings. The setting and structure of InCUS allowed for interaction among groups that infrequently come together: for example, medical trainees with backgrounds in the US and international medical education systems interacted with state medical board members; members of the public interacted with both academic and community-based GME directors; and US medical school deans spoke with international medical school graduates and state medical board members.

In many ways, despite the broad scope and the complexity of the issues under discussion at InCUS, the conversations ultimately returned to **two fundamental questions**:

1. How can the overall process by which medical students apply for, and get selected into, a residency program be improved? At InCUS, this issue was often referred to as the UME-GME transition “system.”
2. Whether considered separately or as part of this broader system, what role can or should USMLE numeric score reporting play?

In grappling with these fundamental questions, a **general consensus among participants emerged** on several fundamental points:

- The current UME-GME transition system is flawed and not meeting the needs of various stakeholders. Over time, various stakeholder groups have tried to optimize the system for their own purposes, but this has left some, including applicants, with an undue burden and at worst negatively impacted diversity.
- Unilateral changes made to USMLE alone will not “fix” the system, absent other changes in other parts of the system.
- Changes, both systemic and those specific to USMLE, must be explored, identified and implemented on a reasonable timeline.

Consensus on these fundamental points, and additional feedback from InCUS participants and external stakeholders, led to the following guiding principles informing subsequent recommendations.

Guiding principles for change

(Note: Guiding principles are not listed in order of prioritization.)

- Improve examinee and physician well-being.
- Improve the reliability of assessments in medical education and licensure overall – making high-quality assessment a shared responsibility for the purposes of advancing the health of the public.
- Institute changes that will improve the overall UME-GME transition “system”. Regarding such changes, and in order to limit harm or disadvantage to any one stakeholder group:
 - Allow time for stakeholders to prepare for, and mitigate, any possible disruption while recognizing that many individuals feel a sense of urgency on this issue.
 - To the degree possible, changes should be based on evidence, and focused on outcomes.
- Maintain the quality and integrity of the U.S. medical licensure system in which both domestic and international graduates have been rigorously assessed.
- Promote holistic review of residency applicants, aligning with goals for diversity in medicine.

Recommendations

The following are preliminary recommendations based on input received prior to and during the InCUS conference. Recommendations are divided into those specific to the USMLE program and those addressing the larger UME-GME transition system. It is anticipated that these

preliminary recommendations may be reassessed and potentially revised after the period of public commentary. The period of public commentary is designed to gather further input from interested individuals or organizations not providing comments in advance of InCUS and those not present at InCUS. Upon conclusion of the period of public commentary and analysis of comments, any revised recommendations specific to USMLE will be drafted and further discussed with USMLE governance and the USMLE's parent boards. (See Timeline below) During this latter period, additional input of various stakeholders may be sought.

USMLE Program

Changes within the USMLE program should take place to meet the following goals.

- 1) *Reduce the adverse impact of the current overemphasis on USMLE performance in residency screening and selection through consideration of changes such as pass/fail scoring.*

The most direct means by which the USMLE program may be able to reduce unintended adverse impact is to consider fundamental changes to its current score reporting practices. Options that garnered significant discussion at InCUS included the following:

- (a) Pass/Fail (of Step 1 alone or the entire USMLE sequence);
- (b) Categorical/tiered scoring of USMLE (e.g., quartiles, quintiles, or some other division);
- (c) A composite score across the assessments within USMLE's Decision Point 1 (DP1), consisting of aggregate performance information from Step 1, Step 2 CK and Step 2 CS.

These options, along with any potential modifications to these options and/or new options resulting from the period of public commentary, will be shared with USMLE governance and the FSMB and NBME governing boards beginning fall 2019. Discussions with USMLE governance and the boards of the two organizations will include information and data addressing the relative strengths and challenges of pursuing particular options as mitigating measures to the current state of the UME-GME transition system. Any final recommendation will likely reach USMLE and FSMB/NBME governance for their consideration by the end of 2019 or early 2020.

- 2) *Accelerate research on the correlation of USMLE performance to measures of residency performance and clinical practice.*

Additional research should commence to identify how performance on various assessments within USMLE is, or is not, related to performance in residency and clinical practice. These studies should be supported by organizations across the house of medicine and be sufficiently large to be generalizable and allow for subgroup analysis. The execution of this research, which by its nature represents a longer-term endeavor, should not delay recommended changes above.

- 3) *Minimize racial demographic differences in USMLE performance.**

The USMLE continues to monitor its content for individual items and cases that have differential performance across certain examinee groups. Despite these efforts, known racial differences continue to exist in USMLE performance (27). The basis for some current group differences are likely outside USMLE's direct sphere of control, but USMLE can strive to be a participant in developing solutions. As for those mitigating strategies that are under more direct control, the USMLE program will commit to such ongoing efforts to minimize group differences. These may include additional collaborative and grant funded research, and working with medical schools and other stakeholders to develop model USMLE preparation programs and curricula. Working with its parent boards and governance, USMLE will prioritize this line of work.

* "Race" is currently categorized in USMLE demographics as "self-identified." The USMLE program is in the process of reevaluating how it asks examinees to identify with demographic groups to align with more contemporary practices, while maintaining the ability to investigate demographic performance differences.

UME-GME Transition System

Participants at InCUS provided important input into possible ways to address the UME-GME transition – but given the limitations of time at InCUS, and the complexity of the topic, suggestions for solutions were not fully formed by the end of InCUS. The InCUS planning committee processed comments in the weeks following the conclusion of the conference. Upon further discussion and reflection since the conclusion of InCUS, the following recommendation is offered.

Changes to the residency application system should take place which meet the following goals.

- 1) *Convene a cross-organizational panel to create solutions for challenges in the UME-GME transition.*

InCUS attendees generally agreed that changes to USMLE score reporting alone – while potentially helpful—would not address other aspects of a UME-GME transition system that need attention. Eliminating USMLE score information, in isolation, could actually create more unintended challenges for applicants and residency programs. It was acknowledged that many organizations and stakeholder groups have responsibility for improving this transition. Yet if many are responsible, a concern exists that no one group will take ownership or feel empowered to carry on the broader conversation necessary to bring about appropriate change.

Therefore, it is recommended that a multi-stakeholder panel be convened to review the UME-GME transition, with the purpose of suggesting systemic improvements for which all relevant stakeholder groups would share accountability. This group would optimally be organized by a coalition within the house of medicine, and could invite subject matter experts and guests into a number of task forces, working under the auspices of the larger group panel. Optimally, an interested coalition could be identified within the house of

medicine to convene this UME-GME transition panel, and have an approved proposal, including scope/timelines, by end of calendar year 2019.

Such a panel might address considerations identified at the InCUS, such as the following:

- *Reducing the number of applications perceived by residency applicants as necessary to obtain a position*

Given no likely end to the growth of applications being submitted, continued efforts must occur to provide students with resources and services to inform their residency application and selection process. Some of these resources already exist (13) and some are in continued development, such as the Residency Explorer, which represents a collaborative effort between AAMC, AMA, the USMLE program (NBME/FSMB), NRMP and NBOME. Other efforts, such as limiting the number of residency applications that an individual can submit per match cycle, should be considered as well. Lastly, residency programs should provide more transparent program summaries and selection criteria (if such criteria exist) to applicants as they embark upon the application process.

- *Improving Residency Program Directors' ability to more holistically evaluate candidates*

In addition to clinical departments providing sufficient time for program directors to review applications, additional efforts are required to allow for more individualized and holistic review of residency applicants. Multiple data sources on applicants from U.S. medical education programs are currently available to residency program directors. These include: USMLE scores; Medical Student Performance Evaluations (MSPE); individual letters of recommendation including, in some specialties, letters of evaluation such as the Standardized Letters of Evaluation (SLOE); personal statements and assessments of applicants' other activities, such as research and community service. The current system for residency application does not provide program directors with sufficient options for combining and sorting on multiple domains. Building additional sorting / analytic tools into the residency application system should be prioritized. Such tools could give program directors the ability to sort applicants on overall/holistic profiles consisting of many measures, not simply single measures (such as USMLE Step 1), with the intent of finding applicants most suitable to their programs - based on that program's objectives and strengths. Examples of solutions shared at InCUS which warrant further consideration included the ability of a program director to "weight" certain characteristics beyond USMLE scores and combine measures to be used in screening and selection.

- *Improving the trust of school-based assessments for residency screening and selection*

InCUS highlighted the current challenges that medical schools (US and international) face with developing reliable assessment systems. These include the need for faculty development, limited preceptor time due to competing demands, and concerns that

conveying any information about a student's areas for growth would be perceived as a "warning sign" to a program director, particularly in such a competitive residency environment. It is natural that such limitations in the school-level system of learner assessment could then lead to conflicts of interest, given a school's advocacy for their own students. At InCUS this was generally discussed as a potential misrepresentation of a student applicant's competency, including behaviors. Program directors at InCUS spoke about these challenges leading to workforce issues that negatively interacted with the overall educational goals of the residency training program. Efforts must be undertaken to create a greater sense of trust in school-level performance information at the point of the GME handoff.

Final Reflection on USMLE numeric scoring

While the validity argument for USMLE performance predicting medical practice performance has limitations, some medical specialties – and some programs within a specialty – place more value than others in how, or to what degree, USMLE relates to measures in residency. This may pertain to program specific goals or contexts. To this end, many InCUS attendees acknowledged that the controversy over USMLE numeric scores might not exist if, for example, USMLE numeric scores were weighted 10%, or even as high as 25% in residency screening and selection. It is the degree to which scores are currently being used (for example, score cut-offs, which amount to 100% weight) at certain points in the screening and selection process, that defines the problem.

Indeed, many medical schools frequently place a 15-25% weight on subject examinations in the context of a clinical clerkship, highlighting the point that there is desire for some standardization. If, for example, certain specialties were to develop consensus opinion on weighting of a numeric score or other performance result from USMLE, and justify this approach, applicants for residency might direct their efforts accordingly, while also focusing on presenting themselves in a holistic manner. Unfortunately, the current challenges in the UME-GME transition environment in large part are not leading to this balanced approach. This is why the USMLE program's parents (FSMB and NBME), the AAMC, AMA, and ECFMG felt urgency to convene InCUS in the first place.

The USMLE program is committed to bringing options for change to parent boards and governance in a near-term timeframe. Participants at InCUS, however, recognized the systems-level problems in the UME-GME transition, and the slower rate of potential change for such systems improvements. Over time, it is conceivable that broad systemic changes to the UME-GME transition may occur, particularly with participation from additional organizations from within the house of medicine. Furthermore, advances over time in learning analytics and school-level assessments will likely occur. As these develop over the longer term, the USMLE program will look forward to ongoing discussions over the role of various score reporting practices to best inform medical regulation and medical education.

Timelines

The full text of this report will be posted to <https://www.usmle.org/inCus/> the week of June 10, 2019. The host site for this report will also include an online mechanism for interested parties to post comment/feedback to the report and the preliminary recommendations. This period for public comment will run for 6 weeks.

Respectfully Submitted,

The InCUS Planning Committee

Michael A. Barone, MD, MPH
Vice President, Licensure Programs
National Board of Medical Examiners (NBME)

Andrew T. Filak, Jr., MD
Interim Senior Vice President for Health Affairs and Dean, College of Medicine
University of Cincinnati
Chair, Board of Directors, Educational Commission for Foreign Medical Graduates (ECFMG)

David Johnson, MA
Senior Vice President, Assessment Services
Federation of State Medical Boards (FSMB)

Susan Skochelak, MD, MPH
Group Vice President, Medical Education
American Medical Association (AMA)

Alison Whelan, MD
Chief Medical Education Officer
Association of American Medical Colleges (AAMC)

References

1. Scoring the Step 2 CS Exam. Available at <https://www.usmle.org/step-2-cs/#scoring> Accessed 4/19/2019.
2. Bowles LT. Use of NBME and USMLE scores. *Acad Med.* 1993 Oct;68(10):778.
3. O'Donnell MJ, Obenshain SS, Erdmann JB. Background essential to the proper use of results of step 1 and step 2 of the USMLE. *Acad Med.* 1993 Oct;68(10):734-9.
4. Hoffman KI. The USMLE, the NBME subject examinations, and assessment of individual academic achievement. *Acad Med.* 1993 Oct;68(10):740-7.
5. Williams RG. Use of NBME and USMLE examinations to evaluate medical education programs. *Acad Med.* 1993 Oct;68(10):748-52.
6. Results of the 2018 NRMP Program Director Survey. Available at <https://www.nrmp.org/wp-content/uploads/2018/07/NRMP-2018-Program-Director-Survey-for-WWW.pdf> Accessed 4/28/2019
7. Prober CG, Kolars JC, First LR, Melnick DE. A Plea to Reassess the Role of United States Medical Licensing Examination Step 1 Scores in Residency Selection. *Acad Med.* 2016 Jan;91(1):12-5.
8. Green M, Jones P, Thomas, JX Jr. Selection criteria for residency: Results of a national program directors survey. *Academic Medicine.* 2009;84(3):362-367.
9. FACTS: Applicants, Matriculants, Enrollment, Graduates, MD-PhD, and Residency Applicants Data. Available at <https://www.aamc.org/data/facts/> Accessed 5/1/2019
10. Trends in Osteopathic Medical School Applicants, Enrollment, and Graduates. Available at https://www.aacom.org/docs/default-source/data-and-trends/2018-trends-com-aeg.pdf?sfvrsn=d2ba4c97_74 Accessed 5/1/2019
11. Results and Data 2019 Main Residency Match. Available at <http://www.nrmp.org/main-residency-match-data/> Accessed 4/30/2019
12. ERAS Preliminary Data (ERAS 2019). Available at <https://www.aamc.org/services/eras/stats/359278/stats.html> Accessed on 4/30/2019
13. Apply Smart: New Data to Consider. Available at <https://students-residents.aamc.org/applying-residency/article/apply-smart-data-consider/> Accessed 4/30/2019
14. Grading Systems Use by US Medical Schools. Available at <https://www.aamc.org/initiatives/cir/406418/11.html> Accessed 4/30/2019
15. Osteopathic Medical College Student Performance Evaluation Methods. Available at https://www.aacom.org/docs/default-source/data-and-trends/2016-17-osteopathic-medical-college-student-performance-evaluation-methods.pdf?sfvrsn=b5232c97_10 Accessed 4/30/2019
16. McDuff SG, McDuff D, Farace JA, Kelly CJ, Savoia MC, Mandel J. Evaluating a grading change at UCSD school of medicine: pass/fail grading is associated with decreased performance on preclinical exams but unchanged performance on USMLE step 1 scores. *BMC Med Educ.* 2014 Jun 30;14:127.
17. Bloodgood RA, Short JG, Jackson JM, Martindale JR. A change to pass/fail grading in the first two years at one medical school results in improved psychological well-being. *Acad Med.* 2009 May;84(5):655-62.
18. Hauer KE, Lucey CR. Core Clerkship Grading: The Illusion of Objectivity. *Acad*

- Med. 2019 Apr;94(4):469-472.
19. The United States Medical Licensing Exam (USMLE) and medical student wellness: an ethnographic qualitative study at Cooper Medical School of Rowan University. Available at <https://rdw.rowan.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=3152&context=etd>. Accessed 4/29/2018
 20. Chen DR, Priest KC, Batten JN, Fragoso LE, Reinfeld BI, Laitman BM. Student Perspectives on the "Step 1 Climate" in Preclinical Medical Education. *Acad Med*. 2019 Mar;94(3):302-304.
 21. Andolsek, K. M. One small step for Step 1. *Academic Medicine*, (2019) 94(3), 309-313.
 22. Moynahan KF. The Current Use of United States Medical Licensing Examination Step 1 Scores. *Acad Med*. 2018;93(7):963-965.
 23. Kumar AD, Shah MK, Maley JH, Evron J, Gyftopoulos A, Miller C. Preparing to take the USMLE Step 1: a survey on medical students' self-reported study habits. *Postgrad Med J*. 2015 May;91(1075):257-61.
 24. Burk-Rafel J, Santen SA, Purkiss J. Study Behaviors and USMLE Step 1 Performance: Implications of a Student Self-Directed Parallel Curriculum. *Acad Med*. 2017 Nov;92
 25. Bowles LT, Melnick DE, Nungester RJ, Golden GS, Swanson DB, Case SM, Dillon GF, Henzel TR, Orr NA, Thadani RA. Review of the score-reporting policy for the United States Medical Licensing Examination. *Acad Med*. 2000 May;75(5):426-31.
 26. Comprehensive Review of USMLE. Available at <https://www.usmle.org/pdfs/cru/CEUP-Summary-Report-June2008.pdf> Accessed 4/30/2019
 27. Rubright JD, Jodoin M, Barone MA. Examining Demographics, Prior Academic Performance, and United States Medical Licensing Examination Scores. *Acad Med*. 2019 Mar;94(3):364-370.