

FY22 AAIM Medical Education Research Workplan

Chair: Chayan Chakraborti, MD

Vice Chair: Margaret Baker, MD

Charge: The medical education research committee will assess the culture underlying medical education by designing, evaluating, and supporting curricular innovations.

List of proposed activities	Map to AAIM Strategic Plan (Goal and Initiative)
<ol style="list-style-type: none"> 1. Develop a plan to conduct a literature review that will focus on USMLE Step 1 scores potential implications for residency program selection, future ABIM exam success, and clinical practice skill. This review should be made available to all key stakeholders. 2. Develop a plan to further study the impact on each stakeholder in the transition to USMLE Step 1 pass/fail <ol style="list-style-type: none"> a. US medical students b. Internal medicine clerkship directors c. IMGs d. Internal medicine program directors e. Fellowship program directors 3. Develop a plan to investigate potential structural issues preventing longitudinal assessment of learners 	<p><i>Goal 2 → AAIM will redesign the transitions across the continuum of internal medicine education</i></p> <ul style="list-style-type: none"> ➤ <i>Initiative 4: Promote innovation in medical education research</i>

Medical Education Research Committee

USMLE Step 1 Workgroup Recommendations:

Members: M. Hayes Baker, MD, Oliver Cerqueira, DO, Anthony Donato, MD, Raj Jagarlamudi, MD, Amalia Landa-Galindez, MD, Suchita Pancholi, MD

Background: Beginning in January, 2022, the United States Medical Licensing Examination (USMLE) Step 1 will transition from a numerically-scored exam to a pass/fail exam. Although not originally intended for use as a screening tool, USMLE Step 1 has become a gauge by which many Internal Medicine residency programs initially rate their applicants. With increasing numbers of applications per program, and without widely accepted, holistic screening tools, Step 1 has become the default screen for many programs. USMLE Step 1 has also been used to predict future skill in clinical practice as well as success on the American Board of Internal Medicine (ABIM) examination. For these reasons, medical students have been under tremendous pressure to perform well on this test. The consequences of the change to pass/fail on key stakeholders have the potential to be far-reaching, both positively and negatively. The purpose of this document is to make recommendations that will allow a smoother, more equitable transition for all involved. While we focus on Internal Medicine groups specifically, we recognize that specialties across Graduate Medical Education will be impacted.

Research on Prior and Future Impact:

While the scoring change from USMLE Step 1 has prompted a variety of discussions among physician training groups, understanding the history and validity behind USMLE Step 1 application for such predictions could prove useful. Also, since the announcement that USMLE Step 1 would become pass/fail, multiple programs and institutions have launched surveys and discussions to better understand the impact of this change.

- 1. We recommend a concise literature review be conducted to focus on: USMLE Step 1 scores potential implications for residency program selection, future ABIM exam success, and clinical practice skill. This review should be made available to all key stakeholders.***
- 2. We recommend the results of any studies, surveys, and other literature surrounding the USMLE Step 1 change to pass/fail be collated and distributed to all key stakeholders.***

Key Stakeholders:

Transition to USMLE Step 1 pass/fail has the potential to have both positive and negative impact on U.S. and international medical students, Internal Medicine Clerkship Directors, Internal Medicine Program Directors, and Fellowship Directors. Understanding how each group is affected will highlight needs for change in medical school curriculum, Internal Medicine program recruitment, and individual resident career planning.

- 3. We recommend further study on the impact on each key stakeholder group individually.***
 - U.S. Medical Students:** While U.S. Medical students may benefit from a lessening of the emphasis on studying for Step 1, it is unknown whether this will increase their engagement in other portions of the curriculum, or whether the same pressures will now be transferred to study for Step 2.

- **Internal Medicine Clerkship Directors:** Less emphasis on Step 1 may place more emphasis on performing optimally on Step 2. This change could potentially pull third year medical students away from clerkship study, curricula, and engagement in clinical environment, thereby creating the need for Clerkship Directors to reevaluate existing curriculum, evaluation tools, and medical student advising.
- **International Medical Graduates:** The change to USMLE Step 1 has the potential to negatively impact International Medical Graduates (IMG) more than any other group. High scores on Step 1 had traditionally been the single most important gateway to a U.S. residency. Without that barometer, program directors may not be able to compare among international graduates due to unfamiliarity with both the schools and their curricula. They may then retreat to selection based on schools only, which would heavily favor U.S. graduates at the expense of international graduates. Because of the resultant inequity created for this group, in addition to studying the impact of the Step 1 change on international graduates, we recommend the following:

4a. AAIM work with ECFMG to create an international medical school guide for U.S. residency programs, to include information about grading scales, clinical experience, and evaluation methods.

4b. ECFMG create an advising system to help international graduates choose specific residency programs to apply to, based on mutual interest and program offerings.

4c. AAIM work with ECFMG to create a holistic application screening tool that is specific to international graduates and their unique needs.

- **Internal Medicine Program Directors:** Because Internal Medicine Program Directors have seen tremendous increases in numbers of applications, the Step 1 score has been used as a continuous measure by which to select applicants who have adequate medical knowledge and good test taking skills, prerequisites to later board passage. Without a continuous score, program directors will likely place greater emphasis on Step 2 scores, potentially creating a similar situation for medical students, just later in their training. Program Directors may also place greater emphasis on particular medical schools or medical students that have rotated with them, thereby creating inequity for both lesser-known schools and applicants without audition rotations.
- **Fellowship Program Directors:** Step 1 has been used similarly for fellowship program directors to compare applicants from different medical schools and Internal Medicine residency programs; without it, Step 2 may become more important. Fellowship program directors may also rely more heavily on familiar or recognizable residency programs and medical schools over those that are unfamiliar.

Screening Tools for Holistic Review:

With many medical schools transitioning to pass/fail assessments and the Step 1 converting to pass/fail, program directors are faced with the challenge of finding alternate ways to screen multiple applicants. Additional holistic screening tools are needed to assist program directors in this challenging task, while still limiting the use of tools that may cause excess hardships and stress on medical students.

5. We recommend working with AAMC to create holistic screening tools for programs to utilize.

Potential tools include:

- **Class Rank Reporting in Terciles or Quartiles:** Rather than reporting a specific numerical rank, tercile or quartile reporting would provide an assessment of long-term performance and commitment to self-study. It would eliminate the possibility of having a single “bad day” while taking a standardized exam. To be most useful, it would require uniformity across medical schools.
- **Simulated Clinical Exam:** Developing a standardized clinical exam that focuses on simulated patient interactions and demonstration of ethics, compassion, and professionalism would aid in finding an applicant to fit with the core values of the residency program. Particularly with the disintegration of USMLE Step 2 Clinical Skills due to COVID, the intent of this exam would be to reflect the student’s general level of clinical performance, interpersonal skills, and professionalism, without requiring significant time for preparation.
- **Standardized Specialty-Specific Exam:** By using a specialty-specific exam, such as the Internal Medicine Shelf exam or Aquifer Internal Medicine Clinical Decision-Making Exam, applicants to Internal Medicine can focus on their specialty of choice, and programs can more accurately gauge their performance in Internal Medicine.
- **Enhanced, standardized Internal Medicine Clerkship Evaluation:** Standardization of Internal Medicine Clerkship grading across clinical sites would help programs more accurately choose applicants.
- **Student Self-Assessment Tool:** Tools used for applicants to self-rate their strongest qualities would help programs get to know the applicant and their values. Categories could include: a) commitment to service, b) clinical skills, c) procedural skills, d) academic performance, e) research engagement, f) interpersonal communication skills. Validated Emotional Intelligence (EI) self-assessment tools would also be useful in this setting. This self-assessment would allow programs to select candidates for interview who best match their individual mission/needs.

We recognize that the tasks above may be difficult to assess, implement and distribute. However, studying the impact of Step 1 on learner performance, and the current and future impact of changing to a pass/fail system will promote meaningful scholarship to help students, medical schools, residency, and fellowship programs. Examining the impact of the change to USMLE Step 1 on each key stakeholder group will empower each group to navigate the change more effectively. Creating, implementing, and advocating for holistic application tools will help create a more manageable application season for Internal Medicine programs and more importantly, a more equitable residency application experience for applicants.

References:

Gauer JL, Jackson JB. The association between United States Medical Licensing Examination scores and clinical performance in medical students. *Adv Med Educ Pract.* 2019 Apr 26;10:209-216. doi: 10.2147/AMEP.S192011. PMID: 31114422; PMCID: PMC6497117.

Hamstra, Stanley J. PhD¹; Cuddy, Monica M. MA²; Jurich, Daniel PhD³; Yamazaki, Kenji PhD⁴; Burkhardt, John MD, PhD⁵; Holmboe, Eric S. MD⁶; Barone, Michael A. MD, MPH⁷; Santen, Sally A. MD, PhD⁸ Exploring the Association Between USMLE Scores and ACGME Milestone Ratings, *Academic Medicine*: June 15, 2021 - Volume - Issue - doi: 10.1097/ACM.0000000000004207

McDonald, Furman S. MD, MPH; Jurich, Daniel PhD; Duhigg, Lauren M. MPH; Paniagua, Miguel MD; Chick, Davoren MD; Wells, Margaret; Williams, Amber MEd; Alguire, Patrick MD Correlations Between the USMLE Step Examinations, American College of Physicians In-Training Examination, and ABIM Internal Medicine Certification Examination, *Academic Medicine*: September 2020 - Volume 95 - Issue 9 - p 1388-1395. doi: 10.1097/ACM.0000000000003382.

McGaghie, William C. PhD; Cohen, Elaine R.; Wayne, Diane B. MD Are United States Medical Licensing Exam Step 1 and 2 Scores Valid Measures for Postgraduate Medical Residency Selection Decisions?, *Academic Medicine*: January 2011 - Volume 86 - Issue 1 - p 48-52 doi: 10.1097/ACM.0b013e3181ffacdb

Rayamajhi S, Dhakal P, Wang L, Rai MP, Shrotriya S. Do USMLE steps, and ITE score predict the American Board of Internal Medicine Certifying Exam results? *BMC Med Educ.* 2020 Mar 18;20(1):79. doi: 10.1186/s12909-020-1974-3. PMID: 32183789; PMCID: PMC7079442.

Medical Education Research Committee

Longitudinal Assessment of Learners Workgroup Recommendations

Members: Margaret H. Baker, Chayan Chakraborti, Christopher Henry, Ben Kinnear, Jill Patton, Carleen Spitzer, Jaini Sutaria

Prior Research Prompting Current Question: Do we trust our graduates to deliver high-quality care?

A study by Jonker et al. asked 26 senior anesthesiologists involved in certifying residents ready to graduate (from 21 European countries) a simple question: “Would you entrust your loved ones to each trainee you certify?”¹ Ten of the 26 participants (38%) said they had certified trainees that they “did not deem competent to anaesthetize their relatives.” Several other respondents were hesitant or unsure if they would trust all graduates or not. In an unscientific replication of the Jonker et al. study with the same [question posted to Twitter](#), over 50% of respondents had also graduated trainees that they would not trust with a loved one’s care.

Data from the National Residency Match Program (NRMP) indirectly suggests the same – Program Directors (both in residency and fellowship) use objective measures to **select** “competent” applicants, but they simultaneously **value** subjective measures as predictors of success (or trust). How can we build a system in which all graduates are trusted to provide high-quality care without supervision? How can we promote and evaluate the subjective qualities with a more standardized, less biased approach? The following recommendations seek to answer these questions.

- 1. Conduct a study similar to Jonker’s, with a focus on Internal Medicine in the United States.**
- 2. Initiate research to better understand the systems and human factors that cause program directors to graduate and certify trainees they do not fully trust.**
- 3.**

Current Metrics for Longitudinal Learner Assessment: Do They Already Suggest Competency-Based Assessments Inadequate?

Many of the questions regarding longitudinal assessment of learners seem to stem from an issue of trust. In promoting one group of learners to the next level, how trustworthy are these handoffs themselves? In this breakdown of trust, information exchange can fail due to errors of commission, errors of omission, or “not assessed”. The multiplicity of employed frameworks has resulted in further fracturing, installing systemic barriers to the issue of trust: one framework does not necessarily map to another. More important than trustworthiness from one level of training to the next is perhaps the ultimate test of the success or character of a program—what are the outcomes of training programs’ graduates beyond training?

When looking at the congregate data presented in the National Resident Matching Program (NRMP) 2020 survey of residency program directors (rPDs) on the process of evaluating applicants, a divide exists between what program directors want and what is available to them.² rPDs cited the most important factors in *selecting applicants* as Step scores, Medical Student Performance Evaluation

¹ Jonker G, Ochtman A, Marty A, Kalkman C, Cate Ot, Hoff R. Would you trust your loved ones to this resident? Certification decisions in postgraduate anesthesiology training. *Under review with the British Journal of Anaesthesia*. 2020.

² National Resident Matching Program, Data Release and Research Committee: Results of the 2020 NRMP Program Director Survey.

(MSPE) and class rank or percentile. Though there is considerable debate regarding the merits of USMLE step 1 scores, with the disappearance of this metric, rPDs will be left without an initial screening tool.

When subsequently asked to rank the important factors used to determine residents' *success*, rPDs ranked professionalism, quality of patient care and ethics as most important. Second to these traits, they noted clinical competence and communication skills. When discussing residents' success, there is a transition from the previously cited competency based measures, some of which are more objective, in selecting applicants now to arguably more subjective qualities. Though these qualities are extremely important and desirable for program directors, it is unclear if these qualities are observable or assessable. The clear gap is that we lack a validity argument in support of selection decisions that would be made using these characteristics.^{3,4,5,6} AAIM has an opportunity to explore evidence to support these decisions, and hence build a needed validity argument.

- 3. We recommend working with the AAMC to develop a holistic screening tool.**
- 4. We recommend additional investigations to determine what characteristics residency program directors value in applicants and how they assess them.**
- 5. We recommend a more standardized approach to promotion and evaluation of subjective measures (such as professionalism and communication skills) to diminish bias in the assessment of each.**

When considering residents' transition to fellowship training, the paucity of objective metrics to assess fellowship applicants becomes apparent. In the NRMP 2016 survey of fellowship program directors (fPDs) encompassing all specialties, fPDs cited letters of recommendation as the most important factors in evaluating applicants, as well as the Program Director letter.⁷ Second to letters, fPDs cited the reputation of the applicants' residency and perceived involvement or interest in research. Again, these subjective letters are unreliable predictors of competency given their self-selected nature (i.e. residents will only seek out references from those who look on their clinical performance favorably). Whereas rPDs may have course and rotation grades to supplement applicants' curriculum vitae (CVs) and interview performances, fPDs oftentimes only have a Program Director Letter stating whether or not a resident is meeting or exceeding performance expectations. Though the Program Director letter should be more objective in its summative assessment, this assessment relies almost entirely upon the residency milestones with an incentive to promote applicants for a more successful match - success on behalf of both the applicant and the program. Some of the pitfalls in looking to the milestones as the singular source of an objective performance measure was described succinctly in a survey of the perception of pediatric fPDs.⁸ This survey study identified concerns that milestones, while appropriate for baseline assessments or identifying certain needs, may ultimately lack validity, relevance, and reliability.

³ Veen M, Skelton J, de la Croix A. Knowledge, skills and beetles: respecting the privacy of private experiences in medical education. *Perspectives on Medical Education*. 2020:1-6.

⁴ Kane MT. An argument-based approach to validity. *Psychological bulletin*. 1992;112(3):527.

⁵ Kane MT. Validating the interpretations and uses of test scores. *Journal of Educational Measurement*. 2013;50(1):1-73.

⁶ Cook DA. When I say... validity. *Medical education*. 2014;48(10):948-949.

⁷ National Resident Matching Program, Data Release and Research Committee: Results of the 2016 NRMP Program Director Survey Specialties Matching Survey.

⁸ Reed S, Mink R, Li ST. Utility of Residency Milestones Reported to Fellowship Directors: A National Survey of Pediatric Fellowship Program Directors. *Acad Pediatr*. 2020 Jul;20(5):696-702. doi: 10.1016/j.acap.2020.01.004. Epub 2020 Jan 21. PMID: 31978601.

- 6. We recommend additional investigations to determine what characteristics fellowship program directors value in applicants and how they assess them.**

Potential Structural Issues Preventing Longitudinal Assessment of Learners: How Can We Build a System in Which we Trust All Graduates?

Competency-based medical education (CBME) is the predominant paradigm for both undergraduate medical education (UME) and graduate medical education (GME).⁹ It is an outcomes-based approach that begins by identifying patient and healthcare system needs, defining competencies required to meet those needs, and then designing curricular and assessment processes to help learners achieve those competencies.¹⁰ CBME arose, in part, from calls in the mid-20th century for physician training programs to be more accountable to society.¹¹ In short, we should keep high-quality patient care at the center of everything we do in education and we should not graduate learners not competent to deliver such care.

As trainees progress throughout their training, there are a number of different methods that programs use for evaluation. Unfortunately, these tools are far from perfect. Entrustable professional activities (EPAs) and Milestones are inherently subjective and there may be significant variation in faculty members' assessments of these metrics, especially when exposure periods are brief.¹² Additionally, though significant time has been spent building evaluation "bridges" across training programs¹, there is little evidence that these proposed structures lead to desired patient outcomes or even that an individual trainee's performance is passed from program to program via these bridges. In fact, in one recent study designed to assess the Milestones' ability to provide a linear path of development from pediatrics residency to trainees' first-year neonatal-perinatal medicine fellowship, researchers found that the average Milestone scores in trainees' last year of residency were significantly higher than scores assigned at their first scoring period in fellowship.¹³ These changes persisted even for trainees who completed residency and fellowship at the same program. These data suggest that fellowship evaluations represent a "reset button" for evaluation and not the desired linear, longitudinal assessment.

- 7. We recommend investigating the number of faculty who participate in faculty development related to trainee assessment utilizing a specific framework (i.e. Milestones, EPAs, OPAs) and determining what such programs look like.**
- 8. We recommend the creation of a faculty development program related to trainee assessment to be distributed to all key stakeholders.**
- 9. We recommend investigating how different training programs "hand off" assessments from residency to fellowship to unify a more longitudinal evaluative system.**

In addition to shortcomings in how we currently evaluate performance, we also need to consider what is not being evaluated. If noncognitive skills are, at least in part, what rPDs and fPDs seek to critically appraise residents' competency and likelihood of success, then we need a better framework within

⁹ Carraccio C, Wolfsthal SD, Englander R, Ferentz K, Martin C. Shifting paradigms: from Flexner to competencies. *Academic medicine: Journal of the Association of American Medical Colleges*. 2002;77(5):361-367.

¹⁰ Frank JR, Snell LS, Cate OT, et al. Competency-based medical education: theory to practice. *Medical teacher*. 2010;32(8):638-645.

¹¹ Frank JR, Snell L, Englander R, Holmboe ES, Collaborators I. Implementing competency-based medical education: Moving forward. *Medical teacher*. 2017;39(6):568-573.

¹² Favreau MA, Tewksbury L, Lupi C, Cutrer WB, Jokela JA, Yarris LM; AAMC Core Entrustable Professional Activities for Entering Residency Faculty Development Concept Group. Constructing a Shared Mental Model for Faculty Development for the Core Entrustable Professional Activities for Entering Residency. *Acad Med*. 2017 Jun;92(6):759-764. doi: 10.1097/ACM.0000000000001511. PMID: 28557935.

¹³ Sawyer T, Gray M, Chabra S, et al. Milestone level changes from residency to fellowship: a multicenter cohort study. *J Grad Med Ed*. 2021:377-384.

which to collect this data and translate it into a meaningful observation. In 2011, the Alliance for Academic Internal Medicine's Education Redesign Committee attempted something similar by creating a list of EPAs.¹⁴ Briefly, EPAs are distinct from discrete milestones in that they underscore the notion of competence as "synthetic and contextual". An EPA is a constellation of multiple competencies or observable behaviors that, in combination, represent the capability of a resident working towards becoming an independent clinician. An extension of this concept was recently proposed as an observable practice activity (OPA).¹⁵ OPAs are similar to nested EPAs, and are specific tasks performed by physicians in their daily work.¹⁶ They are intended to allow for better construct alignment for frontline assessors, particularly clinicians who are assessing. Multiple groups have described specific behaviors and skills that are thought to reflect the non-cognitive attributes that PDs desire.^{17,18} These are, however, rarely explicitly reported to residency and fellowship programs around times of transition.

10. Based on what is desired by rPDs and fPDs, we recommend reporting noncognitive skills (via Milestones, EPAs, OPAs or a novel scoring rubric) in a more standardized fashion to potential residency and fellowship programs within the MSPE or Program Director letter

In summary, though we are making strides to better objectively quantify trainees' specific competencies, these assessments primarily exist in silos. Milestone assessments are not linear as trainees move from UME to residency to fellowship and there are no data to suggest that these evaluations are correlated with patient outcomes. In order to address these breaks in the current system, we need to determine what metrics are most important when determining whether an individual is ready to graduate to the next phase in training and how to best communicate areas for continued improvement. Overall, we need to broaden or redefine the metrics we use to ensure the assessment of competency predicts not only cognitive performance but also implicit trust, or "success", in future patient care.

¹⁴ Caverzagie KJ, Cooney TG, Hemmer PA, Berkowitz L. The development of entrustable professional activities for internal medicine residency training: a report from the Education Redesign Committee of the Alliance for Academic Internal Medicine. *Acad Med.* 2015 Apr;90(4):479-84. doi: 10.1097/ACM.0000000000000564. PMID: 25406600.

¹⁵ Warm EJ, Mathis BR, Held JD, et al. Entrustment and mapping of observable practice activities for resident assessment. *Journal of general internal medicine.* 2014;29(8):1177-1182.

¹⁶ ten Cate O, Chen HC, Hoff RG, Peters H, Bok H, van der Schaaf M. Curriculum development for the workplace using entrustable professional activities (EPAs): AMEE guide no. 99. *Medical teacher.* 2015;37(11):983-1002.

¹⁷ Warm EJ, Kinnear B, Lance S, Schauer DP, Brenner J. What Behaviors Define a Good Physician? Assessing and Communicating About Noncognitive Skills. *Acad Med.* 2021 Jun 22. doi: 10.1097/ACM.0000000000004215. Epub ahead of print. PMID: 34166233.

¹⁸ Pingree EW, Huth K, Harper BD, et al. Encouraging Entrustment: A Qualitative Study of Resident Behaviors That Promote Entrustment. *Academic Medicine.* 2020.