Medical Education's Wicked Problem: Achieving Equity in Assessment for Medical Learners

Catherine R. Lucey, MD, Karen E. Hauer, MD, PhD, Dowin Boatright, MD, and Alicia Fernandez, MD

Abstract

Despite a lack of intent to discriminate, physicians educated in U.S. medical schools and residency programs often take actions that systematically disadvantage minority patients. The approach to assessment of learner performance in medical education can similarly disadvantage minority learners. The adoption of holistic admissions strategies to increase the diversity of medical training programs has not been accompanied by increases in diversity in honor societies, selective residency programs, medical specialties, and medical school faculty. These observations prompt justified concerns about structural and interpersonal bias in assessment. This manuscript characterizes equity in assessment as a "wicked problem" with inherent conflicts, uncertainty, dynamic tensions, and susceptibility to contextual influences. The authors review the underlying individual and structural causes of inequity in assessment. Using an organizational model, they propose strategies to achieve equity in assessment and drive institutional and systemic improvement based on clearly articulated principles. This model addresses the culture, systems, and assessment tools necessary to achieve equitable results that reflect stated principles. Three components of equity in assessment that can be measured and evaluated to confirm success include intrinsic equity (selection and design of assessment tools), contextual equity (the learning environment in which assessment occurs), and instrumental equity (uses of assessment data for learner advancement and selection and program evaluation). A research agenda to address these challenges and controversies and demonstrate reduction in bias and discrimination in medical education is presented.

Increasing diversity in medicine and other health professions improves access and the quality of care provided to minority populations, enhances the educational environment for both minority and majority students, optimizes team-based problem solving, and expands the focus and relevance of medical research.[1–4] Driven in part by accreditation standards and national organizations,[5–8] many medical schools and residency programs have successfully used holistic admissions strategies to increase the diversity of their classes. However, similar increases in diversity have not been realized in honor societies, selective residency programs, and medical specialties, and among faculty in U.S. medical schools.[9] Because entry into competitive careers and programs is often dependent on student grades and academic awards, these observations have prompted concerns about whether the assessment practices of medical schools may be subject to structural and interpersonal bias.

Medical schools have a moral responsibility to respond to these concerns about equity in assessment. Assessment guides learning, enables student progress, and impacts career choice and opportunity while also assuring stakeholders of graduates' competence. Inequitable assessments in medical education perpetuate barriers to advancement and career opportunities for learners from groups underrepresented in medicine (UIM). Even if unfounded, concerns about the possibility of inequitable assessments burden UIM learners and add to the challenging nature of the learning environments in which they are expected to perform.

Equity in assessment meets the definition of a wicked problem.[10] Wicked problems are characterized by conflicts, uncertainty, dynamic tensions, and susceptibility to contextual influences. Many faculty believe that our current approach to teaching and assessing students is defensible, rewarding merit and hard work. Increasingly, others believe that it has been tainted by the same structural racism that has perpetuated a state of educational opportunity and career privilege for populations that have historically constituted the majority of medical students, residents, and faculty. The literature that explores issues of equity in assessment is providing greater clarity into the complexity of the problem but has yet to resolve the fundamental questions about the nature of equity in assessment or propose solutions to observed differences.

Addressing this wicked problem will require concerted work by educators in all medical schools and residency programs. To help catalyze this work, we begin this paper by defining key concepts to establish a shared mental model of equity in assessment.[11] We continue with a review of the literature, exploring concerns about equity in assessment in medical education, and follow with a proposed framework modeled on work in the field of organizational excellence. Finally, we describe challenges and controversies that future research should explore.

What Is Equity in Assessment?

Equity is the state of being fair and achieving social justice.[12,13] An equitable assessment system thereby

Please see the end of this article for information about the authors.

Correspondence should be addressed to Catherine R. Lucey, UCSF School of Medicine, 533 Parnassus Ave., U80, Box 0710, San Francisco, CA 94143; telephone: (415) 502-5153; email: Catherine.Lucey@ucsf.edu.

Acad Med. 2020;95:S98–S108.
First published online September 1, 2020
doi: 10.1097/ACM.0000000000003717

Supplemental digital content for this article is available at http://links.lww.com/ACADMED/B18.

facilitates future educational and career opportunities. Equity in assessment is present when all students have fair and impartial opportunities to learn, be evaluated, coached, graded, advanced, graduated, and selected for subsequent opportunities based on their demonstration of achievements that predict future success in the field of medicine, *and* that neither learning experiences nor assessments are negatively influenced by structural or interpersonal bias related to personal or social characteristics of learners or assessors. An equitable assessment system should enable both majority and minority learners to learn more and learn better.[14]

Components of equity in assessment

There are 3 components to equity in assessment: intrinsic equity, contextual equity, and instrumental equity (see Figure 1). Intrinsic equity means that the design of the program of assessment and the assessment tools used minimize bias against groups historically marginalized by the medical profession. Established criteria define high-quality assessment strategies and also inform efforts to achieve equity in assessment. These criteria include: (1) validity or coherence, (2) reproducibility or consistency, (3) equivalence, (4) feasibility, (5) educational effect, (6) catalytic effect, and (7) acceptability.[15] Adhering to these criteria guides educators to align curriculum and assessment and to ensure that what is measured reflects what is valued, guides learning, and ensures accountability of the system to patients and the public. When these criteria are not met, *inequity* in assessment may result. Table 1 summarizes definitions for these criteria and provides an example of how failing to meet them can enable inequity in assessment.

Contextual equity refers to fairness in the learning experience and the environment in which assessment strategies are implemented. Contextual equity includes fairness in: assigned environments and tasks within those environments, support and social structures available within and outside the learning environments, and the preparation of supervisors who implement assessment procedures. These factors impact the climate experienced by learners and, thus, the opportunity learners have to perform and be assessed at their peak abilities.

Equity in assessment in medical education also requires attention to instrumental equity: How results of assessment processes are shared with and used by stakeholders to create equitable opportunities for all, regardless of their social class or personal characteristics. Instrumental equity is present when the results are shared and used in a manner that neither over- nor underpredicts subsequent performance in the context for which assessment was designed (see Table 2).

Intrinsic, contextual, and instrumental equity are process equity values that collectively contribute to equity in assessment outcomes: the opportunities afforded to individuals (such as selection for a prestigious residency) or populations (such as diversity of faculty in academic health systems) based on the consequences of assessment.

Why a focus on psychometric rigor shortchanges equity in assessment in the clinical learning environment

The definition of contextual equity—fairness in the learning environment in which assessment occurs—is inherently challenging in the complex clinical environment. Learners experience different patients, clinical conditions, team dynamics, time constraints, supervisor skill, and other factors. This variability makes the conditions for assessment similarly variable, and in conflict with traditional expectations for tightly controlled, reproducible conditions sought for high-stakes testing. For example, it is entirely feasible to pursue psychometric rigor—reproducible, reliable results that are often touted as "objective"—in the environments of high-stakes national licensing and certification examinations. However, these examinations only measure some aspects of the competencies relevant to physician performance and thus, used alone, have only modest predictive ability for future patient care.[16–20] For most competencies other than knowledge, patient care skills are an essential component of the valid and equitable assessment of learner performance. Creating the conditions for equitable assessment of clinical skills has thus far been challenging for medical educators. In the psychometric era of assessment, the major focus on reliability or reproducibility of workplace assessments fueled an assumption that fairness would result if raters were sufficiently trained to provide similar ratings of different levels of learner performance. However, assessment of clinical performance has fallen grossly short of this vision of highly reliable ratings. In fact, more recent literature advocates that we embrace the variability in ratings as a meaningful reflection of the complexity of clinical tasks, contexts, and different supervisors whom learners encounter.[21–24] This guidance embraces the importance of contextual equity;

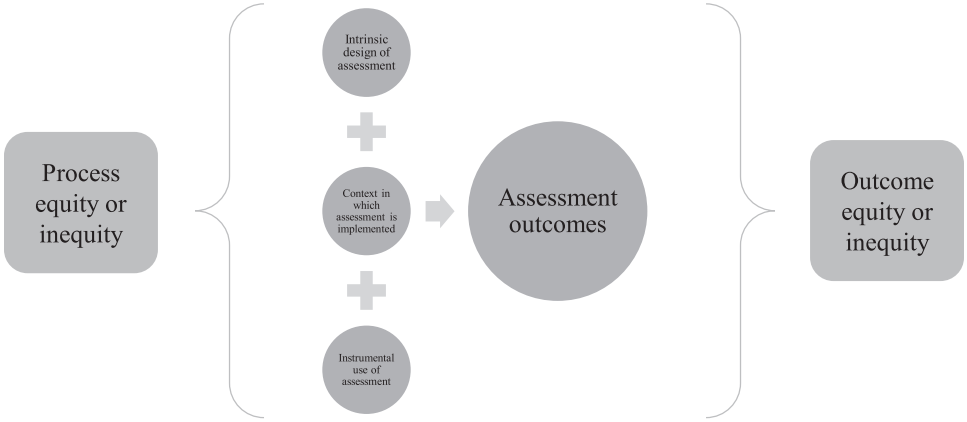

Figure 1 Components of Equity in Assessment

Table 1

Criteria for High-Quality Assessment and Examples of Inequity Resulting From Failure to Meet These Criteria[37]

Criteria for high-quality assessment	Definition	Example of inequity resulting from failure to meet this criterion
Validity or coherence	Assessment tools and strategies measure the intended construct and are used for the purpose for which they are intended.	The numeric results of a licensing examination designed to ensure minimal competence in medical knowledge are used for selection for graduate medical training, despite known structural differences in education that favor majority groups on high-stakes examinations.
Reproducibility or consistency	The same assessment yields the same results under a variety of conditions.	Supervisors are observed to rate the clinical skills of non-UIM and men higher than those of UIM and women learners.
Equivalence	Information is used similarly across settings or institutions to make assessment decisions.	Clerkship directors, grading committees, program directors, or honor society committees synthesize assessment information in ways that consistently favor non-UIM learners.
Feasibility	The assessment tool or strategy is practical to implement.	Clinical faculty do not observe UIM learners' unique contributions to the care of UIM patients through language concordance, advocacy, and doctor–patient relationships due to time demands on a busy clinical service.
Educational effect	Assessment methods motivate learners and drive them to focus on certain activities.	UIM learners are aware of the literature and experience documenting lower clinical scores and grades for members of their group and therefore underperform in the clinical environment because of stereotype threat.
Catalytic effect	The effects of assessment results on learners and the system.	A progress committee reviews differences in performance data for UIM and not-UIM learners and concludes that UIM learners may not be qualified for the program.
Acceptability	Learners and educators find the assessment tools and procedures to be workable and credible.	UIM learners have a blog that discusses the flaws in assessment of UIM learners in the program.

Abbreviation: UIM, underrepresented in medicine.

failure to consider contextual equity perpetuates inequitable assessment practices, as described below.

Existing Literature: Concerns About Equity in Assessment in Medical Education Are Warranted

Exploration into the question of equity in assessment has begun with multiple studies focused on concepts such as bias, fairness, differences, and differential attainment in assessments used for entry to, progress through, and graduation from medical school and successful competition for residency programs and faculty careers.[25–35] These studies consistently document population group differences in standardized examinations, clinical assessments, grading, and academic awards between UIM and well-represented in medicine (WRIM) students and residents, differences that virtually always favor WRIM learners. Furthermore, studies have also documented that minor differences in assessment outcomes can have a powerful impact on residency and career opportunities.[9,30]

Population group differences exist in high-stakes standardized exams

National high-stakes standardized exams are often used to select students for entry into medical school and into the most competitive specialties, residency programs, and careers. Unfortunately, in many of these exams, population group differences exist that consistently favor White applicants over their non-White peers and men applicants over women. The Medical College Admission Test (MCAT) is one of the most important selection criteria used by medical school admissions officers to determine which applicants are offered medical school interviews and acceptance.[36] Although individual women and UIM students score across the range of MCAT scores, population group differences exist in MCAT scores, with Black and Hispanic students, on average, scoring lower than their White peers.[35,37,38] Similarly, the United States Medical Licensing Exam (USMLE) Step 1 exam has played an important role in residency selection.[39] Similar population group differences have been demonstrated on the USMLE Step 1. Studies have consistently shown that Black, Hispanic, and Asian medical students, on average, have lower USMLE Step 1 scores than White students and women, on average, score lower than men.[40–42] These group differences in USMLE Step 1 scores are not explained by students' prior academic achievement, and they persist even after accounting for students' total grade point average and MCAT scores.[43]

Psychometric analysis of the predictive ability of the MCAT on future performance of students in medical school and the USMLE Step 1 exam on performance on future licensing exams shows no evidence of intrinsic bias of the exams.[35] The prevailing theory explaining population group differences in the MCAT, the USMLE, and other high-stakes standardized exams is that the differences result from consequences of structural racism on educational opportunities afforded to UIM students.[35,38] The recent change to pass/fail scoring for the USMLE Step 1 examination represents a promising strategy to address inequity while maintaining assurance of minimal competence in medical knowledge for licensure.

For both the MCAT and USMLE examinations, students with a broad range of scores are capable of future success in medical school, residency, and physician practice. However, many medical schools (for the MCAT) and residency programs (for the USMLE) choose to restrict the candidates they consider for interviews and admission to those with the highest scores.[38,44] They cite concerns about need to efficiently screen large numbers of applicants, false beliefs about the predictive ability of minor differences in scores, and pressure from

Table 2
Examples of Challenges to Instrumental Equity

Assessment	Intended use of assessment information	Instrumental use of assessment information	Inequity resulting from assessment	Consequence of inequity
USMLE Step 1 licensing examination scores	Determination of minimal competence for licensure to practice medicine, to ensure safety of the public	Sorting applicants for residency training into groups who will or will not be invited for interviews	Students from backgrounds underrepresented in medicine are at risk for lower scores due to structural factors throughout their education	UIM students do not match into top residency programs
Clerkship grades	Quantitative ratings and qualitative comments from team members ensure that students have achieved expected competence and inform future learning	Ranking students and sorting students to determine who will receive the highest grades	Bias in quantitative and qualitative ratings favors students who are White; bias exacerbated when faculty providing ratings may comprise a less diverse group than the student population	UIM students are less likely to be elected to the AOA honor medical society
Examinations of medical knowledge during clerkships	Assurance that all medical students achieve the expected minimum medical knowledge across varied clerkship sites and settings	Numerical data serve as easy, "objective" metrics that can be weighted heavily alongside clinical performance data to rank or sort students into groups for purposes of grade assignments	Medical knowledge contributes more to or drives clinical grade assignments rather than other competencies essential for high-quality patient care	UIM students earn fewer honors in core clerkships
Milestone ratings of resident performance	Monitor and support all residents' developmental trajectory	Quantitative milestone ratings enable ranking of residents	Women residents receive lower milestone ratings than men in certain domains that are traditionally valued as "male" characteristics	Women residents are less likely to be selected for chief resident positions or faculty appointments

Abbreviations: UIM, underrepresented in medicine; AOA, Alpha Omega Alpha Honor Society.

leadership to craft a class with the highest exam metrics to increase institutional rankings according to *U.S. News & World Report*. Given the demonstrated population group differences on these exams, this restriction of opportunity to the highest scorers systematically disadvantages UIM students and presents a barrier to diversifying medical school classes and residency programs.[37,45,46]

Population group differences exist in narrative evaluations, grades, and awards

Narrative evaluations and medical school grades are thought to provide a more holistic view of the competencies needed for future success as a medical resident, fellow, or practicing physician. Population group differences also exist in departmental and institutional assessments of students by faculty and administrators. Descriptions of medical students used by faculty in narrative evaluations and letters of recommendation vary by sex, gender, race, and ethnicity. These differences favor White students over non-White students.

Medical student performance evaluations (MSPEs) are the documents prepared by medical schools to summarize student performance for the purpose of residency selection. In a recent study of MSPEs, White students were more likely to be described with "standout" words such as "best," "excellent," and "outstanding" compared with their non-White peers.[32] Further, female students were more likely than males to be described as "caring" and "compassionate" and less likely to be described with words that denote intelligence and competence. Another recent study of language in core clerkship evaluations found that evaluators reinforce gender stereotypes through their choice of words. For example, women were more likely than men to be described as "lovely," while men were more likely to be described as "scientific."[3]

Studies have also demonstrated racial and ethnic differences in awarded grades. A single academic medical center recently found that students historically UIM (Black or African American, Hispanic, American Indian/Alaska Native, or Native Hawaiian/Pacific Islander) received half as many "honors" grades across all clerkships compared with White students.[9] An examination of numeric global assessment rankings in the Standard Letters of Evaluation used to rank 2,884 medical students applying to emergency medicine (EM) programs found that Black applicants were rated lower than White applicants across several domains, including ratings on future success in EM, rank list prediction, and overall applicant ranking.[47]

Population group differences also arise in selection of students for membership in the Alpha Omega Alpha Honor Society (AOA). AOA membership is predictive of future success in academic medicine, and membership often advantages students in obtaining the most competitive residency positions.[39,48,49] Recent studies reveal racial and ethnic disparities in society membership, with Asian medical students being approximately 50% less likely than White students to be selected into AOA. Black medical students are nearly 80% less likely than White students to be elected into AOA, even when accounting for numerous measures of academic achievement including USMLE Step 1 and 2 scores, dual degrees, research productivity, and

hours dedicated to leadership activities and community service.[30] This disparity suggests that racial and ethnic minority medical students are differentially rewarded for comparable achievement. Some institutions have chosen to suspend membership in AOA because of these concerning observations.[50]

These differences in narrative assessments, clerkship grades, letters of recommendation, and honor society membership impart sustained, negative consequences for both individuals and the profession. In an analysis of U.S. medical school graduates from 2005 to 2015, Black or African American, Hispanic, Asian, and American Indian/Alaskan Native graduates were all less likely than White graduates to secure training positions in graduate medical education (GME), even after accounting for USMLE Step 1 scores.[51] Moreover, the number of Black and Hispanic graduates unplaced in GME after medical school graduation increased throughout the study. Because GME training is requisite for medical licensure, delayed entry or inability to enter a residency program after medical school has substantial economic consequences for students who may also have considerable student debt. The consequences on goals to diversify the profession are clear.

Potential causes of population group differences in assessment in medical education

Population group differences exist in medical education assessments between majority and racial, ethnic, and gender minorities; they constrain individuals' opportunities and contribute to challenges in diversifying the profession. It is tempting to seek a simple explanation to this problem. Some have argued that the mere existence of population group differences in assessment confirms overt bias and discrimination in the assessment process. Certainly, some assessments may be poorly designed, and some assessors may be consistently biased. Others have posited that population group differences reflect differential aptitude within affected populations. Indeed, within any medical school, different students from many different social groups and identities demonstrate different aptitudes for and interest in various specialties. If the origins of this problem resided with individual learner or assessor performance, we would not expect to see the same results in studies done at different institutions with different assessment methods. The universality and consistency of differences advantaging those from groups WRIM suggest that systematic forces are likely operational.

As with other wicked problems, solving the problem of inequity in assessment requires a broader view of possible causes and potential solutions. The social–ecological model used in public health recognizes that individual outcomes must be considered in the context of broader organizational and social systems.[52] Applying this model to the issue of equity in assessment provides insights into possible causes of observed population group differences and illustrates the need to plan interventions at multiple levels (see Figure 2).

Our working definition of equity in assessment can be used to test institutional systems for their potential contributions to inequity in assessment. Equity in assessment is present when students have equitable opportunities to: learn (contextual equity); be evaluated, coached (assessment *for* learning); be graded, advanced, graduated (assessment *of* learning); and be selected for subsequent opportunities based on demonstrated achievements that predict future success in medicine (instrumental, assessment for ranking) *and* that neither learning experiences nor assessments are influenced by structural or interpersonal bias related to personal or social characteristics of the learner, assessor, or context of the assessment. Inequity in

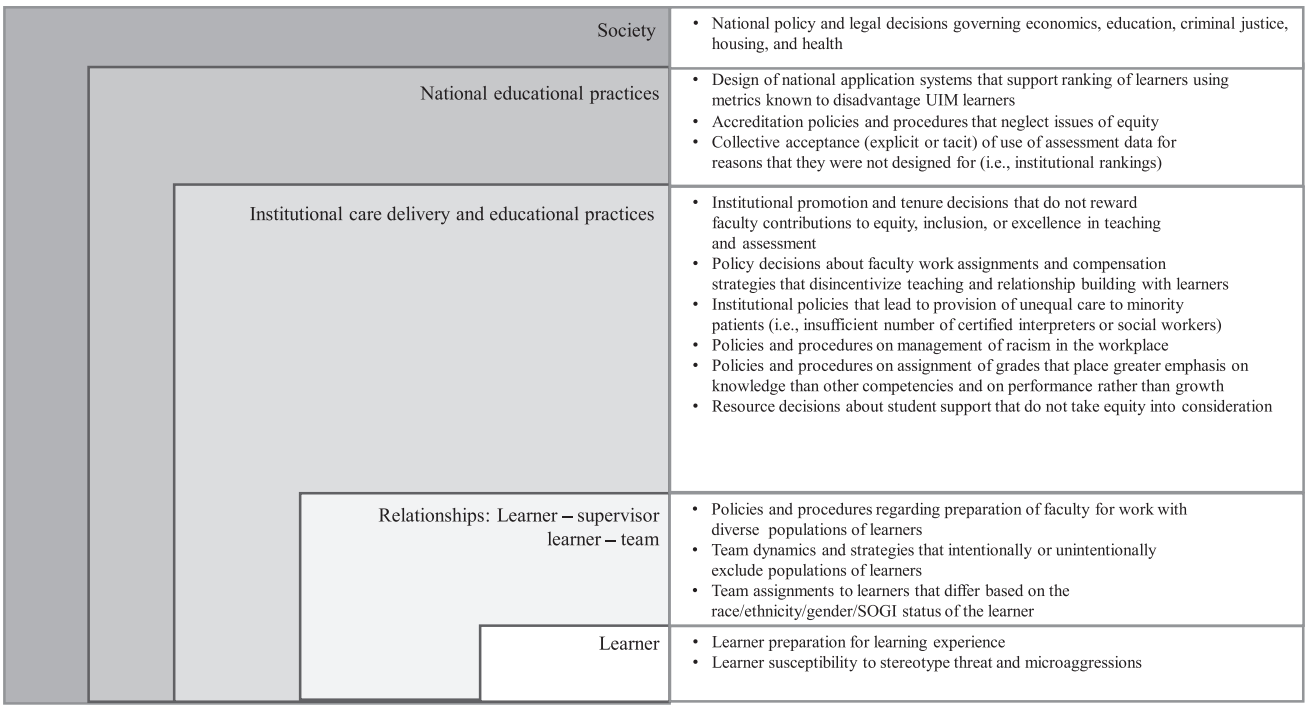

Figure 2 The socioecological model applied to medical education: Examples of decisions at multiple social system levels that impact equity in assessment. Abbreviations: UIM, underrepresented in medicine; SOGI, sexual orientation and gender identity.

assessment exists if these conditions are not met. Supplemental Digital Appendix 1 at http://links.lww.com/ACADMED/B18 summarizes examples of inequity in each of these domains relevant to assessment.

It is vital to consider contextual equity, even though it may not be under the direct control of educators who design programmatic assessment strategies. If work is done to optimize equity in intrinsic and instrumental assessment procedures and policies without addressing challenges in the learning environment that systematically disadvantage one population and not others, then inequity in assessment outcomes will remain. Case studies outlined in Supplemental Digital Appendix 2 at http://links.lww.com/ACADMED/B18 illustrate how the elements of inequity impact learners in the clinical learning environment.

Population group differences may occur for reasons other than inequity or bias. Population group differences may exist because one population is consistently more likely to demonstrate the characteristics valued and measured by the assessment process. Wijesekera and colleagues identified that women were more than twice as likely as men to be inducted into the Gold Humanism Honor Society (GHHS), with criteria emphasizing empathy and patient centeredness.[53] Our working definition of equity in assessment is a lens to explore the question of whether one population group is consistently more likely to demonstrate the characteristics valued or whether the selection process is inequitable. Do women have more empathy and patient centeredness than men (individual aptitude and achievement)? Are the ways that men demonstrate empathy and patient centeredness equally captured by assessors and/or by the assessment method (intrinsic equity)? Do men and women have equal opportunity to receive coaching (context and assessment for learning) in the expected ways to demonstrate these traits? Does the assessment context allow men and women equal opportunities to demonstrate these traits (context and conduct of learning)? And finally, an organizational value statement: Are these traits important enough to future physicians to be measured and used to differentiate among individuals (intrinsic and instrumental equity)? This example both highlights the importance of antecedent institutions, cultures, and experiences that shape men and women students differently and may lead to true differences in demonstration of these traits while also raising valid questions about assessment strategies. How the GHHS responds to these findings will reflect, embody, and ultimately promote institutional values.

A Framework for Equitable Assessment in Medical Education

Drawing on insights from the literature on equity and fairness in medical education assessments, hypotheses for why disparities in assessment exist in medical education assessment, and theories supporting high-quality assessment, we propose a framework for considering equity in assessment. This framework is based on the Shingo model for organizational excellence, which recognizes that effective organizations begin their improvement work from a purpose-driven and principled platform and then move to shape the culture, build the systems, and select the tools that can achieve the results that reflect those principles.[54] List 1 outlines principles that guide application of the Shingo model to equity in assessment.

Guiding principles

Because results of assessments at any given institution will be used by multiple institutions and organizations (residency and fellowship programs, licensing and certifying boards, hospital credentialing units, professional societies, and governmental authorities), a model for equity in assessment in medical education must transcend program and institutional borders. Thus, successfully achieving equity in assessment in medical education requires a nationwide, collective commitment to advance equity as an essential element in our work in health care and medical education.

The aim of all assessment in medical education is to ensure that every graduate of schools and training programs has demonstrated the competencies needed to provide high-quality, patient-centered, equitable care for all patients. This purpose must drive the design, implementation, and continuous improvement of the culture of medicine and medical education; the systems and strategies we use to assess, grade, promote, graduate, and certify learners; and the tools we select to implement our strategies.

Culture

Culture represents an organization's norms, expectations, beliefs, and values. Cultural norms are both explicit and tacit and are reflected in the behaviors exhibited and rewarded by individuals in the organization. Climate is the individual experience of culture.

Achieving a culture that supports equity and inclusion requires recalibration of long-standing beliefs about how we define, develop, and recognize excellence in medicine. In the 21st century, the team-based delivery model needed to provide care for patients with complex chronic diseases requires that all team members be excellent. Therefore, what matters to the health of our patients and communities is not how good the very best physician is, but how good every physician is. This view aligns well with equity in assessment. Our culture of assessment must embrace the belief that all learners can and must grow and develop throughout their medical careers, aided by systems and other professionals committed to supporting this growth. Our focus must be to design assessments to ensure every graduate exceeds the competency thresholds necessary for safe, high-quality patient care, rather than to identify the top 10% of a graduating class.

Systems of assessment

Systems must be designed to support intrinsic equity in assessment. Educators must engineer systems that explicitly articulate criteria for assessment, equalize the learning opportunities, provide formative feedback for all learners, minimize the deleterious impact of unconscious bias of any individual evaluator on a learner's grades, and recognize and reward growth rather than performance mindset. While little empirical data guide the design of equitable programs of assessment, 2 current frameworks for assessment align with and can support development of equitable programs of assessment: competency-based medical education (CBME) and programmatic assessment focused on equity. These approaches have the potential to minimize unconscious bias in individuals and groups that

List 1

Achieving Equity in Assessment: A Model Based on the Shingo Model of Organizational Excellence

Guiding principles:

- The purpose of medical education is to prepare a physician workforce capable of and committed to providing high-quality, safe, and equitable care to our increasingly diverse patients and communities.
- The purpose of assessment in medical education is to ensure that medical education fulfills our social contract by ensuring that all who graduate from a school or training program have the competencies needed to provide excellent and equitable care to all patients.

A culture committed to equity in assessment values:

- Diversity, equity, and inclusion as drivers of a high-quality health care system.
- Equity as an essential characteristic of high-functioning learning and assessment systems.
- Excellence in all individuals and teams as defined by achievements in the comprehensive set of competencies that are required to provide high-quality patient care.
- A commitment to growth and improvement as an essential requirement for sustained excellence over the course of a career.

Equitable systems and programs of learning and assessment:

- Are centrally designed and continuously monitored for evidence of equity, using contemporary theories of assessment and learning.
- Focus on structures and processes that optimize intrinsic equity by mitigating the impact of unconscious bias by individuals and groups in assessment for learning, including:
    - explicit criteria by which achievements are assessed, rather than relying on normative criteria
    - a diversity of assessment strategies and metrics to validly assess the breadth of competencies needed for physicians to provide excellent care
    - preparation of all supervisors who assess learners and the learners themselves to have a clear understanding of assessment procedures and desired competency attainment criteria
    - strategies that ensure that all students receive frequent, actionable, formative assessment before summative, high-stakes decisions are made
- Focus on structures and processes that optimize intrinsic equity by mitigating the impact of unconscious bias by individuals and groups in assessment of learning, including:
    - Summative decisions about competency achievement are based on evidence collected from multiple observers who interact with the learner in a wide range of clinical contexts, with a diversity of patients.
    - Summative decisions about competency achievement are made by committees of diverse individuals, expert in assessment procedures and data analysis, and educated about the ramifications of unconscious bias and dysfunctional group think.
- Focus on structures and processes that optimize contextual equity, including:
    - Curricular environments that afford all learners with the opportunity to learn while participating in clinical situations of varying complexity to optimize their chance of maximizing their achievements.
    - Learning environments designed and monitored to address bias, stereotype threat, unmeasured workload, and other causes of inequitable learning that contribute to inequitable assessment.
    - Faculty work assignments and promotion criteria that enable and incentivize them to dedicate the time and effort needed to optimize assessment for learning.
- Advocate for structures and processes that support instrumental equity and equity in assessment outcomes.

Tools that facilitate equity in assessment include:

- Criterion-based competency descriptions.
- Assessment strategies that incorporate qualitative and quantitative data selected for their ability to predict future performance as a physician, rather than on ease of collection or interpretation.
- Faculty development and just-in-time tools that allow faculty and learners to understand and embrace roles and expectations in learning and assessment.
- Workflow, work assignment, and technologic strategies that support and reward supervisors to use direct observation of learner performance (with patients, on rounds, during chart review) to gather evidence and make formative judgments about competency of a given learner.
- Coaching strategies to enhance a learner's ability to understand and act upon feedback.
- Strategies and technologies to collect and display data on learner progress for use by learners and their coaches and supervisors.

Results that indicate equity in assessment include evidence that:

- Process indicators:
    - Assessment procedures are fully aligned with a clear understanding of their benefits and limitations in predicting future performance.
    - Assessment strategies are employed to increase educational and career opportunities for all learners and only function to prevent these opportunities when their rigor is unquestioned.
    - Assessment data are used as intended and not for situations for which their relevance is unproven.
    - Programs routinely investigate issues of validity, fairness, and equity in their programs of assessment and work to minimize population group differences that unfairly disadvantage any particular group.
- Outcome indicators:
    - Population group differences in educational and career opportunities for groups underrepresented in medicine are eliminated.

contribute to inequities in assessment and assessment outcomes.

CBME defines desired outcomes of medical education and outlines the expected trajectory of competency achievement using milestones.[55–57] Assessment in CBME is based on the belief that all individuals on the continuum of medical school into independent practice are continuing to learn and grow.[58] In contrast, traditional approaches to assessment focus not on supporting growth but on identifying learners with shortcomings.[59–62] This deficit approach disproportionately disadvantages UIM learners and misses the opportunity to use assessment information to foster a growth mindset in all learners. From a learning perspective, a focus on identifying struggling learners typically contradicts a developmental focus in which all learners are assumed to have areas in need of development and to be continuously learning and growing.[63] Any learner with less exposure or less rigorous premedical training, scenarios that disproportionately affect UIM learners, is particularly vulnerable to being labeled a struggling learner. Once this label is applied, the learner suffers the consequences, both personally and professionally.

Programmatic assessment focused on equity strives for a holistic, well-rounded view of individual learners and their trajectory. This aim is achieved by collecting and analyzing many samples of learners' work in multiple different contexts. In addition to quantitative data, qualitative assessments, such as narrative descriptors of performance, provide information about the nuances of performance that can guide learner improvement and contribute to rigorous decisions about learner competence.[64]

For further support of equity, programmatic assessment includes systems that rely on committees rather than individuals to make high-stakes, summative decisions about advancement and graduation.[65,66] Three critical elements of optimal committee structure and function can advance equity: group membership, data management, and decision-making procedures. A group invites opportunity for the "wisdom of the crowd," armed with a large number of data points, to draw on a diversity of opinions to make well-considered decisions. Diverse groups outperform individuals or homogeneous groups because members strengthen the quality of decisions made.[67] Training about common biases brings awareness of the human vulnerability for cognitive shortcuts and personal preferences that can shortchange every learner's opportunity for fair assessment.[68] The quality of data available and ease of accessing well-organized information strengthen group decision making. Absent well-rounded data from multiple sources, group members may default to making decisions based on impressions or limited data, a process that again introduces risk for bias. Finally, structured procedures for data review and group discussions help ensure that all learners are evaluated based on the totality of information available.

Systems must be designed to address contextual equity. Achieving equity in assessment outcomes requires attention to contextual equity. Both curricula and programs of assessment are highly dependent on other institutional systems, specifically those that orchestrate patient care and faculty support for teaching and learning. Educators who design systems of assessment must work with other leaders in academic health systems to optimize the learning environment for minority learners.

Systems must be designed to address instrumental equity and equity in assessment outcomes. Accreditors (Liaison Committee on Medical Education and Accreditation Council for Graduate Medical Education) and leaders of organizations that collect and use assessment data (National Resident Match Program, certifying boards) can support equity in assessment by designing systems of data presentation that align with principles of equity and by holding educational institutions and leaders accountable for using assessment data only for intended purposes.

Tools

Tools are the technologies, processes, incentives, and disincentives employed by individuals within the system to achieve desired results. Tools used in equitable assessment are selected to support and operationalize system priorities. Tools include evidence-seeking tools (direct observation of and discussion with learners about interactions with patients, clinical reasoning, and written work products), data display tools, data analytic tools, faculty development tools, and communication tools.

Results

Results are the measurable outcomes that demonstrate and affirm the assessment model's guiding principles. Outcomes serve an essential feedback mechanism for refining the tools and systems engineered to achieve the desired results. Results reinforce the culture of the institution, organization, or professional medical education community.

To fulfill our commitment to advancing equity in health care delivery and in medical education, individuals, institutions, and national organizations must commit to striving for the results outlined in List 1. Adhering to equity principles and evidence-based strategies for assessment does not guarantee that any given institution will avoid all population group differences in their classes. The small sample size of a given class and variability of student interest and aspirations from year to year may make a goal of eliminating any differences unfeasible. Instead, a better indicator that equity in assessment in medical education has been achieved will be the absence of national-level disparities for populations of learners from groups that are historically URM or that have been marginalized by medicine.

Research Agenda

The above proposed model for equity in assessment demands commitment to addressing inequities that have pervaded the medical education system. A robust research agenda to guide the collection of needed evidence demonstrating reduction in bias and discrimination is outlined below.

Research into intrinsic equity in assessment

Learner assessment: While much has been written about obstacles and barriers to success for UIM learners, the antideficit lens takes a different approach by drawing focus to the strengths and characteristics that position UIM learners for success. Harper's rigorous work used an antideficit lens to describe the factors that contributed to success rather than the shortcomings of Black men preparing

for and pursuing college education. This approach can serve as a model for studying the motivations, performance features, and individual and institutional characteristics that set UIM learners up for success.[69]

Educators will benefit from rigorous studies that explicate the characteristics of learner trajectories that predict success in practice. Educators can then determine which assessment methods can capture valid data about favorable learner trajectories and, thus, shift focus away from normative assessment methods.

Research into contextual equity in assessment

Learner experience: Understanding the influential aspects of the learning environment for UIM learners empowers researchers to examine relationships between the environment, learner satisfaction, and achievement to inform improvements that address any structural barriers impeding UIM learners' opportunities to maximize their achievement.[70] Another approach, with a focus on person, social, physical, and organizational elements of the learning environment, is Gruppen's framework, which holds potential for evaluating unique experiences of UIM learners and proposing solutions to address experiences that are counter to growth.[71]

The current configuration of the clinical learning environment, with dual aims to optimize both education and patient care, creates tensions that deserve investigation to find solutions. Further research is needed into systems and structures that foster meaningful relationships and trust between learners and supervisors, particularly those who may be of different backgrounds or identities, in a rotational model of medication education. This work must include a focus on optimizing organizational culture and climate. An organization may espouse diversity and inclusivity, but struggle to change long-standing ideologies of senior members or minimize microaggressions; evidence-based strategies for intervening on these problems are needed.

Research on instrumental equity in assessment

Selection of learners for positions in training programs is another aspect of learner assessment deserving of further research due to potential for conscious or unconscious bias to influence decision making. Current evidence examining admissions to health professions training programs comprises mainly single-institution studies but encouragingly shows that, in general, interventions are effective at diversifying class composition.[72] Further research is needed through multicenter, longitudinal studies to demonstrate effective approaches to not only selection but also academic support to ensure that UIM and other diverse learners achieve success in health professions careers. Research efforts must acknowledge tensions in assessment.[73] For example, the tension between assessment for learning vs assessment for ranking and selection remains central in discussions regarding equity in assessment.

Research on preparation of faculty for their roles supporting equity in assessment

Faculty development: Studies demonstrate the negative effects of physicians' implicit bias on their patients[74]; similar bias exists in evaluation of medical learners.[9,30] Research is needed to understand how to intervene and measure the benefits of faculty development to conduct high-quality assessment that minimizes the risks of bias.

Research into equity in assessment outcomes

Evidence is needed to confirm whether and how equitable assessment practices produce better outcomes for learners and the patients they serve. For example, interventions to enhance providers' cultural competency, though well-intentioned, do not all achieve desired outcomes.[75] Research is needed to examine what approaches to assessment optimize UIM learners' pursuit of careers in any and all specialties and settings as well as in academic careers.

Research into assessment for program evaluation

Counteracting inequity and bias in assessment requires a programmatic approach with careful attention to program evaluation. A realist approach to program evaluation meets these needs not only by asking what interventions work or don't work but also by considering the context of the interventions. Put simply, realist evaluation asks: "What works for whom in what circumstances, and why?"[76] Understanding the context allows other programs to consider the feasibility and applicability of interventions to their own context.

Challenges and Unintended Consequences

Any discussion of equity assessment creates unease for many and confronts the notion of merit. Termed "the foundational myth" of U.S. society, merit is thought to be a property of the individual, his (sic) work ethic and intrinsic talents.[77] Merit is Horatio Alger's alone—not a property of a social group or a parental legacy. Rivera's research, for example, describes how cultural capital enables children of the elite to: perform better in interviews and receive more coaching before interviews; have childhood experiences that allow for more bonding and mirroring with decision makers; have mistakes discounted (vs students for whom mistakes confirm stereotypes); and otherwise exhibit the fit, drive, skills, and talents necessary to secure a highly competitive job.[78] Challenging the existence of a true meritocracy remains an uphill battle.

Another challenge to discussions of equity assessment is the link to bias; many educators reject the notion of unconscious bias. Equity in assessment is about ensuring that we assess all learners for the skills, knowledge, and competencies required to care for their patients. Finally, no system of assessment will do away with group differences due to social inequities. In part, that is because elites adapt. The SAT test was developed to offer talented youth opportunities to elite colleges. Its developers never dreamed that test prep services would follow. Because the drive for equity in assessment is linked to equity in learning and opportunities that derive from social inequities, it must be a process of continuous quality improvement.

Conclusion

We have surveyed issues of equity in assessment, distinguishing various forms of assessment and various aspects of inequity. Informed by the evidence base

on inequity in assessment, this manuscript puts forth a framework for optimizing assessment to achieve equity. Key issues underlying debates on equity serve as the agenda for ongoing needed research and practice improvement to achieve equity in assessment that will ultimately improve patients' health. Individual medical schools can begin seeking solutions to the wicked problem of equity in assessment by working locally to design and continuously improve our learning and patient care ecosystems and by joining together to make equity in health professions education a national workforce priority.

Acknowledgments: The authors wish to thank the participants in the Josiah Macy Jr. Foundation conference on Equity in Assessment for their valuable insights.

Funding/Support: None reported.

Other disclosures: None reported.

Ethical approval: Reported as not applicable.

Previous presentations: The authors presented this work at the Josiah Macy Jr. Foundation conference on Equity in Assessment, February 24–27, 2020, Atlanta, Georgia.

C.R. Lucey is executive vice dean/vice dean for education and professor of medicine, University of California, San Francisco, School of Medicine, San Francisco, California.

K.E. Hauer is professor of medicine, University of California, San Francisco, School of Medicine, San Francisco, California.

D. Boatright is assistant professor of emergency medicine, Yale University School of Medicine, New Haven, Connecticut.

A. Fernandez is professor of medicine, University of California, San Francisco, School of Medicine, San Francisco, California.

References

1 Shen MJ, Peterson EB, Costas-Muniz R, et al. The effects of race and racial concordance on patient-physician communication: A systematic review of the literature. J Racial Ethn Health Disparities. 2018;5:117–140.
2 McGee R Jr, Saran S, Krulwich TA. Diversity in the biomedical research workforce: Developing talent. Mt Sinai J Med. 2012;79:397–411.
3 Hung R, McClendon J, Henderson A, Evans Y, Colquitt R, Saha S. Student perspectives on diversity and the cultural climate at a U.S. medical school. Acad Med. 2007;82:184–192.
4 Whitla DK, Orfield G, Silen W, Teperow C, Howard C, Reede J. Educational benefits of diversity in medical school: A survey of students. Acad Med. 2003;78:460–466.
5 Lett LA, Murdock HM, Orji WU, Aysola J, Sebro R. Trends in racial/ethnic representation among US medical students. JAMA Netw Open. 2019;2:e1910490.
6 Gonzaga AMR, Appiah-Pippim J, Onumah CM, Yialamas MA. A framework for inclusive graduate medical education recruitment strategies: Meeting the ACGME standard for a diverse and inclusive workforce. Acad Med. 2019;95:710–716.
7 Aibana O, Swails JL, Flores RJ, Love L. Bridging the gap: Holistic review to increase diversity in graduate medical education. Acad Med. 2019;94:1137–1141.
8 Conrad SS, Addams AN, Young GH. Holistic review in medical school admissions and selection: A strategic, mission-driven response to shifting societal needs. Acad Med. 2016;91:1472–1474.
9 Teherani A, Hauer KE, Fernandez A, King TE Jr, Lucey C. How small differences in assessed clinical performance amplify to large differences in grades and awards: A cascade with serious consequences for students underrepresented in medicine. Acad Med. 2018;93:1286–1292.
10 Rittel HW, Webber MM. Dilemmas in a general theory of planning. Policy Sci. 1973;4:155–169.
11 Jonker CM, Van Riemsdijk MB, Vermeulen B, Den Helder FB. Shared mental models: A conceptual analysis. In: De Vos M, et al. Coordination, Organization, Institutions and Norms in Agent Systems VI. New York, NY: Springer; 2010.
12 Gipps C. What do we mean by equity in relation to assessment? Assess Educ Princ Policy Pract. 1995;2:271–281.
13 Braveman P, Gruskin S. Defining equity in health. J Epidemiol Community Health. 2003;57:254–258.
14 Banks CAM, Banks JA. Equity pedagogy: An essential component of multicultural education. Theory Pract. 1995;34:152–158.
15 Norcini J, Anderson B, Bollela V, et al. Criteria for good assessment: Consensus statement and recommendations from the Ottawa 2010 Conference. Med Teach. 2011;33:206–214.
16 Marcus-Blank B, Dahlke JA, Braman JP, et al. Predicting performance of first-year residents: Correlations between structured interview, licensure exam, and competency scores in a multi-institutional study. Acad Med. 2019;94:378–387.
17 Cadieux G, Abrahamowicz M, Dauphinee D, Tamblyn R. Are physicians with better clinical skills on licensing examinations less likely to prescribe antibiotics for viral respiratory infections in ambulatory care settings? Med Care. 2011;49:156–165.
18 McGaghie WC, Cohen ER, Wayne DB. Are United States Medical Licensing Exam Step 1 and 2 scores valid measures for postgraduate medical residency selection decisions? Acad Med. 2011;86:48–52.
19 Wenghofer E, Klass D, Abrahamowicz M, et al. Doctor scores on national qualifying examinations predict quality of care in future practice. Med Educ. 2009;43:1166–1173.
20 Tamblyn R, Abrahamowicz M, Dauphinee WD, et al. Association between licensure examination scores and practice in primary care. JAMA. 2002;288:3019–3026.
21 Gomez-Garibello C, Young M. Emotions and assessment: Considerations for rater-based judgements of entrustment. Med Educ. 2018;52:254–262.
22 Gingerich A, Kogan J, Yeates P, Govaerts M, Holmboe E. Seeing the 'black box' differently: Assessor cognition from three research perspectives. Med Educ. 2014;48:1055–1068.
23 Wood TJ. Exploring the role of first impressions in rater-based assessments. Adv Health Sci Educ Theory Pract. 2014;19:409–427.
24 Gingerich A, Regehr G, Eva KW. Rater-based assessments as social judgments: Rethinking the etiology of rater errors. Acad Med. 2011;86(10 suppl):S1–S7.
25 Shah R, Ahluwalia S. The challenges of understanding differential attainment in postgraduate medical education. Br J Gen Pract. 2019;69:426–427.
26 Woolf K, Viney R, Rich A, Jayaweera H, Griffin A. Organisational perspectives on addressing differential attainment in postgraduate medical education: A qualitative study in the UK. BMJ Open. 2018;8:e021314.
27 Colbert CY, French JC, Herring ME, Dannefer EF. Fairness: The hidden challenge for competency-based postgraduate medical education programs. Perspect Med Educ. 2017;6:347–355.
28 Duffield KE, Spencer JA. A survey of medical students' views about the purposes and fairness of assessment. Med Educ. 2002;36:879–886.
29 King JT Jr, Angoff NR, Forrest JN Jr, Justice AC. Gender disparities in medical student research awards: A 13-year study from the Yale School of Medicine. Acad Med. 2018;93:911–919.
30 Boatright D, Ross D, O'Connor P, Moore E, Nunez-Smith M. Racial disparities in medical student membership in the Alpha Omega Alpha Honor Society. JAMA Intern Med. 2017;177:659–665.
31 Mueller AS, Jenkins TM, Osborne M, Dayal A, O'Connor DM, Arora VM. Gender differences in attending physicians' feedback to residents: A qualitative analysis. J Grad Med Educ. 2017;9:577–585.
32 Ross DA, Boatright D, Nunez-Smith M, Jordan A, Chekroud A, Moore EZ. Differences in words used to describe racial and gender groups in medical student performance evaluations. PLoS One. 2017;12:e0181659.
33 Rojek AE, Khanna R, Yim JWL, et al. Differences in narrative language in evaluations of medical students by gender and under-represented minority status. J Gen Intern Med. 2019;34:684–691.
34 Capers Q 4th, Clinchot D, McDougle L, Greenwald AG. Implicit racial bias in medical school admissions. Acad Med. 2017;92:365–369.
35 Davis D, Dorsey JK, Franks RD, Sackett PR, Searcy CA, Zhao X. Do racial and ethnic group differences in performance on the MCAT exam reflect test bias? Acad Med. 2013;88:593–602.
36 Association of American Medical Colleges. Using MCAT Data in 2019 Medical Student Selection. Washington, DC: Association of American Medical Colleges; 2018.
37 Genao I, Gelman J. The MCAT's restrictive effect on the minority physician pipeline: A legal perspective. Ann Intern Med. 2018;169:403–404.
38 Lucey CR, Saguil A. The consequences of structural racism on MCAT scores and medical school admissions: The past is prologue. Acad Med. 2020;95:351–356.

39 Green M, Jones P, Thomas JX Jr. Selection criteria for residency: Results of a national program directors survey. Acad Med. 2009;84:362–367.
40 Case SM, Becker DF, Swanson DB. Performances of men and women on NBME Part I and Part II: The more things change. Acad Med. 1993;68(10 suppl):S25–S27.
41 Dawson B, Iwamoto CK, Ross LP, Nungester RJ, Swanson DB, Volle RL. Performance on the National Board of Medical Examiners. Part I Examination by men and women of different race and ethnicity. JAMA. 1994;272:674–679.
42 Veloski JJ, Callahan CA, Xu G, Hojat M, Nash DB. Prediction of students' performances on licensing examinations using age, race, sex, undergraduate GPAs, and MCAT scores. Acad Med. 2000;75(10 suppl):S28–S30.
43 Rubright JD, Jodoin M, Barone MA. Examining demographics, prior academic performance, and United States Medical Licensing Examination scores. Acad Med. 2019;94:364–370.
44 Terregino CA, Saguil A, Price-Johnson T, Anachebe NF, Goodell K. The Diversity and success of medical school applicants with scores in the middle third of the MCAT score scale. Acad Med. 2020;95:344–350.
45 Andriole DA, Jeffe DB, Hageman HL, et al. Variables associated with full-time faculty appointment among contemporary U.S. medical school graduates: Implications for academic medicine workforce diversity. Acad Med. 2010;85:1250–1257.
46 Edmond MB, Deschenes JL, Eckler M, Wenzel RP. Racial bias in using USMLE step 1 scores to grant internal medicine residency interviews. Acad Med. 2001;76:1253–1256.
47 Hopson LR, Regan L, Bond MC, et al. The AAMC standardized video interview and the electronic standardized letter of evaluation in emergency medicine: A comparison of performance characteristics. Acad Med. 2019;94:1513–1521.
48 Brancati FL, Mead LA, Levine DM, Martin D, Margolis S, Klag MJ. Early predictors of career achievement in academic medicine. JAMA. 1992;267:1372–1376.
49 DeCroff CM, Mahabir RC, Zamboni WA. The impact of alpha omega alpha membership on successfully matching to residency. Plast Reconstr Surg. 2010;126:113e–115e.
50 Lynch G, Holloway T, Muller D, Palermo AG. Suspending student selections to Alpha Omega Alpha Honor Medical Society: How one school is navigating the intersection of equity and wellness. Acad Med. 2020;95:700–703.
51 Sondheimer HM. Graduating US medical students who do not obtain a PGY-1 training position. JAMA. 2010;304:1168–1169.
52 Centers for Disease Control and Prevention. The social-ecological model: A framework for prevention. https://www.cdc.gov/violenceprevention/publichealthissue/social-ecologicalmodel.html. Accessed April 10, 2020.
53 Wijesekera TP, Kim M, Moore EZ, Sorenson O, Ross DA. All other things being equal: Exploring racial and gender disparities in medical school honor society induction. Acad Med. 2019;94:562–569.
54 Plenert GJ, ed. Discover Excellence: An Overview of the Shingo Model and Its Guiding Principles. New York, NY: Productivity Press; 2017.
55 Frank JR, Snell LS, Cate OT, et al. Competency-based medical education: Theory to practice. Med Teach. 2010;32:638–645.
56 Swing SR; International CBME Collaborators. Perspectives on competency-based medical education from the learning sciences. Med Teach. 2010;32:663–668.
57 Lucey CR, Thibault GE, ten Cate O. Competency-based, time-variable education in the health professions: Crossroads. Acad Med. 2018;93(3 suppl):S1–S5.
58 Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR. The role of assessment in competency-based medical education. Med Teach. 2010;32:676–682.
59 Hauer KE, Teherani A, Kerr KM, Irby DM, O'Sullivan PS. Consequences within medical schools for students with poor performance on a medical school standardized patient comprehensive assessment. Acad Med. 2009;84:663–668.
60 Hauer KE, Ciccone A, Henzel TR, et al. Remediation of the deficiencies of physicians across the continuum from medical school to practice: A thematic review of the literature. Acad Med. 2009;84:1822–1832.
61 Chang A, Boscardin C, Chou CL, Loeser H, Hauer KE. Predicting failing performance on a standardized patient clinical performance examination: The importance of communication and professionalism skills deficits. Acad Med. 2009;84(10 suppl):S101–S104.
62 Kalet A, Guerrasio J, Chou CL. Twelve tips for developing and maintaining a remediation program in medical education. Med Teach. 2016;38:787–792.
63 Hauer KE, Chesluk B, Iobst W, et al. Reviewing residents' competence: A qualitative study of the role of clinical competency committees in performance assessment. Acad Med. 2015;90:1084–1092.
64 Cook DA, Kuper A, Hatala R, Ginsburg S. When assessment data are words: Validity evidence for qualitative educational assessments. Acad Med. 2016;91:1359–1369.
65 Frank AK, O'Sullivan P, Mills LM, Muller-Juge V, Hauer KE. Clerkship grading committees: The impact of group decision-making for clerkship grading. J Gen Intern Med. 2019;34:669–676.
66 Kinnear B, Warm EJ, Hauer KE. Twelve tips to maximize the value of a clinical competency committee in postgraduate medical education. Med Teach. 2018;40:1110–1115.
67 Hauer KE, Cate OT, Boscardin CK, et al. Ensuring resident competence: A narrative review of the literature on group decision making to inform the work of clinical competency committees. J Grad Med Educ. 2016;8:156–164.
68 Dickey CC, Thomas C, Feroze U, Nakshabandi F, Cannon B. Cognitive demands and bias: Challenges facing clinical competency committees. J Grad Med Educ. 2017;9:162–164.
69 Harper SR. An anti-deficit achievement framework for research on students of color in STEM. New Dir Institutional Res. 2010;2010:63–74.
70 Soemantri D, Herrera C, Riquelme A. Measuring the educational environment in health professions studies: A systematic review. Med Teach. 2010;32:947–952.
71 Gruppen LD, Irby DM, Durning SJ, Maggio LA. Conceptualizing learning environments in the health professions. Acad Med. 2019;94:969–974.
72 Simone K, Ahmed RA, Konkin J, Campbell S, Hartling L, Oswald AE. What are the features of targeted or system-wide initiatives that affect diversity in health professions trainees? A BEME systematic review: BEME guide no. 50. Med Teach. 2018;40:762–780.
73 Govaerts MJB, van der Vleuten CPM, Holmboe ES. Managing tensions in assessment: Moving beyond either-or thinking. Med Educ. 2019;53:64–75.
74 Maina IW, Belton TD, Ginzberg S, Singh A, Johnson TJ. A decade of studying implicit racial/ethnic bias in healthcare providers using the implicit association test. Soc Sci Med. 2018;199:219–229.
75 Brottman MR, Char DM, Hattori RA, Heeb R, Taff SD. Toward cultural competency in health care: A scoping review of the diversity and inclusion education literature. Acad Med. 2020;95:803–813.
76 Pawson R, Tilley N. Realistic Evaluation. Thousand Oaks, CA: SAGE Publications; 1997.
77 Markovits D. The Meritocracy Trap: How America's Foundational Myth Feeds Inequality, Dismantles the Middle Class, and Devours the Elite. New York, NY: Penguin Press; 2019.
78 Rivera LA. Pedigree: How Elite Students Get Elite Jobs. Princeton, NJ: Princeton University Press; 2015.