

Expanding the Study of Bias in Medical Education Assessment

Halah Ibrahim , MD, MEHP

Amy Miller Juve , EdD

Alpesh Amin , MD, MBA, MACP

Kenyon Railey , MD

Kathryn M. Andolsek , MD, MPH

Bias critically and adversely impacts the assessment of individuals at many stages in the developmental continuum of a physician's career, including admission to medical school, progression through training, career prospects and advancement, and termination decisions.^{1,2} Studies have focused primarily on disparities for women, and for racial and ethnic groups, but implicit and explicit bias also negatively affects many other groups underrepresented in medicine (UIM), such as LGBTQIA+ individuals and those with different abilities or from nonpredominant religious groups.^{3,4} Standardized test scores, clerkship grades, letters of recommendation, honor society memberships, research opportunities, and formative and summative assessments are all traditionally viewed as indicators of proficiency and predictors of future success, yet are also increasingly recognized as vulnerable to bias.⁵⁻⁷ Despite the prevalence of systemic bias in medical education, there are limited published reports of interventions to minimize effects from bias on important outcomes. Similar to other journals, the *Journal of Graduate Medical Education* (JGME) has received relatively few submissions that examine the extent of a broader, more systemic view of bias in, or promise practices to mitigate bias in, graduate medical education (GME) assessment practices.

Research shows that bias is evident at multiple critical points in medical education, including the initial acceptance decision to medical school, where standardized entrance examinations, like the Medical College Admission Test, disadvantage UIM applicants.^{8,9} Disparities persist throughout undergraduate medical education, where UIM students frequently receive lower clerkship grades.^{1,6} Bias has also been identified in narrative assessments, including the Medical Student Performance Evaluation (former Dean's letter) and letters of recommendation to GME programs.¹⁰ Letters written for women and UIM students contain fewer standout adjectives such as "exceptional"

or "outstanding" than those written for men and racial majority counterparts.¹⁰ During residency, many of these assessment biases persist. Racial and gender bias appears to exist in Accreditation Council for Graduate Medical Education competencies and Milestone achievement, as some studies have found that White residents attain a higher level of Milestone achievement than non-White trainees.¹¹ Similarly, women GME learners receive conflicting feedback regarding autonomy and assertiveness, whereas men residents and fellows receive more constructive feedback, progress through training at a faster pace, and are granted more autonomy than women.¹¹ These disparities can lead to what Teherani and colleagues describe as an "amplification cascade," in which small differences in assessment accumulate longitudinally and result in enduring disparities throughout later training and a physician's career.¹² Far fewer studies compare assessments of those of different abilities, gender identities, ethnic, or religious backgrounds in comparison to the majority group in a GME program, institution, or specialty. We have little information, aside from opinion pieces and personal essays, as to the perceived and actual effects of bias on assessments and careers. Given the evidence that shows important benefits—to health professionals as well as patients—of enhanced diversity in medicine, ensuring successful career growth for nonmajority individuals should be a priority.

Promising Areas for Future Study Individual Focus

Adapting some interventions that have been effective in decreasing bias in clinical encounters to the educational environment may prove beneficial in mitigating assessment bias. Many clinical strategies involve implicit bias training, with the hope that if physicians recognize their own biases, there will be a reduction in health care disparities.¹³ However, social science research has shown that recognizing implicit bias is not sufficient.¹⁴ Implicit bias training should also include

DOI: <http://dx.doi.org/10.4300/JGME-D-23-00027.1>

concrete strategies to reduce bias, such as perspective taking, stereotype replacement, and counterstereotype imaging.^{13,14} For an assessment example, an attending could utilize the perspective taking strategy when completing an assessment form by imagining themselves as the resident. Professional development exercises could include faculty members role-playing the part of a resident during a simulated feedback session.

Hagiwara et al suggest that minimizing the *impact* of bias through improved clinician communication may be a more realistic intervention target than reducing implicit bias.¹⁵ Studies demonstrate that implicit bias can manifest itself through body language, such as eye contact and body distance, and how one speaks to patients, rather than the content of speech. These strategies might be adapted from the clinician-patient interaction to teaching experiences and be practiced and reinforced. When providing critical feedback, educators could practice substituting more inclusive behavioral choices for negative nonverbal (gestures, eye contact, body stance) and paraverbal (tone pitch and volume of speech) behaviors.¹⁵ For example, body postures, such as crossing arms, leaning away from the trainee, or avoiding eye contact might be replaced by more inclusive behaviors, including maintaining an open body posture and comfortable eye contact, and leaning forward slightly to express engagement and interest. In this way, essential feedback might be provided in a manner that is more supportive and respectful of the trainee, and perhaps less influenced by bias. If practiced, over time these strategies might become more natural and automatic. Even when not explicitly linked to assessment, studies could examine whether nonverbal and paraverbal behaviors strengthen the trainee-educator relationship and provide positive role modeling for trainees who will deliver feedback to others.

Promising work contributed by Gonzalez et al for patient care may have applicability to inclusive teaching as well.¹⁶ Their findings suggest that, even when patients perceive bias, the outcome of the clinical encounter may still be positive depending upon the physician's subsequent actions.¹⁶ In focus groups with Black and Latinx patients, most participants reported that, after an incident of perceived bias, they most wanted acknowledgement of the biased behavior, followed by an apology. "Restoring the relationship...can lead to the same outcome as never having demonstrated bias in the first place."¹⁶ Similarly, educators can focus on repairing relationships with trainees after instances of perceived bias if open dialog is encouraged. Educators who recognize or are told of perceived bias can apologize to trainees, remain nondefensive, engage in faculty development

to learn different approaches, and intentionally practice more inclusive behaviors in the future.

Institutional Focus

Even when methods directed at individuals improve individual behaviors, institutional and organizational changes are likely needed to ensure equity in assessment. It is straightforward to recommend or even mandate GME institutional improvements in reducing assessment bias, but how can this be accomplished through feasible—as well as effective—strategies? There is minimal to no evidence demonstrating how residency programs can routinely evaluate their assessment modalities and practices for potential inequities among subgroups of learners and use continuous quality improvement methods to address bias.¹⁷ For example, the Canadian GME Competency by Design programmatic assessment initiative has had mixed results.¹⁸⁻²⁰ Faculty could be trained in how to use standardized rubrics or other tools that are criterion referenced, competency-based, and nonnormative, but it is unclear how to accomplish this, consistently over time, in a sustained manner.¹⁷ Furthermore, educators need to focus their assessments on direct observation of authentic work-based skills, such as entrustable professional activities, but studies report many barriers to direct observation assessments.^{20,21} Some programs have had success in inserting frequent, competency-based, and directly observed assessments, with variable acceptance by trainees and faculty; these innovations could be studied in other settings and programs.²² Recommendations to improve assessment also include that assessors “slow” down when assessing learners.²³ Bias is more likely in stressful settings with time pressures and fatigue.²³ Studies that feasibly introduce methods to reduce faculty stress, perhaps through teaching deceleration strategies—eg, taking a deep breath and centering oneself before assessment—could be a time-efficient strategy to study, through comparing assessments done with and without this brief maneuver.²³ Artificial intelligence and natural language models are also beginning to be used to aggregate assessments, but how these can best be adapted to eliminate or mitigate bias is yet unknown.²⁴ Kiyasseh and colleagues address the elephant in the room and the question to which, right now, we have no answer: “how much bias mitigation is sufficient.”²⁵

Conclusion

Bias in medical education assessment endures despite enhanced awareness. For those committed to reducing its influence and enhancing careers for diverse trainees, there are some individual and institutional

approaches that need to be studied and then disseminated for GME contexts. Single interventions will generally not work or not be sustained, and bias might paradoxically be increased.²⁶ There are insufficient studies overall and their findings are mixed, with disagreements regarding which strategies have merit. Multiple longitudinal interventions are likely to be the most effective but will be difficult and expensive to study. Gonzalez and colleagues' suggestion that "implicit bias recognition and management must be reframed as an epistemology of practice...essential to the professional identity of medical learners to be effective"¹⁶ may provide a useful, perhaps inspirational construct as we consider next steps in assessment bias. JGME welcomes your thoughts on this important topic.

References

- Low D, Pollack SW, Liao ZC, et al. Racial/ethnic disparities in clinical grading in medical school. *Teach Learn Med.* 2019;31(5):487-496. doi:10.1080/10401334.2019.1597724
- Klein R, Julian KA, Snyder ED, et al. Gender bias in resident assessment in graduate medical education: review of the literature. *J Gen Intern Med.* 2019;34(5):712-719. doi:10.1007/s11606-019-04884-0
- Fallin-Bennett K. Implicit bias against sexual minorities in medicine: cycles of professional influence and the role of the hidden curriculum. *Acad Med.* 2015;90(5):549-552. doi:10.1097/ACM.0000000000000662
- Schwarz CM, Zetkovic M. You belong in the room: addressing the underrepresentation of physicians with physical disabilities. *Acad Med.* 2019;94(1):17-19. doi:10.1097/ACM.0000000000002435
- Teherani A, Perez S, Muller-Juge V, Lupton K, Hauer KE. A narrative study of equity in clinical assessment through the antideficit lens. *Acad Med.* 2020;95(suppl 12):121-130. doi:10.1097/ACM.0000000000003690
- Ramakrishnan D, Van Le-Bucklin K, Saba T, Levenson G, Kim JH, Elfenbein DM. What does honors mean? National analysis of medical school clinical clerkship grading. *J Surg Educ.* 2022;79(1):157-164. doi:10.1016/j.jsurg.2021.08.022
- Burkhardt JC, Parekh KP, Gallahue FE, et al. A critical disconnect: residency selection factors lack correlation with intern performance. *J Grad Med Educ.* 2020;12(6):696-704. doi:10.4300/JGME-D-20-00013.1
- Robinet K, Kareem R, Reavis K, Quezada S. A multi-pronged, antiracist approach to optimize equity in medical school admissions. *Med Educ.* 2021;55(12):1376-1382. doi:10.1111/medu.14589
- Rubright JD, Jodoin M, Barone MA. Examining demographics, prior academic performance, and United States Medical Licensing Examination scores. *Acad Med.* 2019;94(3):364-370. doi:10.1097/ACM.0000000000002366
- 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(8):e0181659. doi:10.1371/journal.pone.0181659
- Dayal A, O'Connor DM, Qadri U, Arora VM. Comparison of male vs female resident Milestone evaluations by faculty during emergency medicine residency training. *JAMA Intern Med.* 2017;177(5):651-657. doi:10.1001/jamainternmed.2016.9616
- Teherani A, Hauer KE, Fernandez A, King TE Jr, Lucey CR. 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(9):1286-1292. doi:10.1097/ACM.0000000000002323
- Ogunyemi D. A practical approach to implicit bias training. *J Grad Med Educ.* 2021;13(4):583-584. doi:10.4300/JGME-D-21-00598
- Devine PG, Forscher PS, Austin AJ, Cox WTL. Long-term reduction in implicit race bias: a prejudice habit-breaking intervention. *J Exp Soc Psychol.* 2012;48(6):1267-1278. doi:10.1016/j.jesp.2012.06.003
- Hagiwara N, Kron FW, Scerbo MW, Watson GS. A call for grounding implicit bias training in clinical and translational frameworks. *Lancet.* 2020;395(10234):1457-1460. doi:10.1016/S0140-6736(20)30846-1
- Gonzalez CM, Deno ML, Kintzer E, Marantz PR, Lypson ML, McKee MD. Patient perspectives on racial and ethnic implicit bias in clinical encounters: implications for curriculum development. *Patient Educ Couns.* 2018;101(9):1669-1675. doi:10.1016/j.pec.2018.05.016
- McClintock AH, Fainstad T, Jauregui J, Yarris LM. Countering bias in assessment. *J Grad Med Educ.* 2021;13(5):725-726. doi:10.4300/JGME-D-21-00722.1
- Thoma B, Hall AK, Clark K, et al. Evaluation of a national competency-based assessment system in emergency medicine: a CanDREAM study. *J Grad Med Educ.* 2020;12(4):425-434. doi:10.4300/JGME-D-19-00803.1
- Yilmaz Y, Carey R, Chan TM, et al. Developing a dashboard for program evaluation in competency-based training programs: a design-based research project. *Can Med Educ J.* 2022;13(5):14-27. doi:10.36834/cmej.73554
- Szulewski A, Braund H, Dagnone DJ, et al. The assessment burden in competency-based medical education: how programs are adapting [published online ahead of print June 21, 2023]. *Acad Med.* doi:10.1097/ACM.0000000000005305

21. Paterson QS, Alrimawi H, Sample S, et al. Examining enablers and barriers to entrustable professional activity acquisition using the theoretical domains framework: a qualitative framework analysis study. *AEM Educ Train*. 2023;7(2):e10849. doi:10.1002/aet2.10849
22. Burk-Rafel J, Sebok-Syer SS, Santen SA, et al. Trainee Attributable & Automatable Care Evaluations in Real-time (TRACERs): a scalable approach for linking education to patient care. *Perspect Med Educ*. 2023;12(1):149-159. doi:10.5334/pme.1013
23. Moulton CA, Regehr G, Mylopoulos M, MacRae HM. Slowing down when you should: a new model of expert judgment. *Acad Med*. 2007;82(suppl 10):109-116. doi:10.1097/ACM.0b013e3181405a76
24. Yilmaz Y, Jurado Nunez A, Ariaeinejad A, Lee M, Sherbino J, Chan TM. Harnessing natural language processing to support decisions around workplace-based assessment: machine learning study of competency-based medical education. *JMIR Med Educ*. 2022;8(2):e30537. doi:10.2196/30537
25. Kiyasseh D, Laca J, Haque TF, et al. Human visual explanations mitigate bias in AI-based assessment of surgeon skills. *NPJ Digit Med*. 2023;6(1):54. doi:10.1038/s41746-023-00766-2
26. Railey K, Barnett J. Is one lecture enough? Self-perception of bias and cultural training in medical education. *J Physician Assist Educ*. 2022;33(3):222-228. doi:10.1097/JPA.0000000000000452



Halah Ibrahim, MD, MEHP, is Associate Professor, Department of Medicine, Khalifa University College of Medicine and Health Sciences, Abu Dhabi, United Arab Emirates, and Associate Editor, Journal of Graduate Medical Education (JGME); **Amy Miller Juve, EdD**, is Professor of Anesthesiology and Perioperative Medicine and Professional Development, and Program Improvement Specialist for Graduate Medical Education, Oregon Health & Science University, Portland, Oregon, USA; **Alpesh Amin, MD, MBA, MACP**, is Professor, Department of Medicine, University of California Irvine, Irvine, California, USA; **Kenyon Railey, MD**, is Associate Professor, Department of Family Medicine and Community Health, School of Medicine, Duke University School of Medicine, Durham, North Carolina, USA; and **Kathryn M. Andolsek, MD, MPH**, is Assistant Dean for Premedical Education and Professor, Department of Family Medicine and Community Health, Duke University School of Medicine, Durham, North Carolina, USA, and Associate Editor, JGME.

Corresponding author: Halah Ibrahim, MD, MEHP, Khalifa University College of Medicine and Health Sciences, Abu Dhabi, United Arab Emirates, halah.ibrahim@ku.ac.ae, X @HalahlbrahimMD