

Let's Face It: We Are Biased, and It Should Not Be That Way

Paul A. Hemmer, MD, MPH¹ and Reena Karani, MD, MHPE²



¹Department of Medicine, Uniformed Services University of the Health Sciences, Bethesda, MD, USA; ²Departments of Medical Education, Medicine, and Geriatrics and Palliative Medicine, Institute for Medical Education, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

J Gen Intern Med 34(5):649–51

DOI: 10.1007/s11606-019-04923-w

© Society of General Internal Medicine (This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply) 2019

Let's face it—we're all biased. For example, in studies of letters of evaluation and recommendation in academia, men are more often described with words considered “stand-out”¹ (e.g., excellent, superb, exceptional), and/or related to “ability”² (e.g., talent, intelligent, capacity, analytical), and with an emphasis on “research”,³ while women are more often described using words considered “grindstone”^{1, 2} (e.g., hard-working, conscientious, diligent), and with an emphasis on “teaching”;³ other studies have found that women are more commonly described with “communal” words (e.g., compassionate, caring, empathetic).³

In our educational system, descriptive evaluations of trainees have been a hallmark of assessment in the clinical setting. Gingerich et al.⁴ explored differing perspectives on the role of assessors (e.g., faculty) (as trainable, fallible [biased], and meaningfully idiosyncratic). In the discussion of fallibility, the authors note that humans are not (and cannot be) passive observers, but activate cognitive processes to help remember and make sense of their observations, although these cognitive processes are the sources of biases.

Two articles in this issue of JGIM address the issue of bias (gender and race) in the assessment of medical students and residents. Rojek and colleagues⁵ conducted an observational study of over 87,000 core clerkship evaluations of medical students at two geographically different medical schools. They employed natural language processing to identify differences in the text of evaluations of women compared to men and of underrepresented in medicine (URIM) compared with non-URIM students. Their analysis demonstrated significant differences in narrative language associated with gender and URIM status, even among students receiving the same grade. Female students and URIM students were significantly more likely to be described using personal attributes as opposed to competency-based behaviors. Specifically, for words that differed between men and women, 62% represented personal attributes and were found more commonly for women, while 19% represented competency-related behaviors and were

more commonly applied to men. For words that differed between URIM and non-URIM students, comments about personal attributes were more common for URIM students, while competency-based comments were more frequent in evaluations of non-URIM students. The study also shed light on the ten most commonly used words that did not differ between the groups, providing insight that these words may not convey meaningful information about a student. The findings highlight the presence of bias within the descriptive evaluation of medical students, raise questions about the validity of these assessments in the evaluation process, including the Medical Student Performance Evaluation (MSPE), and serve as a call-to-action for concerted efforts to address and work to eliminate bias in all forms across the learning environment.

Klein et al.⁶ performed a systematic review of the literature from 1998 to 2018 to examine the presence and influence of gender bias on resident assessment. Five of the nine studies uncovered evidence of gender bias, with women receiving lower ratings in domains of performance and differences in qualitative comments. In the largest study,⁷ these differences in ratings were equivalent to adding several months of training for women to “catch up” to their male counterparts. In the study⁸ of the qualitative comments made by faculty in these emergency medicine (EM) programs, the “ideal” trainee had more stereotypically masculine traits, there was discordance in the feedback provided to struggling trainees (men received more consistent feedback while women received more discordant feedback, especially with regard to autonomy and assertiveness), and comments about a tension between autonomy and openness to feedback were only found in narratives of women. In other included studies, qualitative comments women received from faculty more often focused on communal or warmth-based descriptors and less often on agentic (related to agency) or competency-related descriptors. The authors of this review concluded that gender bias poses a potential threat to the integrity of resident assessment and there is a need to seek and understand the source(s) of the bias (e.g., in the assessment, the learner, the faculty), as well as the longer term consequences.

Since conducting their review, two articles have addressed gender bias in GME. In an observational study of one academic internal medicine training program,⁹ women scored higher in two domains (medical interviewing, and interpersonal and communication skills), although the absolute differences were small (effect sizes of approximately 0.12–0.15). A second qualitative study¹⁰ of the narrative comments of

resident trainees across nine surgical specialties at one institution found that men were consistently described by more positive comments than women. The authors noted that in their study, "...women were often described as having a potential to succeed while men were simply expected to succeed."¹⁰ In addition, some phrases were only written about women, notably in the area of disposition and humanism (e.g., "always smiling"). Finally, while there were no differences in the relative use of communal, grindstone, and ability words, men were more likely to receive "standout" words, with the word "leader" used three times more for men than women. The differences in the use of standout words, feedback to women about assertiveness, and speaking with confidence are also experienced by medical students.¹¹

Similarly, there are differences in how different racial groups among medical trainees are described. In a study of MSPEs¹² for 6000 applicants from 134 medical schools applying for 16 residency programs at one institution, White students were more likely to be described using "standout" or "ability" words while Black students were more likely to be described as "competent." In another single-institution study representing 4655 medical students from 123 US medical schools, Black students were less likely to be members of the Alpha Omega Alpha Medical Honor Society than White students, even after controlling for other measures of achievement.¹³ In a qualitative study of 27 underrepresented (Black, Hispanic, Native American) residents in GME training,¹⁴ themes emerged around microaggressions and bias, being tasked to be ambassadors, and challenges of personal and professional identity formation. The practice of routinely citing race (about which we typically don't ask patients) in introductory statements¹⁵ ("This is a 65 year old African-American woman....") reflects the difficulty of shifting views of race from a long-held and perpetuated biologic perspective to that of a social construct.¹⁶ With such deeply engrained specious beliefs, it is not surprising, but very concerning, to see these influences in the descriptions of trainees.

While it would be naïve of us to believe we will eliminate all bias in assessment programs, we must commit ourselves to action. We must address the gender and racial bias evident in the descriptive evaluations faculty provide about trainees, since fairness and equity in the process of assessment (fairness to trainees, faculty, and society) are essential for the integrity of our learning environment. What does this mean for our educational, including assessment, processes, programs, and leaders? The answer: Let's face it. And, we have to change it.

Morgan et al.¹⁷ outlined a series of steps in addressing gender-based inequity, which are equally applicable to other forms of bias. The papers in this issue help with addressing the first step, which is to be explicit and name the systematic nature of the bias. In leading and participating in educational programs, understanding and highlighting the words people use in their assessments are part of making this more explicit, as is empowering trainees to come forward when they read these

words about themselves. Implementing educational interventions that focus on dialogue among participants is another means of making the underlying bias explicit.¹⁸ Other steps the authors suggest include advancing efforts to promote equity, incorporating explicit evidence-based training, and increasing transparency on issues such as academic promotion.

We would not recommend eliminating or marginalizing descriptive evaluation of trainees as that would not address the fundamental underlying issue of bias (e.g., gender, race). We should be guiding faculty on what to say in their comments. We can illustrate the words that are helpful to explain their observations (including the 10 words shown in the Rojek⁶ study that might be of limited help). But we should not mislead ourselves and think that a new form or a new online training program will solve the issue. Forms help communicate goals but dialogue and conversation lead to understanding. If decisions are to be made by someone other than the author about what comments to include or not include in a final narrative (course, clerkship, rotation, MSPE), we need to ask: Who will make the decision about what to keep, delete, or change? How will the faculty member be provided feedback? How do we know the person editing is not influenced by their own bias(es)? Do we have a responsibility to the readers of the narrative to inform them if changes have been made? Should personal attributes be eliminated from comments or could they be desirable to help more fully describe an individual? Do we risk swinging the pendulum too far, yielding descriptions that risk becoming too formulaic and impersonal?

Collectively, we have to help one another to say what matters. Our best chance of doing this is to find and take the time to talk with one another. Ultimately, it is in our daily dialogue and through naming the problem that we will mitigate bias and not perpetuate it.

Corresponding Author: Paul A. Hemmer, MD, MPH; Department of Medicine Uniformed Services University of the Health Sciences, Bethesda, MD, USA (e-mail: paul.hemmer@usuhs.edu).

Compliance with Ethical Standards:

Conflict of Interest: The authors do not have any conflicts of interest to declare.

Disclaimer: The opinions and assertions expressed herein are those of the authors and do not necessarily reflect the official policy or position of the Uniformed Services University or the Department of Defense.

REFERENCES

1. **Trix F, Psenka C.** Exploring the Color of Glass: Letters of Recommendation for Female and Male Medical Faculty. *Discourse & Society*. 2003;14(2):191-220.
2. **Schmader T, Whitehead J, Wysocki VH.** A Linguistic Comparison of Letters of Recommendation for Male and Female Chemistry and Biochemistry Job Applicants. *Sex Roles*. 2007;57(7-8):509-514
3. **Axelson RD, Solow CM, Ferguson KJ, Cohen MB.** Assessing implicit gender bias in medical student performance evaluations. *Evaluation & the health professions*. 2010;33(3):365-85.

4. **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(11):1055–68. <https://doi.org/10.1111/medu.12546>.
5. **Rojek AE, Khanna R, Yim JW, Gardner R, Lisker S, Hauer KE, Lucey C, Sarkar U.** Differences in Narrative Language in Evaluations of Medical Students by Gender and Under-represented Minority Status. *J Gen Intern Med* <https://doi.org/10.1007/s11606-019-04889-9>
6. **Klein R, Julian KA, Snyder ED, Koch J, Ufere NN, Volerman A, Vandenberg AE, Schaeffer S, Palamara K.** Gender Bias in Resident Assessment in Graduate Medical Education: Review of the Literature. *J Gen Intern Med* 2017 10(4):411–415. <https://doi.org/10.1007/s11606-019-04884-0>
7. **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.
8. **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(5):577–585.
9. **Sulistio MS, Khera A, Squiers K, Sanghavi M, Ayers CR, Weng W, Kazi S, de Lemos J, Johnson DH, Kirk L.** Effects of gender in resident evaluations and certifying examination pass rates. *BMC Med Educ.* 2019;19(1):10. <https://doi.org/10.1186/s12909-018-1440-7>.
10. **Gerull KM, Loe M, Seiler K, McAllister J, Salles A.** Assessing gender bias in qualitative evaluations of surgical residents. *Am J Surg.* 2019;217(2):306–313.
11. **Treadway N. A** Student Reflection on Doctoring With Confidence: Mind the Gap. *Ann Intern Med.* 2018;169(8):564–565. <https://doi.org/10.7326/M18-0865>
12. **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. <https://doi.org/10.1371/journal.pone.0181659>. eCollection 2017
13. **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(5):659–665. <https://doi.org/10.1001/jamainternmed.2016.9623>.
14. **Osseo-Asare A, Balasuriya L, Huot SJ, Keene D, Berg D, Nunez-Smith M, Genao I, Latimore D, Boatright D.** Minority Resident Physicians' Views on the Role of Race/Ethnicity in Their Training Experiences in the Workplace. *JAMA Netw Open.* 2018;1(5):e182723. <https://doi.org/10.1001/jamanetworkopen.2018.2723>
15. **Finucane TE.** Mention of a Patient's "Race" in Clinical Presentations. *Virtual Mentor.* 2014;16(6):423–427. <https://doi.org/10.1001/virtualmentor.2014.16.6.ecas1-1406>.
16. **Jones CP.** Levels of racism: a theoretic framework and a gardener's tale. *Am J Public Health.* 2000;90(8):1212–5
17. **Morgan AU, Chaiyachati KH, Weissman GE, Liao JM.** Eliminating Gender-Based Bias in Academic Medicine: More Than Naming the "Elephant in the Room". *J Gen Intern Med.* 2018;33(6):966–968. <https://doi.org/10.1007/s11606-018-4411-0>.
18. **Bakke K, Sidhar K, Kumagai AK.** Exploring Matters of Race through Dialogue in the University of Michigan Medical School's Longitudinal Case Studies Program. *Virtual Mentor.* 2014;16(6):442–449. <https://doi.org/10.1001/virtualmentor.2014.16.6.medu1-1406>.

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