

Proceed, With Caution: Unconscious Bias in Technical Assessment

Danielle S. Walsh, MD, FACS, FAAP

In 1964, Supreme Court Justice Potter Stewart made famous the expression, “I know it when I see it,” while admitting difficulty accurately converting the human impression of a visual depiction into written form. Lawyer William T. Goldberg went on to critique Stewart’s statement by stating, “I know it when I see it, and someone else will know it when they see it, but what they see and what they know may or may not be what I see and what I know.”¹ The challenges described by these esteemed jurists are not dissimilar to those facing clinical faculty in the operating room who need to assess competency and autonomy in surgical trainees. Unfortunately, the intersection point between impression and competency assessment proves to fluctuate based on gender of the subject, gender of the evaluator, and bias, requiring the implementation of tools and training to reset objectivity.

At the turn of the century, increasing effort on the part of surgical educators went into transitioning the subjective technical performance of a trainee into quantifiable format. While prior technical assessments were largely end-of-rotation summative impressions, operative performance rating systems were developed to quantify performance in a specific operation immediately following the event.² Over time, technical assessments have been refined and corroborated with performance over a variety of procedures and specialties. In recognition that technical competency and the ability to autonomously complete an operation are distinct though interrelated skills, the Zwisch scale for assessing operating room autonomy was introduced and widely adopted to complement assessments of technique.³ Most, if not all, trainees in accredited, technically based training programs now receive assessments of both competencies using numerical scoring systems. Additionally, the conversion from paper-based assessment forms to digital apps immediately available on smartphones has increased adoption and application of these valuable assessment tools.⁴

From the earliest proposals for use of such assessment tools, enthusiasm has grown for a transition from time-based medical education to

competency-based completion of training. However, to truly attain this Holy Grail, it is incumbent on medical educators that the tools and selected terminus for completion are proven not only feasible, but also reproducible, correlative with ability, and free of bias. Failure to do so in the transition to a “meritocratic” training paradigm risks further lowering the glass ceiling on women surgeons while opening windows for early exit to men who better align with the ideal mental model of their faculty assessors. Two articles from recent issues of the *Journal of Graduate Medical Education* demonstrate the need to further optimize this transition to competency assessments. Chen et al report a multicenter review of nearly 95 000 assessments and identified the development of a gap in autonomy ratings for women over the course of surgical trainees, with the largest gap occurring in the most complex cases.⁵ Additionally, women trainees self-rated as having less autonomy and worse performance than men, even after controlling for most other potential confounding elements. Cooney et al’s review of 8377 assessments of plastic surgery residents in 3 different programs found women trainees were scored significantly lower than men in the first 4 years of residency when rated by men attending surgeons, but not when rated by women attending surgeons.⁶ Both articles add to the large body of evidence that bias creeps into clinical assessment, and mitigation is required before transition to graduation based on competency assessments alone.

Implicit bias, or the impact of unconscious beliefs, attitudes, or stereotypes on human perception, has increasingly been implicated as negatively impacting the health care environment. While initially emphasizing the impact of such bias on patient care, societal changes are now putting the spotlight on implicit bias in the workplace as well. In 2018, the Association of Women Surgeons convened a task force and published best practices for addressing implicit bias in surgical departments.⁷ Following acknowledgement that eradication is necessary, leaders are encouraged to introduce departmental and individual assessment for bias, education for recognition and understanding, exposure to counter stereotypic individuals for resetting of impressions, and deliberate policy for

bias constraint. In a similar vein, Ogunyemi demonstrated increased perception and knowledge around implicit biases in graduate medical education programs through a 90-minute workshop.⁸ Whether through standardized courses or internal efforts, it is critical that faculty and trainees completing both self-assessments and assessments of others develop awareness of the impact of their own biases on scoring. With the rollout of any assessment tool, education on not only how to use the tool but also how to avoid bias in the application of the tool is needed. Following pilot implementations, an intentional search for unconscious bias in the outcomes should be obligatory prior to wide application of the tool for high-stakes utilization, such as assessment of physician competency to practice.

As pointed out by Goldberg when considering the comments of Supreme Court Justice Stewart,¹ the different life experiences of diverse people create different biases and interpretations. As educators, it is incumbent that we work to widen the lens in which we look at others in the face of evidence that there is still work to be done in the identification and elimination of gender bias in training and assessment.

References

1. Goldberg W. Two Nations, One Web: Comparative Legal Approaches to Pornographic Obscenity By The United States And The United Kingdom. *Boston University Law Review*. http://www.bu.edu/law/journals-archive/bulr/documents/goldberg_000.pdf. Accessed August 12, 2021.
2. Larson JL, Williams RG, Ketchum M, Boehler ML, Dunnington GL. Feasibility, reliability, and validity of an operative performance rating system for evaluating surgery residents. *Surgery*. 2005;138(4):640–649. doi:10.1016/j.surg.2005.07.017
3. DaRosa DA, Zwischenberger JB, Meyerson SL, et al. A theory-based model for teaching and assessing residents in the operating room. *J Surg Educ*. 2013;70(1):24–30. doi:10.1016/j.jsurg.2012.07.007
4. Reynolds K, Barnhill D, Sias J, Young A, Polite FG. Use of the QR reader to provide real-time evaluation of residents' skills following surgical procedures. *J Grad Med Educ*. 2014;6(4):738–741. doi:10.4300/JGME-D-13-00349.1
5. Chen JX, Chang EH, Deng F, et al. Autonomy in the operating room: a multicenter study of gender disparities during surgical training. *J Grad Med Educ*. 2021;13(5):666–672. doi:10.4300/JGME-D-21-00217.1
6. Cooney CM, Aravind P, Hultman CS, et al. An analysis of gender bias in plastic surgery resident assessment. *J Grad Med Educ*. 2021;13(4):500–506. doi:10.4300/JGME-D-20-01394.1
7. DiBrito SR, Lopez CM, Jones C, Mathur A. Reducing implicit bias: Association of Women Surgeons #HeForShe Task Force best practice recommendations. *J Am Coll Surg*. 2019;228(3):303–309. doi:10.1016/j.jamcollsurg.2018.12.011
8. Ogunyemi D. Defeating unconscious bias: the role of a structured, reflective, and interactive workshop. *J Grad Med Educ*. 2021;13(2):189–194. doi:10.4300/JGME-D-20-00722.1



Danielle S. Walsh, MD, FACS, FAAP, is Associate Professor of Surgery and General Surgery Program Director Emeritus, East Carolina University and Vidant Medical Center.

Corresponding author: Danielle S. Walsh, MD, FACS, FAAP, East Carolina University, walshd@ecu.edu, Twitter @walshds