Reducing Bias in Evaluations

DR VYDIA PERMASHWAR

DR EMILY NGUYEN

DEPARTMENT OF PEDIATRICS

CARILION CHILDREN HOSPITAL

Conflicts

We have no conflicts of interest

Goals and Objectives

Goals: Explore and address issues relating to bias in evaluations of medical trainees and its potential effect on the learning environment, satisfaction and career implications

Objectives:

- Review current literature on extent of bias in assessment of learners in medical education
- Discuss implications of bias in evaluations
- Provide strategies and tips to improve objectivity and mitigate bias in written evaluations through interactive exercises

Question

One of the educators is working on their assessment methods. They ask for your feedback:

"Is there any evidence of racial/ethnic or gender bias in educational assessment?"

How do you answer?

A. Only in AOA selection
B. Yes, in Clerkship grades only
C. Yes, in Clerkship evaluation narrative language only
D. Yes, in Clerkship grades, evaluations, AOA selection, MSPE letters, and Resident performance evaluation

Question: What percentage of medical students believe they are evaluated based on performance only?

A. 0-15%

B. 15-30%

C. 30-45%

D. 45-60%

E. 60-75%

Question

"Physicians have similar levels of bias as lay persons"

1) T

2) F

Learner Assessments & Evaluations

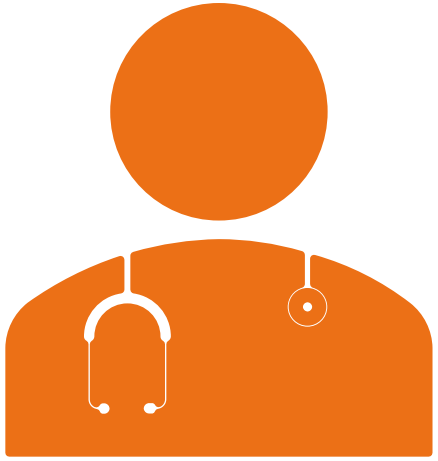
What is the purpose of our evaluations?

- Provide constructive feedback to our learners to help them improve
- Ensure competencies are met
- Students: determine letters of distinction and forms Dean's letter and MSPE in residency applications
- Residents: determine milestone attainment and potentially fellowship and career selection

What evaluations are used and how are these obtained?

- Most evaluations come from attending faculty in several settings (inpatient, outpatient, subspecialties)
- Time spent with learner can vary from an afternoon, to weeks or even a month
- Usually come from individual preceptors, though faculty may seek input from residents/students, nursing, or others to form consensus evaluations
- Categories include: medical knowledge, interpretation, management, data-gathering, reporting, procedural skills, communication, relationships, professionalism, educational attitudes, systems-based

Students' perceptions of grading



-In a survey sent to a sample of 4th year medical students: 67% reported concerns that clinical clerkship grades, narrative comments, and/or feedback were NOT based solely on performance

-In another survey, only 44% of medical students believe that clerkship grading is fair.

What did they consider the two most important factors?

- **Being liked and which doctor students worked with — not clinical reasoning or medical knowledge**

Least important in determining final grades?

- Improvement

"Many students believe that if they are not perfect on day one, they cannot earn top grades — and that hard work, learning, and improving are not rewarded as much as the good luck to work with team members who like you."

Healing a broken clerkship grading system
Justin Bullock, MD, MPH; Karen E. Hauer, MD, PhD
February 20, 2020 AAMC
Healing a broken clerkship grading system | AAMC

The problem(s)

Preceptors tasked with educating and evaluating many learners at different levels of training

- Medical students from multiple institutions
- Residents may be from different specialties

The need to provide helpful detailed formative feedback in a busy clinical environment with competing demands

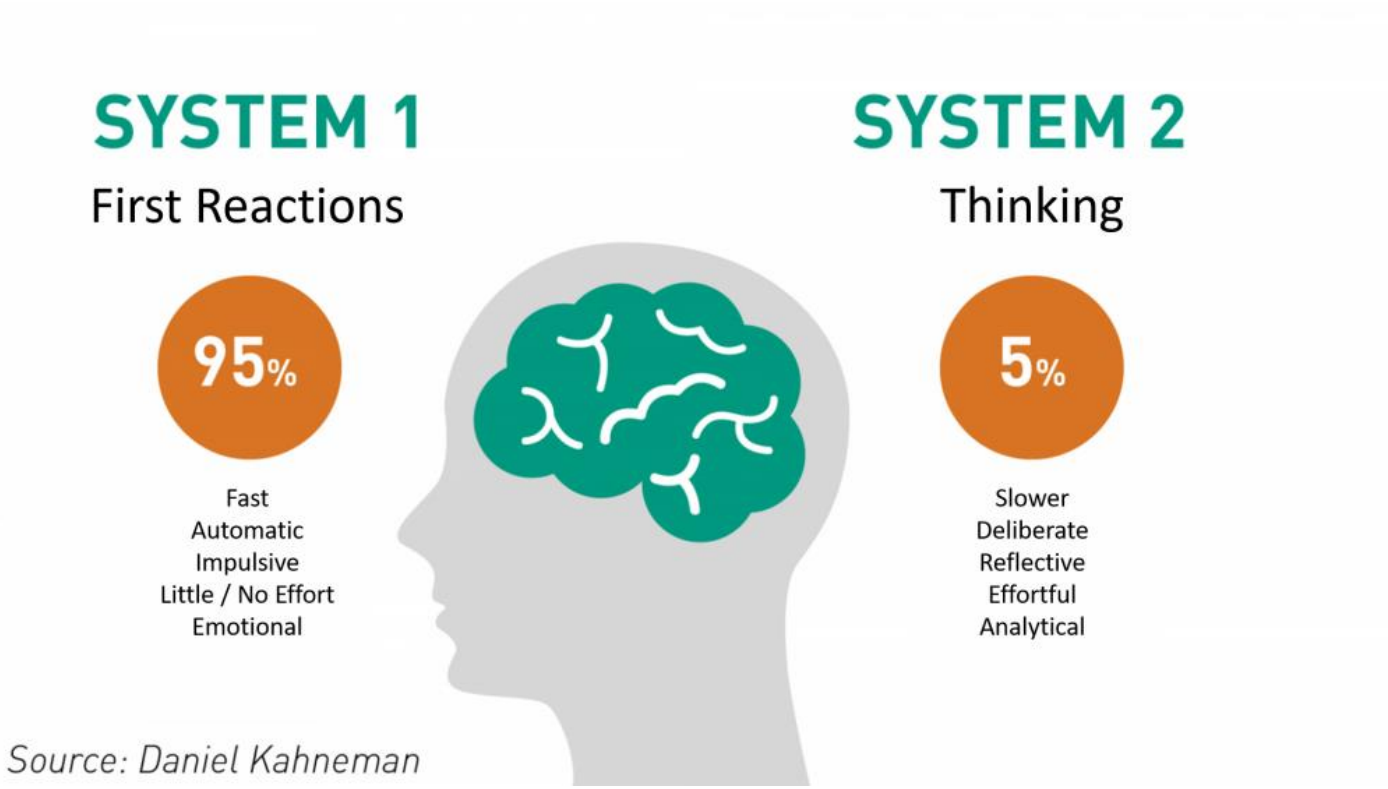
Lack of objectivity/standardization

- Many different preceptors/evaluators at many different sites
- Varying time spent with learners
- Evaluations may be written weeks after interacting with learners
- Lack of standardized training in evaluations

Unconscious bias

The problem: unconscious or implicit bias

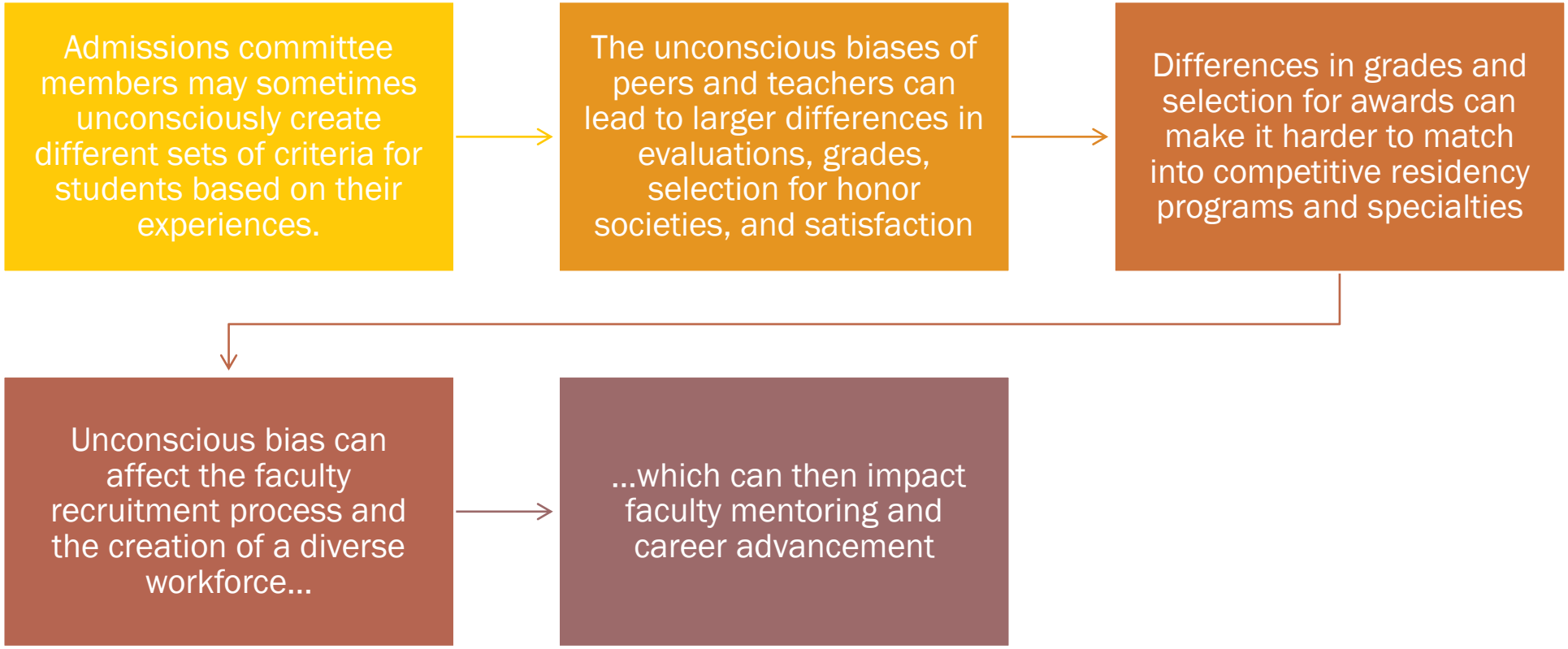
- “Caused by well-intentioned people with blind spots” – Howard Ross
- Automatic mental shortcut that falls back on previous experiences and exposures
- Unconscious bias can extend to race/ethnicity, gender, sexual orientation, weight, height, age, social class, and more



Types of Bias

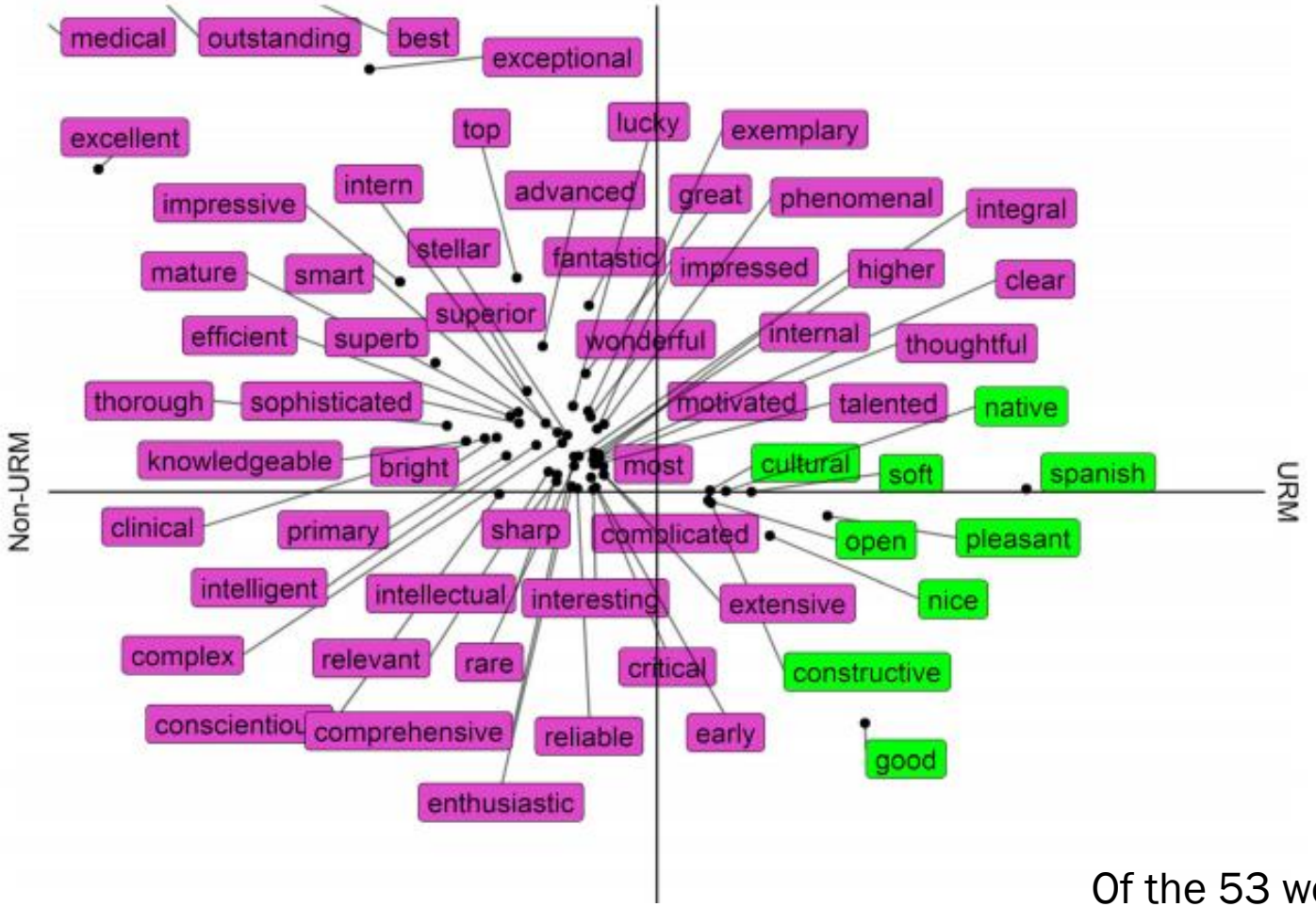
- Confirmation bias: searching for evidence that confirms your initial answer or impression
- Implicit bias: unconscious attitudes, assumptions or stereotypes that may be outside one's awareness or control
- Availability bias: basing an assessment off a specific easily remembered instance that may not accurately represent the learner's performance
- Hawk/dove effect: an evaluator is stricter (hawk) or easier (dove) overall compared to others

Why It Matters: The Amplification Cascade



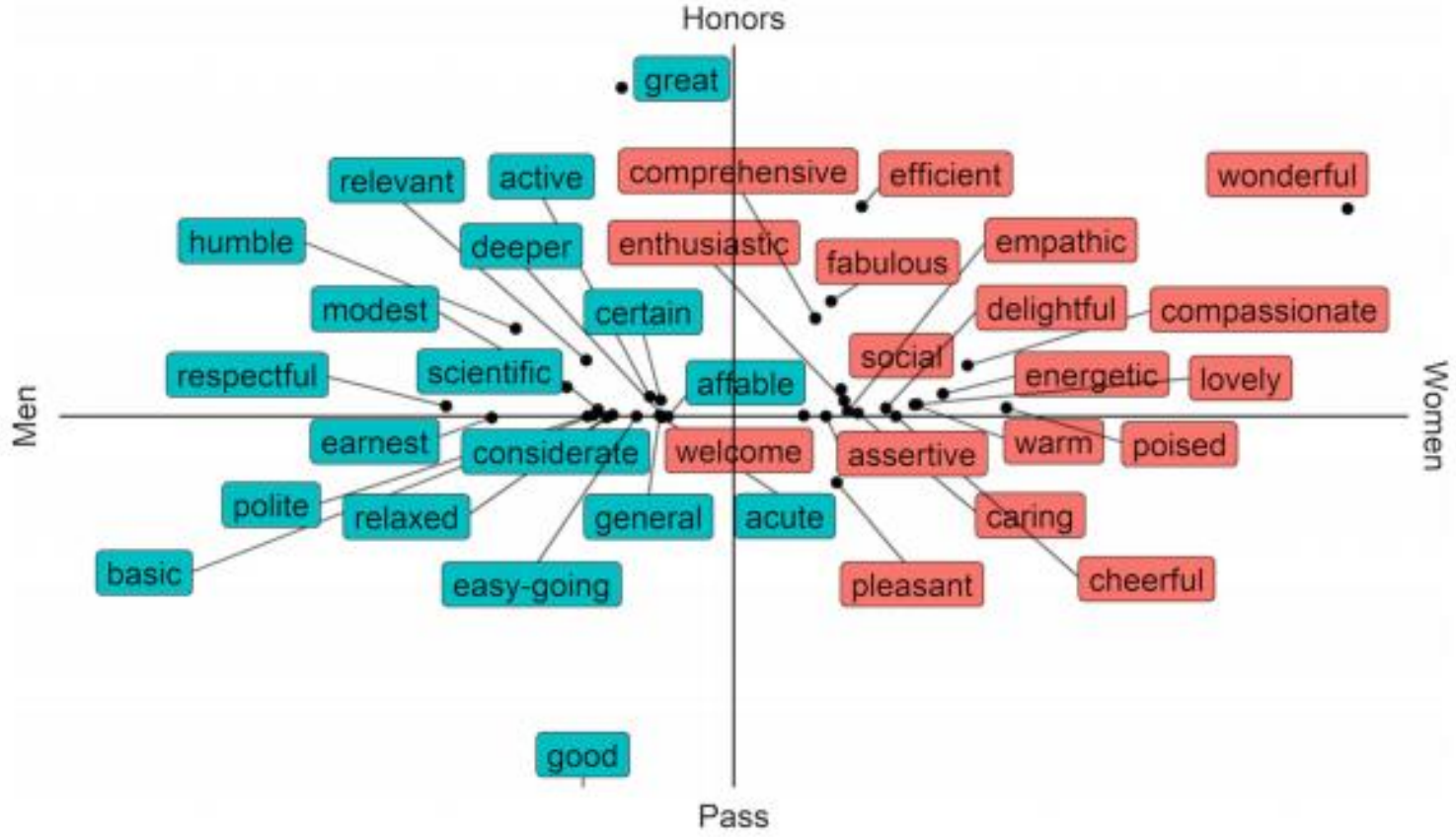
Words we use to describe our learners

Ability/competency	Talented, smart, capable, knowledgeable
Personality traits	Pleasant, lovely
Standout traits	Excellent, outstanding
Grindstone adjectives	Hardworking, diligent
Doubt raisers	Only required minimum supervision



Differences in Narrative Language in Evaluations of Medical Students by Gender and Under-represented Minority Status (UCSF) 2019

Of the 53 words that differed by URM status, 28% were competency-related, all of which were used more in evaluations of non-URM students.



Differences in Narrative Language in Evaluations of Medical Students by Gender and Under-represented Minority Status (UCSF) 2019

Of the 37 words that differed by gender, 62% represented personal attributes. Of these, 57% were used more frequently in evaluations of women.

In another study, women were more likely than men to be described with words relating to compassion, such as “kind,” “caring,” and “empathic” (Ross et al 2017).

What about differences in clerkship grades and honor society selection?

Differences between URM and non-URM students

- Overall students who identified as white consistently received higher final clerkship grades and more honors (Low 2019, Tehereani et al, 2018), whereas URM students receive lower grades across clerkships (Lee et al 2007).
- URM, Black, and Asian students were 2-6x less likely to be selected for honor society membership such as AOA, after adjusting for Step 1 score, research publications, leadership activity (Teherani et al, 2018, Wijesekera, et al 2019, Boatright, et al 2017)

Gender-based differences in clerkship are less clear-cut

- Women were more likely to receive honors in pediatrics, OB/GYN, neurology and psychiatry
- Men were more likely to receive honors in surgery and anesthesia (Rojek et al, 2019)

GME Evidence- Gender Difference

Comparison of Male vs Female Resident Milestone Evaluations by Faculty During Emergency Medicine Residency Training .

Who Gets the Benefit of the Doubt? Performance Evaluations, Medical Errors, and the Production of Gender Inequality in Emergency Medical Education

- Analysis of 2,765 performance evaluations in EM no gender bias in year 1, however, in year 3, men were perceived as outperforming women.
- In 3rd year but not the 1st, women received more harsh criticism and less supportive feedback than men for medical errors of similar severity
- Although male and female residents received similar evaluations at the beginning of residency, the rate of milestone attainment throughout training was higher for male than female residents across all EM sub-competencies in 8 EM programs.

Brewer A, et al. American Sociological Review. 2020; 85(2):247-270.
Dayal A, et al. JAMA Intern Med. 2017 May 1;177(5):747

Evaluation examples

“She was warm, caring and empathetic”

- Analysis: These are characteristics that evaluators tend to focus more on for women than men; the evaluator should comment on other competencies as well
- Instead: describe clinical skills, knowledge and interactions with patients and team

“He worked hard through the rotation”

- Analysis: Effort is commendable; it is also important to describe performance and connect effort to accomplishments
- Instead, consider using this language: “Due to his hard work creating a discharge plan, our team was able to discharge the patient safely home”

Watch for and ask the student and other team members about student contributions you may not have observed.

Student contribution	Competencies	Description
Spending time with a patient explaining a diagnosis that was unclear to the patient on rounds	Interpersonal and communication skills Patient care	'The student spent extra time with the patient explaining his diagnosis and answering questions to ensure his understanding and provide reassurance.'
Working on discharge planning to ensure that the patient will receive all of her medications, have secure housing, and understand her follow up appointments	Systems-based practice Interprofessional collaboration	'The student coordinated discharge planning for a complex patient discharge by working with the with the pharmacist, case manager, and resident to ensure that the patient understood the discharge plans. This included finding a pharmacy to provide all of the prescribed medications, securing temporary housing and explaining the plans to the patient to confirm her understanding.'
Answering questions from a patient's family about an upcoming procedure	Interpersonal and communication skills	'When a patient's family had questions about an upcoming procedure, the student listened to their questions and concerns with empathy. The student then coordinated with the resident to answer all of their questions accurately

Breakout rooms

1

Review example of student evaluation

2

Discuss how it could contain potential biased language and/or be less helpful for student

3

Make recommendations on how to edit or improve it

Evaluation example

Jamie did well in this clerkship. In written notes, Jamie demonstrated good application of medical knowledge but often seemed aloof, not participating in rounds or answering questions. Despite this, patients seemed to respond well. Written notes adequately addressed the issues for the day. Residents found that Jamie was cheerful, enthusiastic and hardworking.

What can we do to improve?

On an individual level

- Being intentional with language used to describe learner
- Using competency-based language as opposed to personal attributes
- Use of specific and objective examples in evaluations
- Consider getting input from multiple team members when writing evaluation
- Consider your implicit biases and use strategies to mitigate them
- Consider using free online tool to look for gender bias in eval: https://www.tomforth.co.uk/genderbias/
- Analyze trends and your practices
- Cognitive control maneuvers

Systematically

- Increase the number and types of assessments (360 approach)
- Consider changes to evaluation forms
- Consider having a separate faculty member/committee edit or review evaluations in a blinded fashion, narratives
- Consider bias training, such as use of the Implicit Association Test and follow up
- Train all faculty members on how to evaluate learners effectively and objectively
- Consider using prompts about intentional language and implicit biases in evaluation forms
- Review trends of institution

More on Intentional Language

Study of video recordings of Grand Rounds at 2 institutions

- Women nearly always used the title “doctor” to introduce speakers (96%)
- Men who made introductions used it 66% of the time:
- When men introduced men, they used formal titles 73% of the time
- When men introduced women this dropped to 49%

Files JA, Mayer AP, Ko MG, et al. Speaker introductions at internal medicine grand rounds: forms of
address reveal gender bias. J Womens Health (Larchmt)2017;26:4139.doi:10.1089/jwh.2016.6044

Systematic Approaches

- “Student evaluations of teaching play an important role in the review of faculty. Your opinions influence the review of instructors that takes place every year. Iowa State University recognizes that student evaluations of teaching are often influenced by students’ unconscious and unintentional biases about the race and gender of the instructor. Women and instructors of color are systematically rated lower in their teaching evaluations than white men, even when there are no actual differences in the instruction or in what students have learned. As you fill out the course evaluation please keep this in mind and make an effort to resist stereotypes about professors…”

Peterson DAM, Biederman LA, Andersen D, Ditonto TM, Roe K. Mitigating gender bias in student evaluations of teaching. PLoS One 2019; 14:e0216241.doi:10.1371/journal.pone.0216241

Strategies to Reduce Implicit Bias

Implicit bias in individual interactions can be addressed and countered if we become aware of our bias and take actions to redirect our responses.(Devine and colleagues offer six strategies to reduce implicit bias):

Stereotype replacement — Recognizing that a response is based on stereotype and consciously adjusting the response

Counter-stereotypic imaging — Imagining the individual as the opposite of the stereotype

Individuation — Seeing the person as an individual rather than a stereotype (e.g., learning about their personal history and the context that brought them in contact with you)

Perspective taking — “Putting yourself in the other person’s shoes”

Increasing opportunities for contact with individuals from different groups — Expanding one’s network of friends and colleagues or attending events where people of other racial and ethnic groups, gender identities, sexual orientation, and other groups may be present

Partnership building — Reframing the interaction with the person as one between collaborating equals, rather than between a high-status person and a low-status person

1) Beware of biases that exist with grading/evaluations

Consider taking the Implicit Association Test to increase awareness of your own biases

2) Be specific, detailed, and give examples

Use observations instead of inferences

Give actionable feedback in real time (shorter but more frequent)

Ask other team members for more well-rounded feedback

Describe specific examples of competency- and behavior-related skills

3) Take your time

Start early. Take notes.

Give yourself dedicated time and space to complete evaluations.

4) Pause before hitting "submit"

If I were the student reading this evaluation, is this helpful in improving my clinical skills or learning?

Is my evaluation biased—based on gender, race/ethnicity, or how likeable the student is?

References

Bullock, Justin, and Karen E. Hauer. “Healing a Broken Clerkship Grading System.” AAMC, 20 Feb. 2020, www.aamc.org/news-insights/healing-broken-clerkship-grading-system.

Dolan, Brigid, and Khalilah Gates. Addressing Bias in Learner Assessment, http://idd.northwestern.edu/elm/addressing_bias/story.html

“Equity in Assessment Guidelines and Checklist.” Equity in Assessment Guidelines and Checklist | UCSF Medical Education, https://meded.ucsf.edu/faculty-educators/equity-assessment-guidelines-and-checklist

Hauer, Karen E. MD, PhD; Lucey, Catherine R. MD Core Clerkship Grading: The Illusion of Objectivity, Academic Medicine: April 2019 - Volume 94 - Issue 4 - p 469-472 doi: 10.1097/ACM.0000000000002413

Lee KB, Vaishnavi SN, Lau SK, Andriole DA, Jeffe DB. "Making the grade: noncognitive predictors of medical students' clinical clerkship grades.” J Natl Med Assoc. 2007 Oct;99(10):1138-50. PMID: 17987918; PMCID: PMC2574397.

Low D, Pollack SW, Liao ZC, Maestas R, Kirven LE, Eacker AM, Morales LS. Racial/Ethnic Disparities in Clinical Grading in Medical School. Teach Learn Med. 2019 Oct-Dec;31(5):487-496. doi: 10.1080/10401334.2019.1597724. Epub 2019 Apr 29. PMID: 31032666.

Rojek AE, Khanna R, Yim JWL, 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. 2019 May;34(5):684-691. doi: 10.1007/s11606-019-04889-9. PMID: 30993609; PMCID: PMC6502922. https://link.springer.com/content/pdf/10.1007/s11606-019-04889-9.pdf

Ross DA, Boatright D, Nunez-Smith M, Jordan A, Chekroud A, Moore EZ (2017) Differences in words used to describe racial and gender groups in Medical Student Performance Evaluations. PLoS ONE 12(8): e0181659. https://doi.org/10.1371/journal.pone.0181659

Resources

Northwestern module: http://idd.northwestern.edu/elm/addressing_bias/story.html

Core Clerkship Grading: The Illusion of Objectivity : Academic Medicine (lww.com)

Avoiding Stereotypes and Bias in Assessment of Learner Performance: file:///C:/Users/amars/Downloads/Avoiding%20stereotypes%20in%20assessment.pdf

Equity in Assessment Checklist: file:///C:/Users/amars/Downloads/Equity_in_Assessment_Checklist_2020.pdf

Good assessment practice: file:///C:/Users/amars/Downloads/Good%20assessment%20practice%20-%20evalution%20examples.pdf

Lehigh Best Practices for Reading and Writing Letters of Recommendation: file:///C:/Users/amars/Downloads/Lehigh%20Best%20Practices%20for%20Reading%20and%20Writing%20Letters%20of%20Recommendation.pdf

Educational equity

-Many learners and educators have turned their attention the parallels between disparities in health care and disparities in access to education and opportunity within medicine.

-Educational equity is now recognized as a core principle of undergraduate medical education.

Thank you!

Implicit Biases, Interprofessional Communication, and Power Dynamics

Erin Stephany Sanchez, MD, Melody Tran-Reina, MD, Kupiri Ackerman-Barger, PhD, RN, Kristine Phung, MD, Mithu Molla, MD, MBA, and Hendry Ton, MD, MS | April 29, 2020

m& m case study, patient safety network, root cause analysis

Give yourself ***dedicated time and space*** to complete student evaluations.

We all experience ***decision fatigue*** (quality of decision deteriorates as day progresses).

2.) GIVE EXAMPLES

Describe student performance in the setting of a specific setting. Examples include:

- **Oral presentations**
- **Patient encounters**
- **Actions within the team / team dynamics**

Knowledge:

Accurate
Articulate
Competent
Deductive
Logical
Systematic

Professionalism:

Adaptable
Appropriate
Organized
Productive
Reliable
Respectful

Patient Care:

Compassionate
Effective
Efficient
Insightful
Methodical
Skilled

Interpersonal:

Attentive
Collaborative
Comprehensive
Humble
Non-judgmental
Perceptive

Learn More at * link to website *

All Other Things Being Equal: Exploring Racial and Gender Disparities in Medical School Honor Society Induction

Results:

-Women were more likely than men to be inducted into GHHS (odds ratio 1.84, P < .001) but did not differ in their likelihood of being inducted into AOA.

-Black medical students were less likely to be inducted into AOA (odds ratio 0.37, P < .05) but not into GHHS

-Adjusting for Step 1 score, research publications, citizenship status, training interruptions, and year of application.

-Gender- and race-matched samples to account for differences in clerkship grades and to test for bias.

Wijesekera et al, Acad Med. 2019; 94(4):562-569.

Diversity among physicians is critically important to reducing health disparities, and reducing implicit bias is key to a diverse medical workforce.

Rooting out implicit bias in admissions Quinn Capers 4,AAMC

-**Annual, mandatory implicit bias mitigation training sessions:** All our application screeners and admissions committee members participate in 45-minute moderated discussions of implicit bias vignettes and evidenced-based strategies to reduce bias. In addition, the admissions dean leads 2 1/2-hour implicit bias workshops throughout the year for the entire medical center community, which admissions committee members often voluntarily attend.

-**Recommended readings on implicit bias**

-**Interview day "cheat sheet":** Before meeting a candidate, interviewers review a bulleted list of strategies to reduce implicit bias. Strategies include "Consider the Opposite," in which the reader decides about an applicant's qualifications but then re-reviews the file looking for evidence supporting the opposite conclusion before making a final decision.

Avoiding the Virtual Pitfall: Identifying and Mitigating Biases in Graduate Medical Education Videoconference Interviews – Marbin, J

This article identifies some of the biases VCIs can introduce to the recruitment process and offers strategies for programs to mitigate them.

These include making interviewers aware of potential technology-based inequities, encouraging interviewers to minimize multitasking, and offering guidance on use of standardized backgrounds.

The authors also recognize the limitations of offering behavioral strategies to mitigate systemic inequities and suggest that structural changes are needed to ensure equitable access to technology.

The Consequences of Structural Racism on MCAT Scores and Medical School Admissions: The Past Is Prologue

Lucey, Catherine Reinis MD; Saguil, Aaron MD, MPH

The AAMC and The Ohio State University Kirwan Institute for the Study of Race and Ethnicity convened a forum in 2014 to examine how unconscious bias affects academic medicine and to identify strategies to mitigate the impact.

Unconscious Bias in Academic Medicine: How the Prejudices We Don't Know We Have Affect Medical Education, Medical Careers, and Patient Health.

Healing a broken clerkship grading system

ur institution, for example, we found that African American, Latinx/Hispanic, and Native American/Pacific Islander students consistently receive slightly lower average scores (about one-tenth of a point) in all clerkships, a difference that was magnified as it translated into their receiving half as many top grades. Similarly, both URM and non-URM (such as Asian) minority students at another institution received lower grades than white students in most clerkships, even after adjusting for confounding variables, suggesting that implicit racial bias likely played a role. Discrimination is not uncommon in medical school: More than 40% of graduates report experiencing bias based on their race, gender, or other personal trait, according to AAMC data.

An analysis of more than 87,000 written evaluations showed that, although there were no differences by race, gender, or ethnicity in the 10 words supervisors used most often, other important words did show such differences. Men and non-URM students were more often described based on their competence, with words like “scientific” and “knowledgeable,” while women and URM students were more often described by their personality, with words such as “pleasant” and “lovely.

Justin Bullock, MD, MPH
Karen E. Hauer, MD, PhD
https://www.aamc.org/news-insights/healing-broken-clerkship-grading-system

Reimagining Merit and Representation: Promoting Equity and Reducing Bias in GME Through Holistic Review

Nicolás E Barceló 1 Acad Psychiatry 2021

Relative to Traditional, Holistic Review significantly increased the odds of URM applicant selection for interview (TR-OR: 0.35 vs HR-OR: 0.84, $p < 0.01$).
Assigning value to lived experience and de-emphasizing USMLE STEP1 scores contributed to the significant changes in odds ratio of interview selection for URM applicants.

Conclusions: Traditional interview selection methods systematically exclude URM applicants from consideration without due attention to applicant strengths or potential contribution to clinical care. Conversely, holistic screening represents a structural intervention capable of critically examining measures of merit, reducing bias, and increasing URM representation in residency recruitment, screening, and selection.

Medical schools implemented holistic review more than a decade ago, which led to more deliberate consideration and inclusion of applicants historically underrepresented in medicine

Trends in Race/Ethnicity of Pediatric Residents and Fellows: 2007-2019

Kimberly Montez[1], Emma A Omoruyi[2], Kenya McNeal-Trice[3], Wendy J Mack[4], Lahia Yemane[5], Alissa R Darden[6], Christopher J Russell[7 8] Pediatrics 2021

Results: Trends in URiM proportions were unchanged in residents (16% in 2007 to 16.5% in 2019; *P* = .98) and, overall, decreased for fellows (14.2% in 2007 to 13.5% in 2019; *P* = .002).

- URiM fellow trends significantly decreased over time in neonatal-perinatal medicine (*P* < .001), infectious diseases (*P* < .001), and critical care (*P* = .006) but significantly increased in endocrinology (*P* = .002) and pulmonology (*P* = .009).
-Over time, the percentage of URiM pediatric trainee representation was considerably lower compared to the US population

Differences in Narrative Language in Evaluations of Medical Students by Gender and Under-represented Minority Status

Of the 37 words that differed by gender, 62% represented personal attributes… more frequently in evaluations of women ($p < 0.001$), while 19% represented competency-related behaviors… more frequently in evaluations of men ($p < 0.001$).

Adjectives describing personal attributes are more likely to be ascribed to women (“lovely”) or URM students (“pleasant”)

Of the 53 words that differed by URM status, 30% represented personal attributes… more frequently in evaluations of URM students ($p < 0.001$), and 28% represented competency-related behaviors… more frequently in evaluations of non-URM students ($p < 0.001$).

Words describing competence are more likely to be ascribed to men (“scientific”) or non-URM students (“knowledgeable”)

Gender/Race

For the first time 2019- women accounted for more than half of students entering medical school. Yet the number of practicing female physicians over the past decade has remained below 35 percent.

Medical school faculty continued to be predominantly White (63.9%) and male (58.6%) overall, and especially so at the professor and associate professor ranks (AAMC).

Is there gender bias in medical student evaluations and/or the training process?

There remains persistent underrepresentation of certain racial and ethnic minority groups and women in medical school faculty positions

PERSPECTIVES VIEWPOINTS Robust evidence demonstrates inequities in clerkship grades that may be related to individual and structural bias, such as those seen in faculty evaluations and the clerkship grading process. Evidence for effective strategies to minimize bias in clerkship assessment is limited.

Examples of recommendations include:

- faculty/resident development,

- workplace-based assessments with criterion-based rubrics,

-competency-based non-normative grading,

-grading committees,

- limiting weight of standardized knowledge-based examination scores

- eliminating standardized examination score cut-offs for Honors grades,

- increasing the number and types of assessments,

- examining inequities in clerkship grades within the organization

Strategies to Minimize Narrative Bias

- Intentional Narrative language– table of options
- Intentional competency inclusion
- Group decision making
- Blinded editing of narratives
- Implicit bias training
- Systematic approaches

Group Decision Making

- Synthesize multiple data points in a standardized and consistent manner
- Social decision scheme theory
- Sharing and processing information-> better decisions
- Examples: CCC or Grading committee

Blinded Evaluations

- Can someone edit or review evaluations in a blinded fashion to create summative narratives?

Systematic approaches

- Consider changes to your evaluation form or evaluation system
- Prompts about intentional narrative language
- Prompts to consider implicit biases that may be present
- Requiring all competencies be evaluated

How Do I Talk about Implicit Bias Without Making People Defensive?

Ways to improve grading

Know the extent of the problem:

Consider pass/fail grading: frequent feedback to students. In our new system, faculty coaches who do not participate in high-stakes assessment help students interpret feedback from residents and attendings and set learning goals in the context of trusting relationships. Since implementation, we have noticed a visible decrease in the stress of our medical students around assessment

address broader issues of bias

Train assessors in student evaluation In addition, training should include completion of an Implicit Association Test (IAT) to increase awareness of implicit biases we all hold and how these biases may impact assessments of students.

Give better-quality feedback: actionable, real time ?? App for real time feedback

Core Clerkship Grading: The Illusion of Objectivity

Hauer, Karen E. MD, PhD; Lucey, Catherine R. MD
Author Information Academic Medicine: April 2019 - Volume 94 - Issue 4 - p 469-472