Intealth™

Advancing the Global Health Workforce

ECFMG FAIMER

Addressing and Reducing Bias in Assessment for Health Professions Education

Eric Holmboe

Copyright © 2024 by Intealth. All rights reserved.

Disclosures

- Eric Holmboe works for Intealth – however, the majority the work presented here was developed during his time at ACGME. He also receives royalties for a textbook on assessment from Elsevier Publishing.
- Portions of this presentation were developed by Dr. Dowin Boatright MD

Copyright © 2024 by Intealth. All rights reserved.

Agenda

- What is bias?
- Brief review of research on bias in assessment
- Interventions and approaches to mitigate bias

Copyright © 2024 by Intealth. All rights reserved.

Forms of Bias

- **Structural bias:**
  - Involves institutional (e.g. medical schools, hospitals) patterns and practices that advantage some groups and disadvantage other groups based on personal and demographic characteristics and identity
    - Represents a group level effect produced by institutional norms
- **Explicit bias:**
  - Refers to ***conscious*** beliefs and attitudes one possesses about another person or groups.

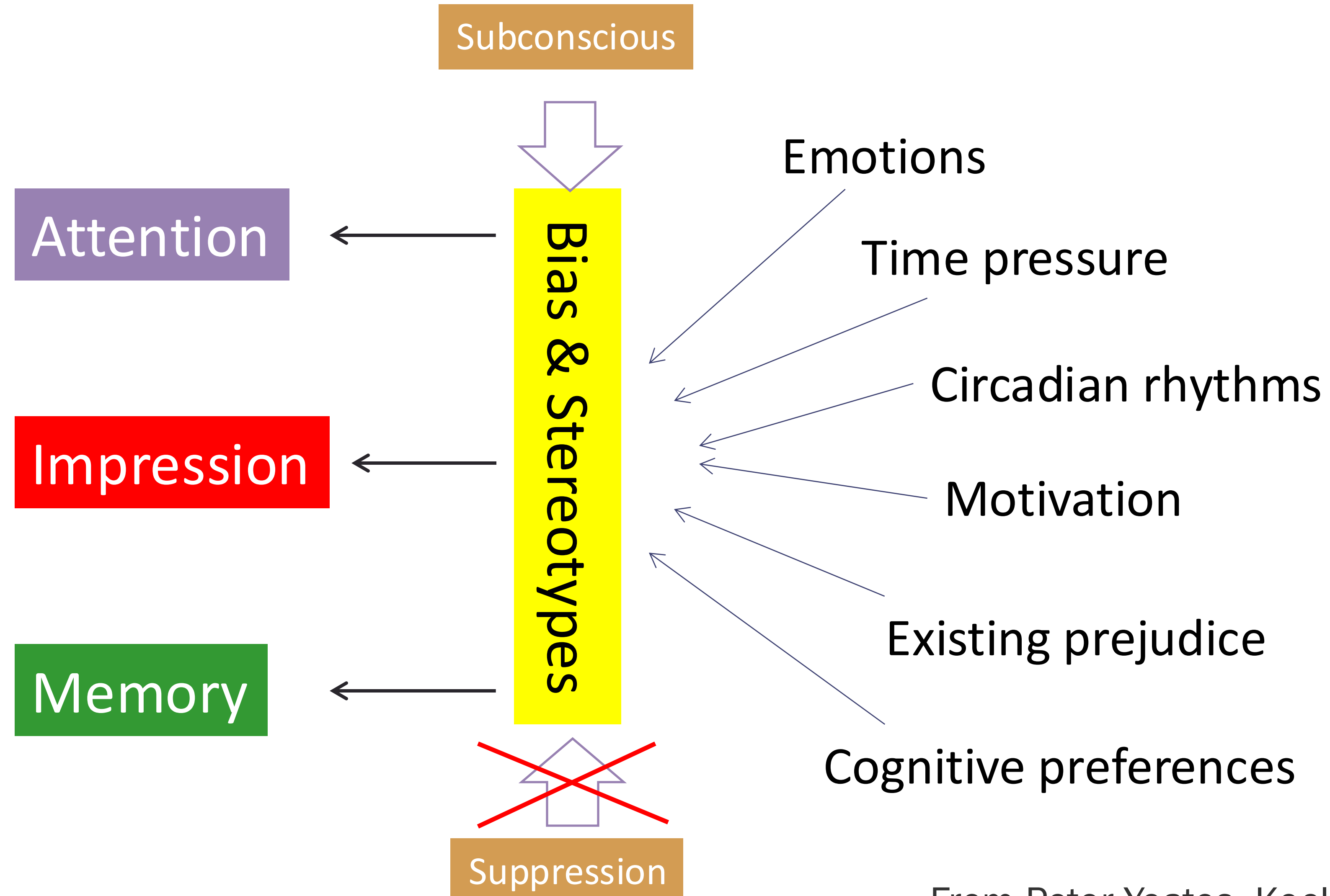
Copyright © 2024 by Intealth. All rights reserved.

Forms of Bias

- Implicit bias:
  - An individual's "prejudicial attitudes towards and stereotypical beliefs about a particular social group or members therein."
    - Often ***subconscious***
    - *Prejudice* relates to the negative <u>attitudes</u> individuals form toward other persons or groups, often in advance or without any actual experience with the affected individuals.
    - *Stereotyping* refers to rigid, fixed, and overgeneralized <u>beliefs</u> about a specific group of people without any actual experience with the affected individuals

Copyright © 2024 by Intealth. All rights reserved.

Bias and Stereotypes



From Peter Yeates. Keele University

Copyright © 2024 by Intealth. All rights reserved.

Reflection Exercise

- Think about the most recent episode of bias you witnessed in assessment.
  - What was the context?
  - What was the impact/consequence of the bias?
  - Do you feel comfortable addressing a bias? (i.e. psychological safety in your environment)

Copyright © 2024 by Intealth. All rights reserved.

Intealth™

Advancing the Global Health Workforce

ECFMG FAIMER

Research on Assessment Bias UME and GME

Copyright © 2024 by Intealth. All rights reserved.

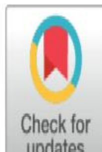
Medical Student Performance Evaluations

RESEARCH ARTICLE

Differences in words used to describe racial and gender groups in Medical Student Performance Evaluations

David A. Ross[1]*, Dowin Boatright[2], Marcella Nunez-Smith[3,4], Ayana Jordan[1], Adam Chekroud[5], Edward Z. Moore[6]

1 Department of Psychiatry, Yale University School of Medicine, New Haven, CT, United States of America, 2 Department of Emergency Medicine, Yale University School of Medicine, New Haven, CT, United States of America, 3 Department of General Internal Medicine, Yale University School of Medicine, New Haven, CT, United States of America, 4 Department of Epidemiology, Yale School of Public Health, New Haven, CT, United States of America, 5 Department of Psychology, Yale University, New Haven, CT, United States of America, 6 Department of Engineering, Central Connecticut State University, New Britain, CT, United States of America

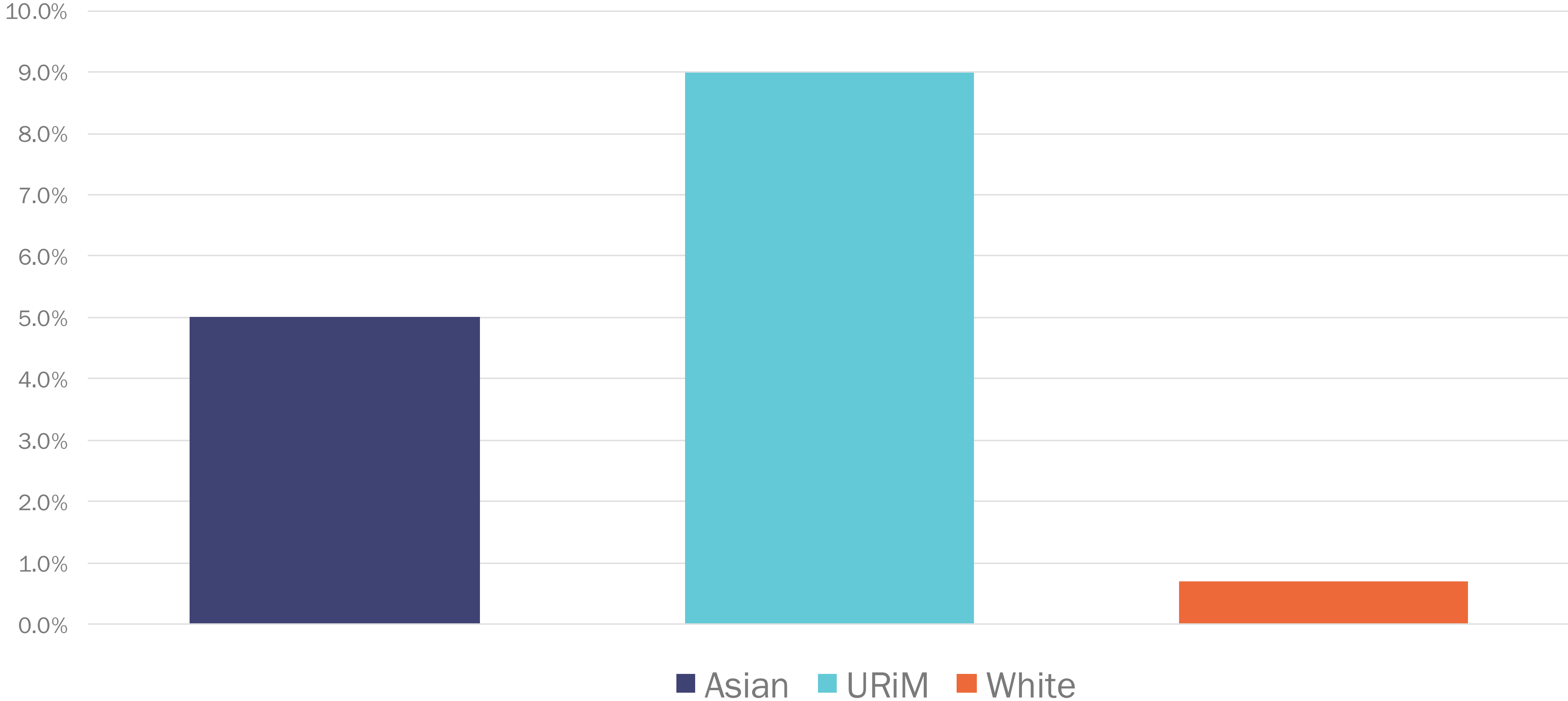


- White applicants more likely to be described using standout keywords (e.g. “exceptional” or “best”) while Black applicants more likely to be described as “competent” in the medical student performance evaluation (MSPE)
- Female applicants were more frequently described as “caring” or “empathetic”

From D. Boatright, NYU

Copyright © 2024 by Intealth. All rights reserved.

Students Reporting Bias in Assessment*



Slide courtesy of D. Boatright

*Hill, JAMA Internal Medicine, 2017

Copyright © 2024 by Intealth. All rights reserved.

The Amplification Cascade

PERSPECTIVES

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

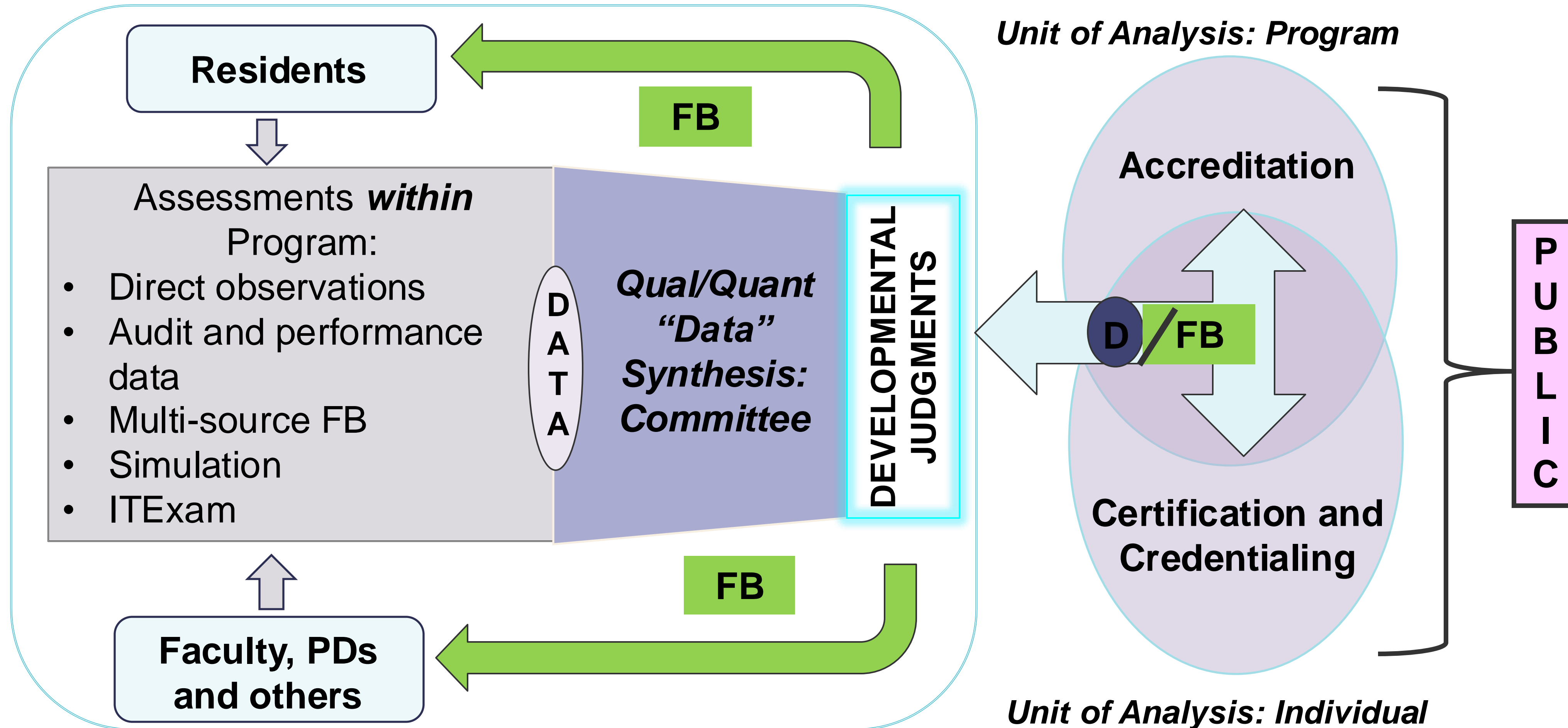
Teherani, Arianne PhD; Hauer, Karen E. MD, PhD; Fernandez, Alicia MD; King, Talmadge E. Jr MD; Lucey, Catherine MD

Academic Medicine 93(9):p 1286-1292, September 2018. | *DOI:* 10.1097/ACM.0000000000002323

Slide courtesy of D. Boatright

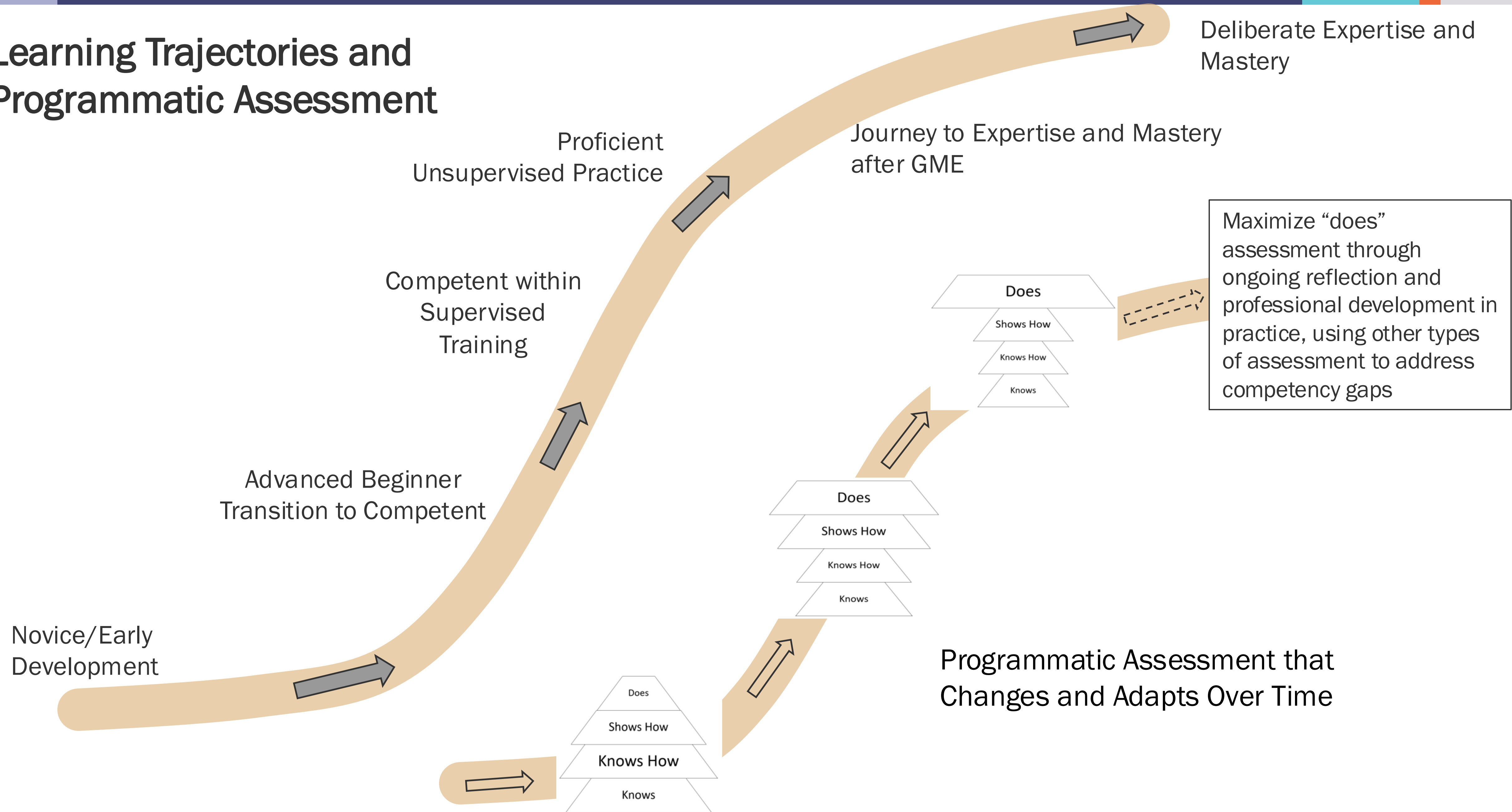
Copyright © 2024 by Intealth. All rights reserved.

The GME Assessment "System"



Copyright © 2024 by Intealth. All rights reserved.

Learning Trajectories and Programmatic Assessment



Copyright © 2024 by Intealth. All rights reserved.

GME Milestones

Practice-Based Learning and Improvement 2: Reflective Practice and Commitment to Personal Growth				
Level 1	Level 2	Level 3	Level 4	Level 5
Demonstrates an openness to performance data (feedback and other input)	Demonstrates an openness to performance data and uses it to develop personal and professional goals	Seeks and accepts performance data for developing personal and professional goals	Using performance data, continually improves and measures the effectiveness of one's personal and professional goals	Acts as a role model for the development of personal and professional goals
	Identifies the factors that contribute to the gap(s) between expectations and actual performance	Analyzes and reflects upon the factors that contribute to gap(s) between expectations and actual performance	Analyzes, reflects on, and institutes behavioral change(s) to narrow the gap(s) between expectations and actual performance	Coaches others on reflective practice

Comments:

Not Yet Completed Level 1

Selecting a response box in the middle of a level implies that milestones in that level and in lower levels have been substantially demonstrated.

Selecting a response box on the line in between levels indicates that milestones in lower levels have been substantially demonstrated as well as **some** milestones in the higher level(s).

Copyright © 2024 by Intealth. All rights reserved.

Comparing MK Milestone Ratings to Past ABIM RAES System

- Gray, et. al. (2024):
  - Knowledge ratings bias against URiM and Asian residents was ameliorated with the adoption of the Milestone ratings system in internal medicine.
  - However, substantial ratings bias against U.S.-born Black residents persisted in IM programs.

- Gray BM, Lipner RS, Roswell RO, Fernandez A, Vandergrift JL, Alsan M. Adoption of Internal Medicine Milestone Ratings and Changes in Bias Against Black, Latino, and Asian Internal Medicine Residents. Ann Intern Med. 2024 Jan;177(1):70-82. doi: 10.7326/M23-1588

Copyright © 2024 by Intealth. All rights reserved.

Comparing MK Milestone Ratings to ABIM RAES System

Figure 2. Yearly bias estimates for residents who are underrepresented in medicine versus U.S.-born non-Latino White residents with no other race/ethnicity.

RAES overall clinical knowledge rating:

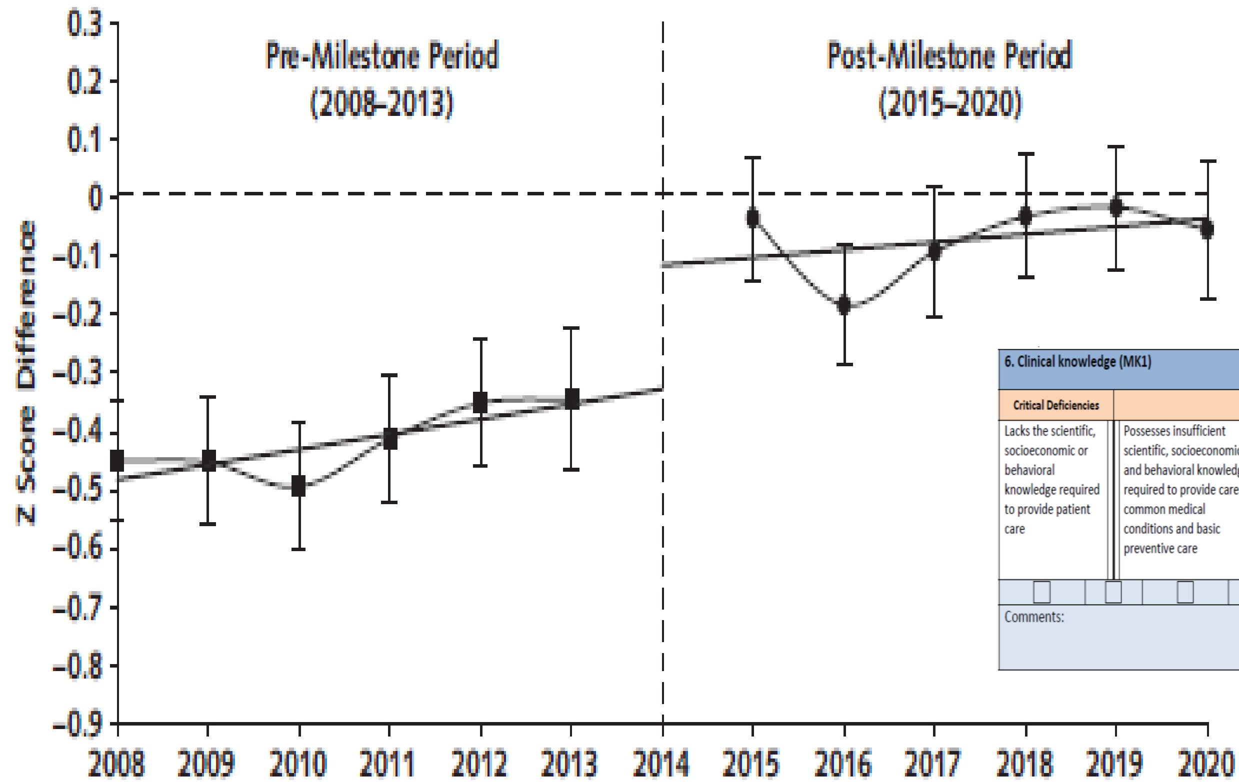
Unsatisfactory (<4)
Unsatisfactory: Limited knowledge of basic and clinical sciences; minimal interest in learning; does not understand complex relationships, mechanisms of disease

Satisfactory

Superior
Exceptional knowledge of basic and clinical sciences; highly resourceful development of knowledge; comprehensive understanding of complex relationships, mechanisms of disease

```
                                    Must receive score =>4 to qualify for exam
         ---Unsatisfactory---    ---Satisfactory---      -----Superior----
Ratings=>[1]      [2]      [3]     [4]      [5]     [6]     [7]      [8]     [9]
```

1-9 Scale: Unsat to Superior

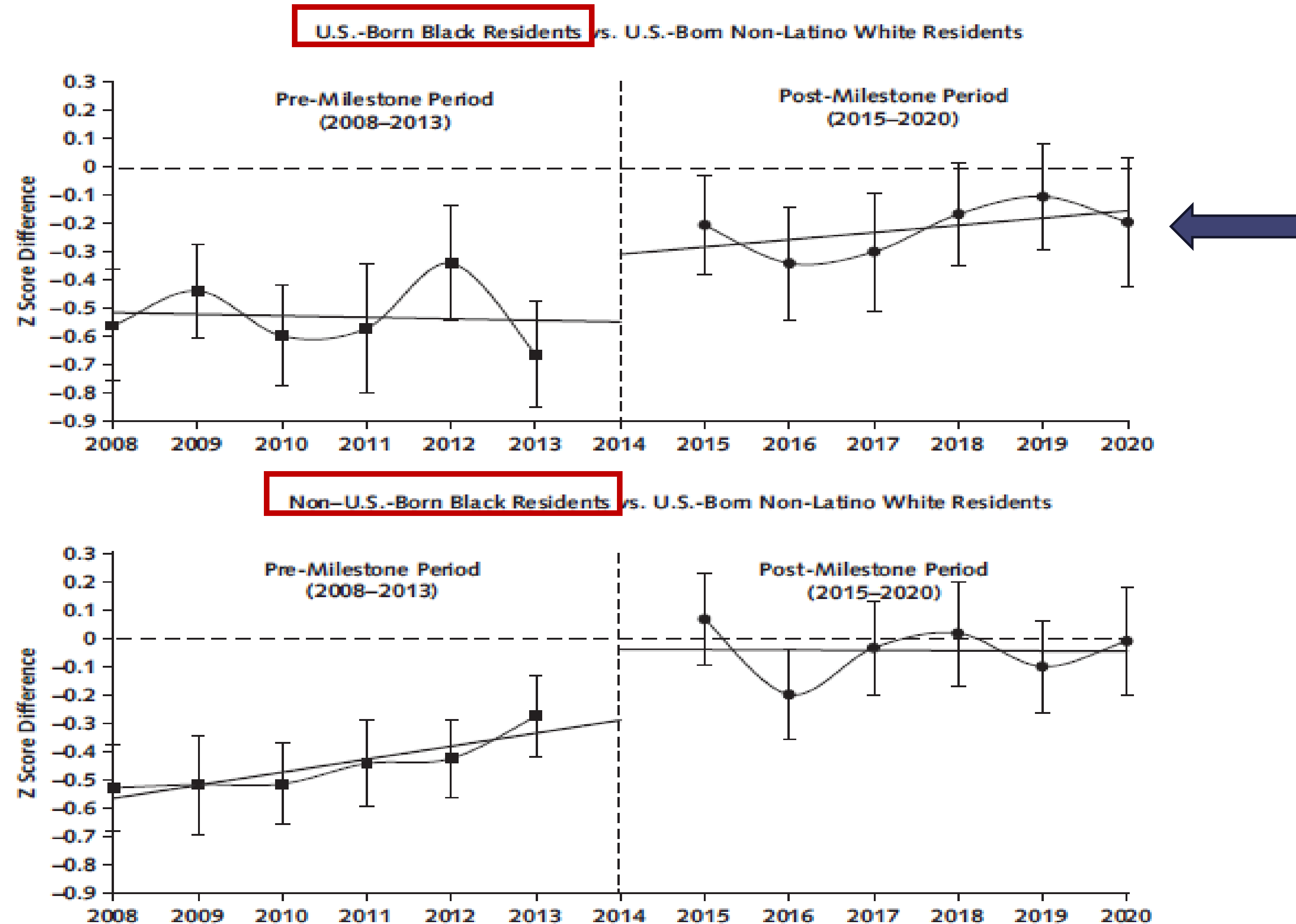


6. Clinical knowledge (MK1)				
Critical Deficiencies			Ready for unsupervised practice	Aspirational
Lacks the scientific, socioeconomic or behavioral knowledge required to provide patient care	Possesses insufficient scientific, socioeconomic and behavioral knowledge required to provide care for common medical conditions and basic preventive care	Possesses the scientific, socioeconomic and behavioral knowledge required to provide care for common medical conditions and basic preventive care	Possesses the scientific, socioeconomic and behavioral knowledge required to provide care for complex medical conditions and comprehensive preventive care	Possesses the scientific, socioeconomic and behavioral knowledge required to successfully diagnose and treat medically uncommon, ambiguous and complex conditions
Comments:				

Copyright © 2024 by Intealth. All rights reserved.

Comparing Milestone Ratings to Past Rating Systems: Black versus White IM Residents

Figure 3. Yearly bias estimates for U.S.-born and non-U.S.-born Black residents with no other race/ethnicity versus U.S.-born non-Latino White residents with no other race/ethnicity.



Copyright © 2024 by Intealth. All rights reserved.

Studies of Gender Bias in Milestone Judgments

- Published:
  - No evidence of bias: internal medicine and ophthalmology
  - Mixed small effects: emergency medicine and general surgery
- Unpublished:
  - No evidence of bias: family medicine and pediatrics

Copyright © 2024 by Intealth. All rights reserved.

Internal Medicine Milestone Ratings by Race/Ethnicity

Interpersonal and Communications Skills

	Difference(SD)	P value
Asian in residentyear 1 Mid	-0.32(0.06)	<0.001
Asian in residentyear 1 End	-0.38(0.06)	<0.001
Asian averaged in year 1	-0.35(0.05)	<0.001
Asian in residentyear 2 Mid	-0.3(0.06)	<0.001
Asian in residentyear 2 End	-0.22(0.05)	<0.001
Asian averaged in year 2	-0.26(0.04)	<0.001
Asian in residentyear 3 Mid	-0.01(0.06)	0.87
Asian in residentyear 3 End	0.01(0.06)	0.88
Asian averaged in year 3	0(0.04)	0.97
URiM in residentyear 1 Mid	-0.16(0.08)	0.05
URiM in residentyear 1 End	-0.34(0.08)	<0.001
URiM averaged in year 1	-0.25(0.06)	<0.001
URiM in residentyear 2 Mid	-0.27(0.08)	0.001
URiM in residentyear 2 End	-0.28(0.08)	<0.001
URiM averaged in year 2	-0.27(0.06)	<0.001
URiM in residentyear 3 Mid	-0.1(0.08)	0.24
URiM in residentyear 3 End	0.04(0.08)	0.61
URiM averaged in year 3	-0.03(0.06)	0.64

Boatright D, Anderson N, Kim JG, Holmboe ES, McDade WA, Fancher T, Gross CP, Chaudhry S, Nguyen M, Nguemeni Tiako MJ, Colson E, Xu Y, Li F, Dziura JD, Saha S. Racial and Ethnic Differences in Internal Medicine Residency Assessments. JAMA Netw Open. 2022 Dec 1;5(12):e2247649.

Copyright © 20[illegible] by [illegible]. All rights reserved.

Intersectionality of Race/Ethnicity and Gender in Emergency Medicine Milestones Ratings*

SEX

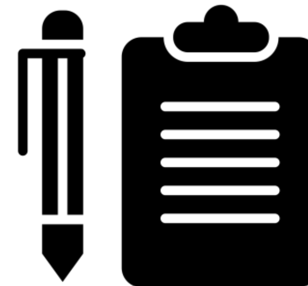
32.1% Female

RACE/ETHNICITY

17.2% Asian
70.1% White
14.3% URiM

EM Programs

128



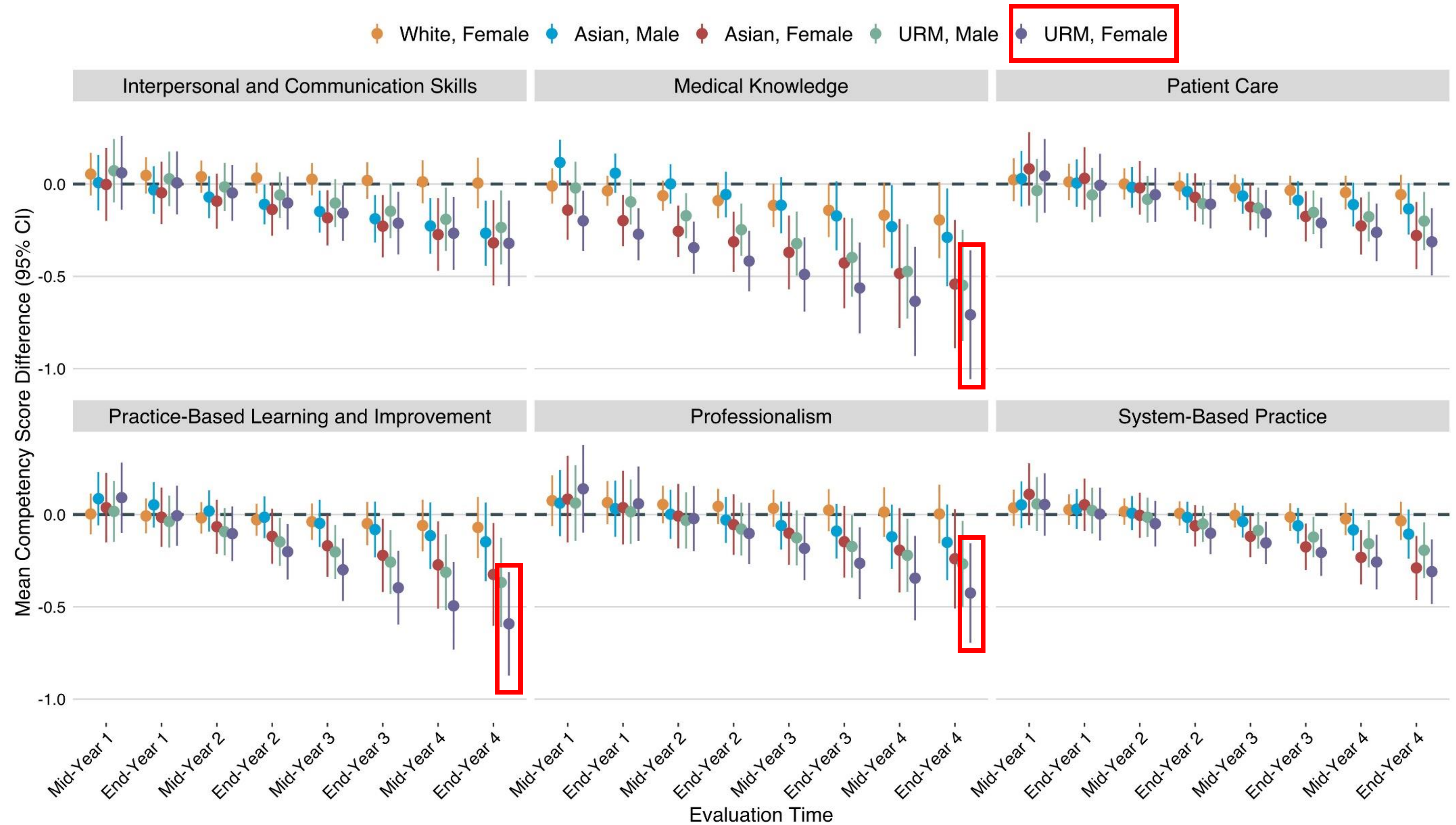
Evaluations

16,248

Copyright © 2024 by Intealth. All rights reserved.

*Lett E, Tran NK, Nweke N, Nguyen M, Kim JG, Holmboe E, McDade W, Boatright D. Intersectional Disparities in Emergency Medicine Residents' Performance Assessments by Race, Ethnicity, and Sex. JAMA Netw Open. 2023 Sep 5;6(9):e2330847

EM: Differences in Assessment by Race, Ethnicity, and Sex*



*Lett E, et. al. Intersectional Disparities in Emergency Medicine Residents' Performance Assessments by Race, Ethnicity, and Sex. JAMA Netw Open. 2023 Sep 5;6(9):e2330847

Copyright © 2024 by Intealth. All rights reserved.

Reflection Exercise

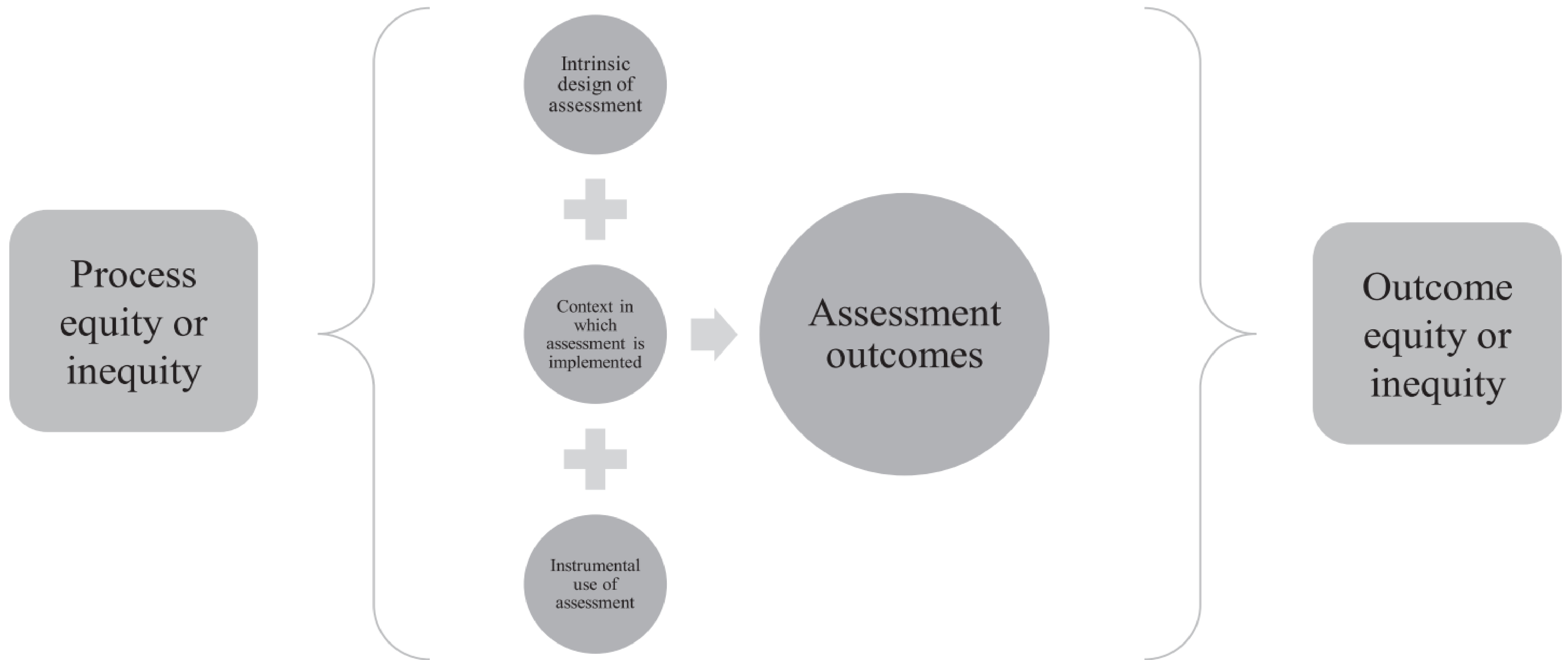
How do you detect and address assessment bias in your own program?

Copyright © 2024 by Intealth. All rights reserved.

Interventions that May Help Mitigate Bias

Copyright © 2024 by Intealth. All rights reserved.

Components of Equity in Assessment



From D. Boatright

Copyright © 2024 by Intealth. All rights reserved.

Lucey, Catherine R. MD; Hauer, Karen E. MD, PhD; Boatright, Dowin MD; Fernandez, Alicia MD. Medical Education's Wicked Problem: Achieving Equity in Assessment for Medical Learners. Academic Medicine 95(12S):p S98-S108, December 2020.

Strategy	Description	Assessment Example
Stereotype replacement	Recognizing when a stereotype has been activated, thinking about why, and then actively substituting non- stereotypical thoughts	When completing a narrative assessment of a female learner, the assessor stops to consider if they may be using gender-laden language or uses an online tool to assess for gender bias. If bias is found, the assessor substitutes evidence-based behavioral skills that are more neutral.
Perspective taking	Considering what it would be like to be a member of the minoritized group	During rounds faculty witness a difficult interaction between a learner from a URiM group with a discriminatory patient. Faculty should ask themselves: What must that be like for the learner? How will I intervene in this situation?
Individuation	Recognizing when you have stereotyped someone according to their group affiliation and instead thinking about what makes them an individual	A faculty member watches a learner from another country struggle to interview a patient with a possible sexually transmitted disease and initially stereotypes the learner as from a group "uncomfortable talking about sex." Instead, the faculty sees an individual learner struggling and seeks to understand why they are struggling as an individual.

Copyright © 2024 by Intealth. All rights reserved.

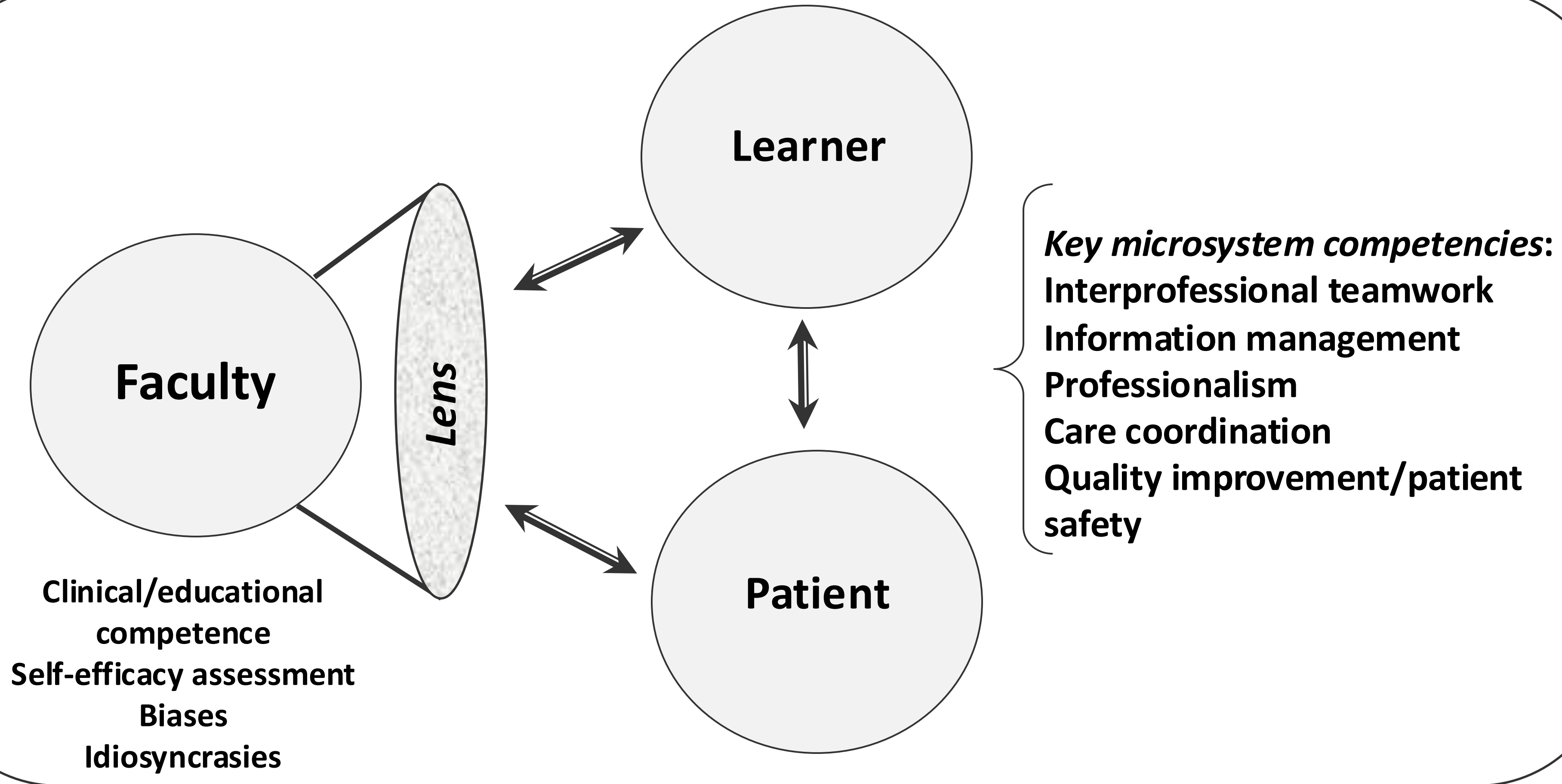
Holmboe, Eric S. MD; Osman, Nora Y. MD; Murphy, Christina M.; Kogan, Jennifer R. MD. The Urgency of Now: Rethinking and Improving Assessment Practices in Medical Education Programs. Academic Medicine 98(8S):p S37-S49, August 2023.

Strategy	Description	Assessment Example
Counter-stereotypic imaging	Imagining an individual or situation that counteracts a stereotypical reaction in detail	A faculty member starts with an assumption that women are not strong enough to perform orthopedic procedures and then instead thinks about successful women who are orthopedic surgeons.
Increased opportunities for contact	Increasing opportunities for contact with members of a stereotyped group	Programs and faculty can spend meaningful time with URiM trainees to listen and learn more about their lived experiences and their path to the current training program.

Holmboe, Eric S. MD; Osman, Nora Y. MD; Murphy, Christina M.; Kogan, Jennifer R. MD. The Urgency of Now: Rethinking and Improving Assessment Practices in Medical Education Programs. Academic Medicine 98(8S):p S37-S49, August 2023.

Copyright © 2024 by Intealth. All rights reserved.

Assessment: Complex and Situated in Context

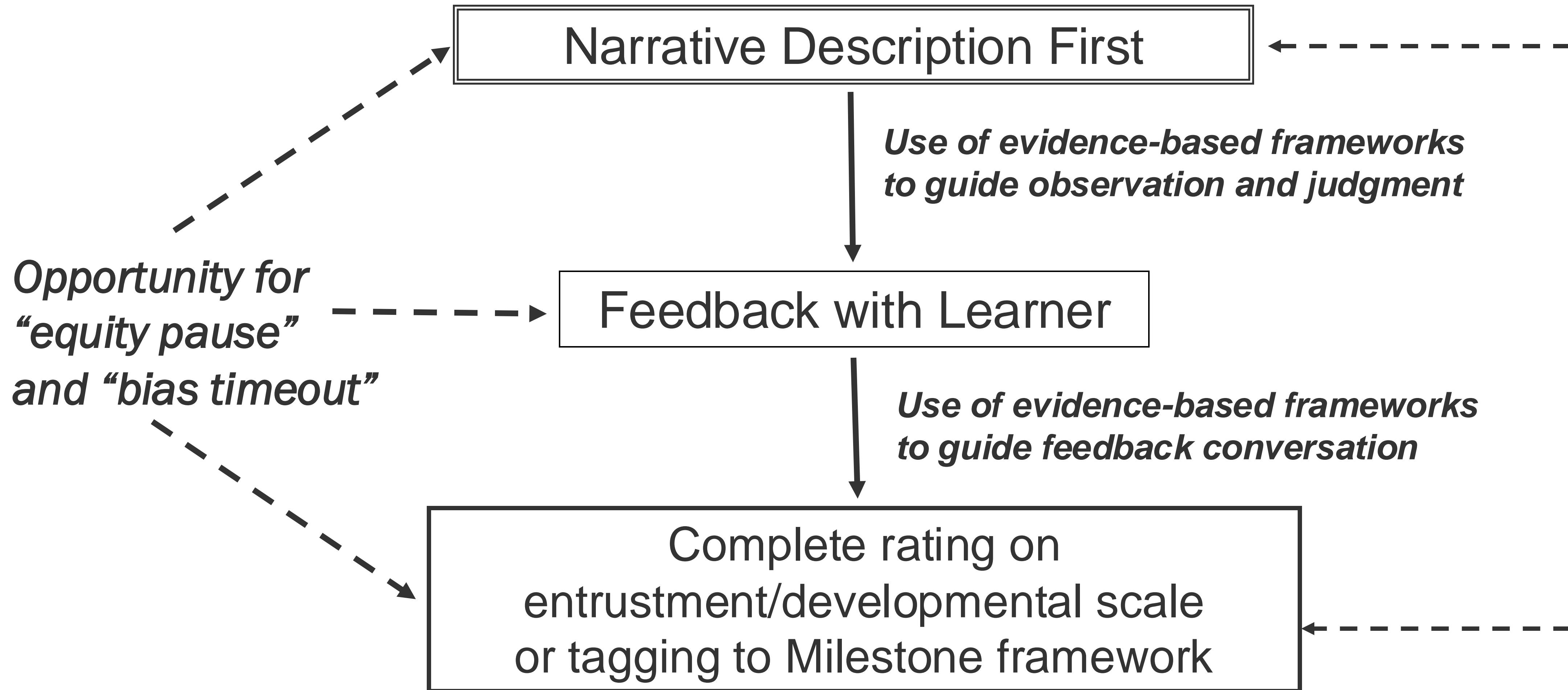


Microsystems: Clinic, Hospital Ward, Operating Room

Institution and the Clinical Learning Environment

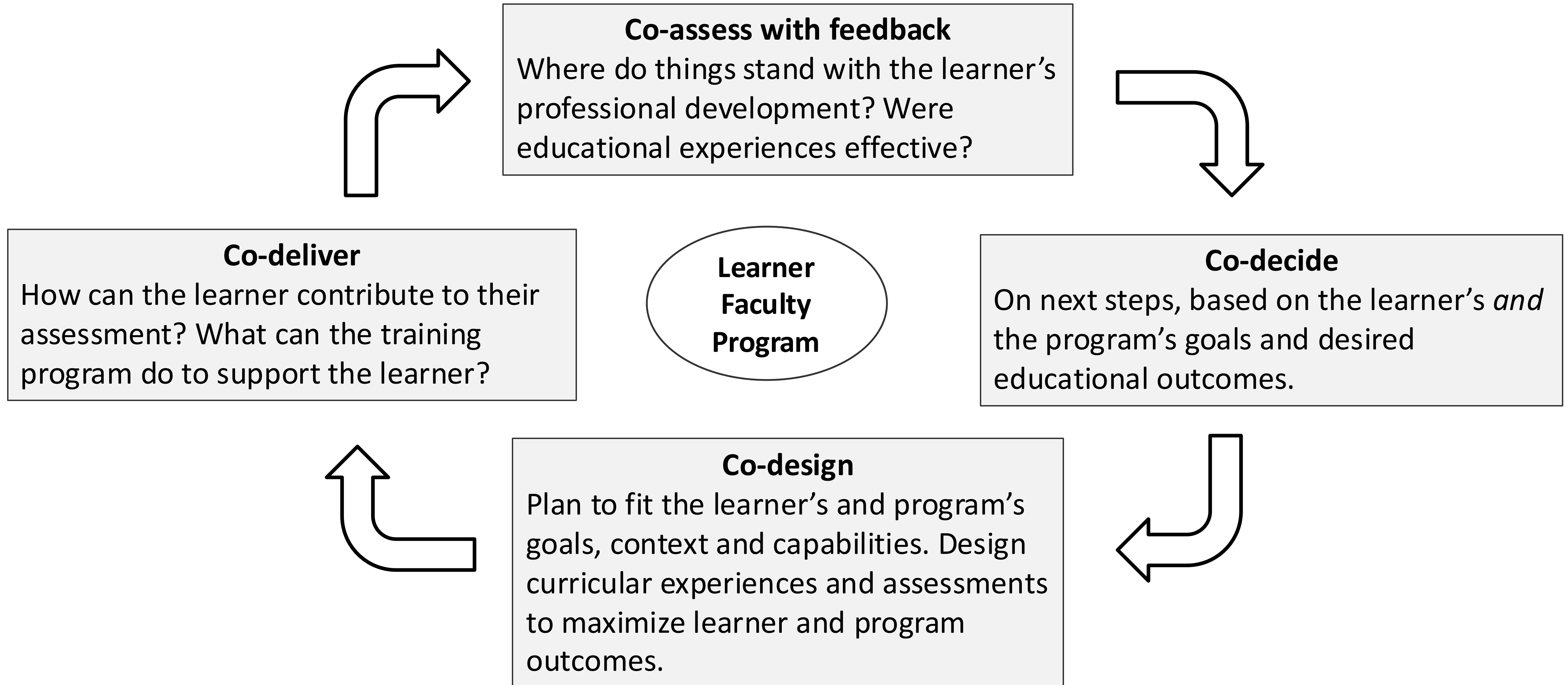
Copyright © 2024 by Intealth. All rights reserved.

Re-thinking the Assessment Process

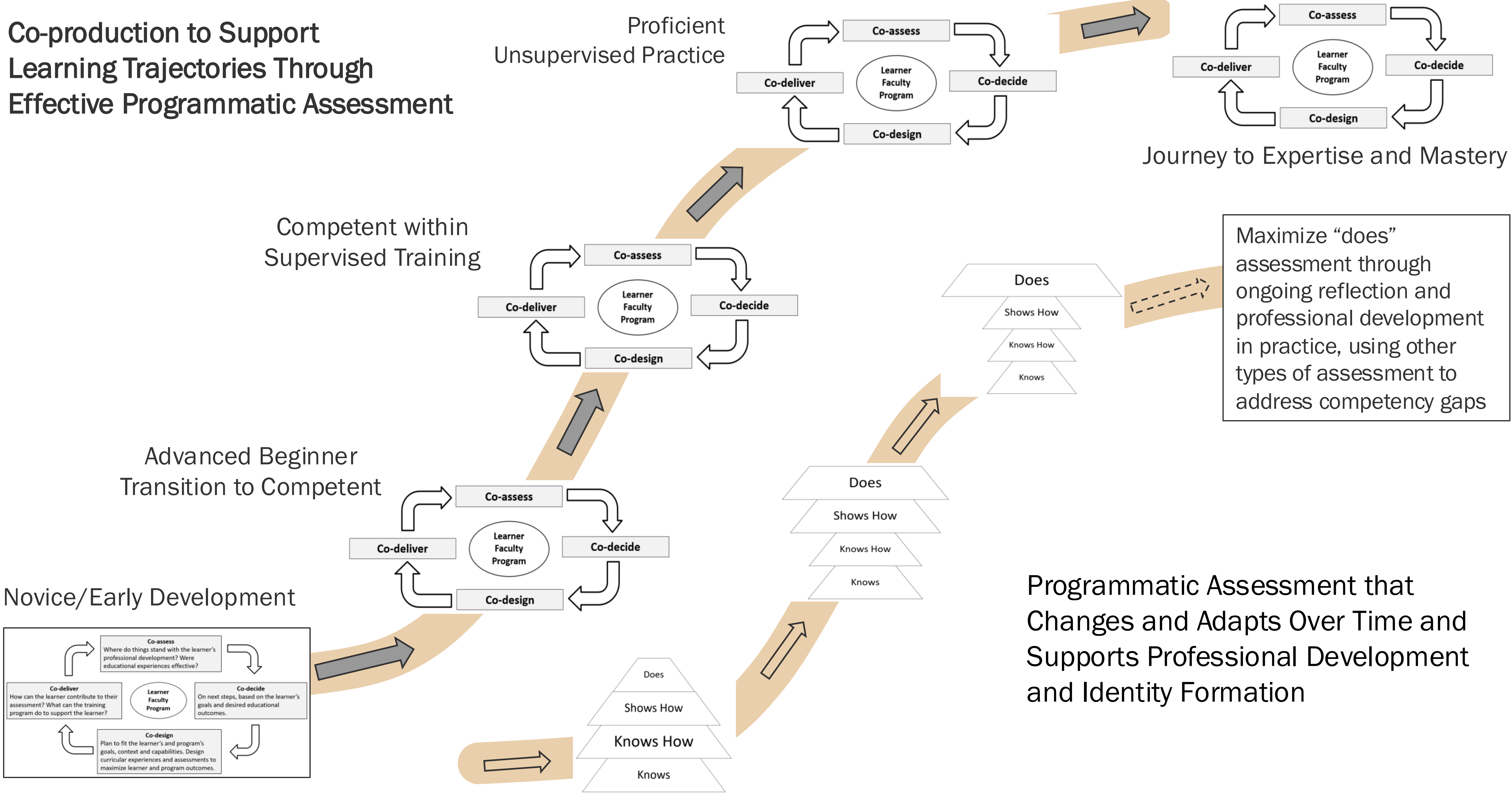


Copyright © 2024 by Intealth. All rights reserved.

Co-production Learning Cycle



Copyright © 2024 by Intealth. All rights reserved.



Copyright © 2024 by Intealth. All rights reserved.

A (Partial) List of Actions Medical Education Programs Can Do Now to Reduce Bias in Assessment

- Use assessment data to investigate, understand, and address sources of bias in the assessment program.
  - Seek to understand the effects of program culture and the institutional learning environment on learners' professional development.
- Develop and implement programmatic assessment.
  - Experiment with bias-reducing interventions to improve assessments.

Copyright © 2024 by Intealth. All rights reserved.

A (Partial) List of Actions Medical Education Programs Can Do Now to Reduce Bias in Assessment

- **Investigate, understand, and address sources of unwarranted variation in the assessment program.**
  - **Use psychometrics to identify rater “harshness” and possible bias**
- **Leverage coproduction to support assessment practices and learners’ professional development.**
- **Honestly assess and confront inertia in changing assessment practices.**

Copyright © 2024 by Intealth. All rights reserved.

Reflection Exercise

- What is one thing you could do right now in your program to reduce bias(es) in assessment?

Copyright © 2024 by Intealth. All rights reserved.

Conclusions

- Bias in assessment remains a persistent and pernicious problem but…
  - We have tools and approaches that can mitigate bias
  - All programs should build into the assessment programs continuous monitoring and improvement practices

Contact: eholmboe@intealth.org

Copyright © 2024 by Intealth. All rights reserved.