



# HHS Public Access

Author manuscript

*Acad Med.* Author manuscript; available in PMC 2023 September 01.

Published in final edited form as:

*Acad Med.* 2022 September 01; 97(9): 1351–1359. doi:10.1097/ACM.0000000000004743.

## Association Between Resident Race and Ethnicity and Clinical Performance Assessment Scores in Graduate Medical Education

**Robin Klein, MD, MEHP [associate professor],**

Department of Medicine, Division of General Internal Medicine, Emory University School of Medicine, Atlanta, Georgia.

**Nneka N. Ufere, MD, MSCE,**

Department of Medicine, Division of Gastroenterology, Massachusetts General Hospital, Boston, Massachusetts.

**Sarah Schaeffer, MD [associate professor],**

Department of Medicine, Division of Hospital Medicine, University of California, San Francisco, San Francisco, California.

**Katherine A. Julian, MD [professor],**

Department of Medicine, Division of General Internal Medicine, University of California, San Francisco, San Francisco, California.

**Sowmya R. Rao, PhD [statistician],**

Department of Global Health, Boston University School of Public Health and Massachusetts General Hospital Biostatistics Center, Boston, Massachusetts.

**Jennifer Koch, MD [professor],**

Department of Medicine, University of Louisville, Louisville, Kentucky.

**Anna Volerman, MD [associate professor],**

Departments of Medicine and Pediatrics, University of Chicago, Chicago, Illinois.

**Erin D. Snyder, MD [professor],**

Department of Medicine, Division of General Internal Medicine, University of Alabama Birmingham School of Medicine, Birmingham, Alabama.

**Vanessa Thompson, MD [associate professor],**

Department of Medicine, Division of General Internal Medicine, University of California, San Francisco, San Francisco, California.

**Ishani Ganguli, MD, MPH [associate professor],**

---

Correspondence should be addressed to Robin Klein, Department of Medicine, Emory University School of Medicine, 49 Jesse Hill Jr. Dr. SE, Atlanta, GA 30303; telephone: (404) 778-1645; rklein3@emory.edu.

*Other disclosures:* None reported.

*Ethical approval:* The institutional review boards at each of the participating institutions reviewed and deemed the study protocol exempt.

Supplemental digital content for this article is available at <http://links.lww.com/ACADMED/B273>.

*Previous presentations:* This work was presented at the Association of American Medical Colleges Group on Educational Affairs virtual meeting in March 2021.

Department of Medicine, Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital, Boston, Massachusetts.

**Sherri-Ann M. Burnett-Bowie, MD, MPH [associate professor],**

Department of Medicine, Division of Endocrinology, Massachusetts General Hospital, Boston, Massachusetts.

**Kerri Palamara, MD [associate professor]**

Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts.

## Abstract

**Purpose**—To assess the association between internal medicine (IM) residents' race/ethnicity and clinical performance assessments.

**Method**—The authors conducted a cross-sectional analysis of clinical performance assessment scores at 6 U.S. IM residency programs from 2016 to 2017. Residents underrepresented in medicine (URiM) were identified using self-reported race/ethnicity. Standardized scores were calculated for Accreditation Council for Graduate Medical Education core competencies. Cross-classified mixed-effects regression assessed the association between race/ethnicity and competency scores, adjusting for rotation time of year and setting; resident gender, postgraduate year, and IM In-Training Examination percentile rank; and faculty gender, rank, and specialty.

**Results**—Data included 3,600 evaluations by 605 faculty of 703 residents, including 94 (13.4%) URiM residents. Resident race/ethnicity was associated with competency scores, with lower scores for URiM residents (difference in adjusted standardized scores between URiM and non-URiM residents, mean [standard error]) in medical knowledge (−0.123 [0.05],  $P = .021$ ), systems-based practice (−0.179 [0.05],  $P = .005$ ), practice-based learning and improvement (−0.112 [0.05],  $P = .032$ ), professionalism (−0.116 [0.06],  $P = .036$ ), and interpersonal and communication skills (−0.113 [0.06],  $P = .044$ ). Translating this to a 1 to 5 scale in 0.5 increments, URiM resident ratings were 0.07 to 0.12 points lower than non-URiM resident ratings in these 5 competencies. The interaction with faculty gender was notable in professionalism (difference between URiM and non-URiM for men faculty −0.199 [0.06] vs women faculty −0.014 [0.07],  $P = .01$ ) with men more than women faculty rating URiM residents lower than non-URiM residents. Using the 1 to 5 scale, men faculty rated URiM residents 0.13 points lower than non-URiM residents in professionalism.

**Conclusions**—Resident race/ethnicity was associated with assessment scores to the disadvantage of URiM residents. This may reflect bias in faculty assessment, effects of a noninclusive learning environment, or structural inequities in assessment.

While equity is a core professional value in medical education, disparities due to race and ethnicity persist.<sup>1,2</sup> Black, Hispanic/Latinx, and Native American physicians remain underrepresented in medicine (URiM) relative to the general population and experience significant disparities in career advancement, pay equity, and leadership achievement.<sup>3–5</sup> Furthermore, multiple aspects of one's social identity, such as race/ethnicity and gender, may give rise to interrelated systems of inequity.<sup>6</sup>

Inequities (i.e., differences linked to unfairness in a system) related to race/ethnicity in medical education are concerning. The literature suggests that URiM students and trainees

experience microaggressions, racial bias, and discrimination during education and training.<sup>7-11</sup> In a national survey of graduating medical students in the United States, nearly a quarter of URiM students (23.3%) reported discrimination based on race/ethnicity, including receiving lower grades or evaluations (9.6%).<sup>7</sup>

Disparities associated with race/ethnicity in the assessment of learners in medical education is of particular concern. Evidence suggests differences in assessment exist in undergraduate medical education in the United States, with URiM students receiving lower clerkship grades and differences in language used in student performance assessments.<sup>12-18</sup> Studies from the United Kingdom and the Netherlands have also reported differences in medical school grades associated with race/ethnicity.<sup>19,20</sup> Qualitative studies exploring the experiences of URiM learners in the United States and United Kingdom suggest that URiM learners view workplace-based clinical assessments as vulnerable to bias.<sup>21,22</sup>

Studies exploring race/ethnicity and assessment in medical education are often hampered by differences in assessments across programs and low representation of URiM learners. Much of the literature to date is in undergraduate medical education and limited to single-institution settings or institution-specific assessments.<sup>13-16</sup> Literature examining the impact of resident race/ethnicity on assessment in graduate medical education is lacking.

Identifying and addressing disparities in learner assessment are important. Evidence suggests that small differences in performance assessment metrics can give rise to larger disparities in later outcomes.<sup>23</sup> In medical education, assessments inform decisions about time-in-training, selection for honors or need for remediation, and access to and caliber of postgraduate training opportunities, including in highly competitive fields within medicine, many of which have long-standing gaps in the representation of URiM physicians.<sup>24-27</sup>

This study explores the relationship between resident race/ethnicity and Accreditation Council for Graduate Medical Education (ACGME) core competency scores as employed in graduate medical education assessment.

## Method

### Study design and population

We conducted a retrospective, cross-sectional analysis of resident assessment scores at 6 ACGME accredited internal medicine residency training programs in the United States: Emory University, Massachusetts General Hospital, University of Alabama Birmingham, University of California San Francisco, University of Chicago, and University of Louisville.

In the United States, medical education and training involves 4 years of medical school followed by 3 or more years of residency training, depending on the field. Internal medicine residency training involves clinical rotations during which trainees provide care for patients in a variety of clinical settings under the supervision of teaching faculty. Faculty assess learner performance during these clinical rotations.

Graduate medical education in the United States uses a competency-based medical education framework, which focuses on learners' progression of competence using explicit

outcome goals. Assessment in this framework focuses on learners demonstrating skills and knowledge and meeting progressive developmental markers to support their progression from novice to mastery.<sup>28,29</sup> Accredited U.S. residency training programs use the ACGME's competency-based assessment framework, which includes 6 core competencies, each of which is composed of multiple subcompetencies or milestones.<sup>30</sup>

We focused on clinical performance assessments of internal medicine residents by faculty from inpatient general medicine rotations during the 2016–2017 academic year. At each program, inpatient general medicine teams include a postgraduate year (PGY) 2 or 3 resident leading a team of PGY-1 interns and medical students in providing patient care under the supervision of 1 to 2 teaching faculty. Residents participate in multiple inpatient medicine rotations each year, spending 2 to 4 weeks on each of these rotations. Faculty evaluate each resident under their supervision, and these clinical performance assessment data are routinely collected by training programs.

### Data and data collection

We collected assessment metrics data and assessment characteristics including rotation setting and time of year; resident characteristics including self-reported race/ethnicity, gender, PGY, and baseline Internal Medicine In-Training Examination (IM-ITE) percentile rank; and faculty characteristics including gender, specialty, academic rank, and residency educational role where applicable. Resident and faculty gender was determined by participants' professional gender identity using institutional data. Resident race/ethnicity was determined by residents' self-reported race/ethnicity information obtained from residency applications.<sup>31</sup>

We used the Association of American Medical Colleges definition of URiM as those who are underrepresented in medicine relative to national and local demographics.<sup>32</sup> This includes those who identify as African American and/or Black, Hispanic/Latinx, Native American (American Indian, Alaska Native, and Native Hawaiian) and Pacific Islander and locally underrepresented racial/ethnic groups as defined by each program in our study (see Supplemental Digital Appendix 1 at <http://links.lww.com/ACADMED/B273>). We included as URiM those residents who identified with 2 or more races/ethnicities where at least 1 race/ethnicity was underrepresented. All other residents were categorized as not underrepresented in medicine (non-URiM). Non-URiM residents were further divided into those identifying as White (non-URiM, White) and those identifying as non-White (non-URiM, non-White).<sup>13</sup> This approach allowed us to explore assessment patterns of URiM residents and non-White residents identifying with races/ethnicities not underrepresented in medicine. We did not obtain faculty race/ethnicity.

Assessment metrics data included faculty assessments of residents' performance in the ACGME's core competencies (Patient Care [PC], Medical Knowledge [MK], Systems-Based Practice [SBP], Practice-Based Learning and Improvement [PBLI], Professionalism [PROF], and Interpersonal and Communication Skills [ICS]) and their respective internal medicine-specific milestones.<sup>30</sup>

Each program in our study used a unique evaluation tool to assess resident performance. We employed an approach used in prior work using this same cohort to contend with differences across evaluation tools.<sup>33</sup> We masked and independently matched question stems to the appropriate competency. To account for differences in rating scales, we converted rating scores to a standardized score.<sup>34</sup> Similar to a z score, the standardized score is a measure of how many standard deviations a data point is from the mean and was calculated from the following formula:  $\text{standardized score} = (\text{raw score} - \text{mean score}) / \text{standard deviation}$ , where raw score was the score in a specific competency obtained from the resident's evaluation and the mean score and standard deviation were the mean and standard deviation of the scores in a specific competency for all resident evaluations in the participant's program.<sup>34</sup> Competency scores were computed as the arithmetic mean of the relevant subcompetency scores. We calculated standardized scores for each competency at each program based on the rating distribution at that program. Standardized scores were used in our analysis and are expressed as standard deviations from the mean.<sup>34,35</sup>

Assessment and demographic data were extracted from education management systems at each program. Then, research team members from that program deidentified the evaluation data, including removing faculty and resident names. We used the deidentified data in aggregate for our analysis.

## Data analysis

We evaluated the relationship between resident race/ethnicity and standardized core competency scores with a multivariable, random-intercept, generalized linear mixed-effects regression model with crossed random effects where evaluation ratings/scores were cross-classified with both resident and faculty within training programs.<sup>36</sup>

First, we examined the distribution of resident and faculty characteristics across our data and assessed differences using chi-square tests. For each competency, we then fit separate unadjusted and adjusted regression models to assess a main effect for race/ethnicity. Adjusted models were conditioned for the following covariates: resident gender, baseline IM-ITE percentile rank, and PGY; faculty gender, academic rank (professor, associate professor, assistant professor/instructor/chief resident, no rank/clinical associate), and specialty (general medicine, hospital medicine, subspecialty); and rotation setting (university, Veterans Administration, public or community hospital) and time of year (July–September, October–December, January–March, April–June). We included all available covariates we believed to be conceptually important. After testing for the main effect of race/ethnicity, we then fitted models to explore the interaction of race/ethnicity and gender. We derived mean adjusted standardized core competency scores and *P* values for the difference in least square means from our model. Finally, we tested to ensure our model met assumptions of regression including linearity, normality, and homogeneity of variance. Model description and fit measures are included in Supplemental Digital Appendix 2 at <http://links.lww.com/ACADMED/B273>.

Following this analysis, we converted the standardized scores to a scale used frequently in assessment to allow for more intuitive interpretation.<sup>34,35</sup> To do this, we established a

representative scale (rating of 1 to 5 in 0.5 increments) using data from 3 programs in our study that used this scale. We calculated the distribution of ratings in the 3 programs that used this scale, including mean and standard deviation. Rescaling was done by multiplying the standardized score from our analysis by the standard deviation of this distribution.<sup>35</sup>

We present our outcomes in 2 ways: (1) mean adjusted standardized core competency scores and associated standard errors (SE) from our final model and (2) mean core competency scores using the representative scale of 1 to 5 in 0.5 increments.

We present *P* values unadjusted for multiple comparisons given the exploratory nature of our study and that we prioritized not missing true differences associated with race/ethnicity.<sup>37</sup> Analyses were conducted in SAS 9.4 (SAS Institute, Cary, North Carolina).

We present differences in scores between groups in the 6 core competencies. Data are presented in aggregate to ensure anonymity of participants and programs. The institutional review boards at each of the participating institutions reviewed and deemed the study protocol exempt. Funding sources were not involved in the study design, data analysis and interpretation, manuscript preparation, or the decision to approve publication of the manuscript.

## Results

Table 1 details the characteristics of the resident and faculty participants. Data included 3,600 evaluations by 605 faculty of 703 residents. Of faculty, 318 (52.6%) were men and 287 (47.4%) were women; 387 (55.0%) residents were men and 316 (45.0%) were women. Among residents, 94 (13.4%) identified with racial/ethnic groups that are underrepresented in medicine, and 609 (86.6%) identified with groups not underrepresented in medicine. In turn, this included 365 (51.9%) non-URiM, White residents and 244 (34.7%) non-URiM, non-White residents.

URiM and non-URiM residents had similar distributions of resident gender, faculty gender, and PGY. There was a difference in baseline IM-ITE percentile rank between URiM and non-URiM residents (median IM-ITE percentile rank for URiM 60.0 vs non-URiM 73.0, *P* < .001).

Using the representative scale of 1 to 5 in 0.5 increments, the mean score (standard deviation) in each competency was: PC 3.518 (0.658), MK 3.599 (0.645), SBP 3.706 (0.640), PBLI 3.721 (0.605), PROF 3.819 (0.673), and ICS 3.749 (0.672).

### Influence of resident race/ethnicity

Resident race/ethnicity was associated with competency scores, with lower scores for URiM residents compared with non-URiM residents (see Figure 1 and Supplemental Digital Appendix 3 at <http://links.lww.com/ACADMED/B273>). This included scores (difference in adjusted standardized scores between URiM and non-URiM residents, mean [SE]) for MK (-0.123 [0.05], *P* = .021), SBP (-0.179 [0.05], *P* = .005), PBLI (-0.112 [0.05], *P* = .032), PROF (-0.116 [0.06], *P* = .036), and ICS (-0.113 [0.06], *P* = .044). Using the 1 to 5 scale,



the ratings of URiM residents were 0.07 to 0.12 points lower than the ratings of non-URiM residents in these 5 competencies.

Scores for URiM residents were lower than scores for both non-URiM, non-White and non-URiM, White residents (see Figure 2 and Supplemental Digital Appendix 4 at <http://links.lww.com/ACADMED/B273>). This included scores (adjusted standardized scores, mean [SE]) for MK (URiM 0.050 [0.07] vs non-URiM, non-White 0.210 [0.06] vs non-URiM, White residents 0.146 [0.05],  $P = .02$ ), SBP ( $-0.135$  [0.07] vs  $0.073$  [0.06] vs  $0.024$  [0.05],  $P = .001$ ), and PBLI ( $-0.0004$  [0.07] vs  $0.141$  [0.06] vs  $0.092$  [0.06],  $P = .04$ ). Scores for URiM residents were lower than those for non-URiM, non-White residents in 4 of 6 competencies (difference in adjusted standardized scores between URiM and non-URiM, non-White residents, mean [SE]), including MK ( $-0.160$  [0.06],  $P < .01$ ), SBP ( $-0.208$  [0.06],  $P < .001$ ), PBLI ( $-0.142$  [0.06],  $P = .013$ ), and PROF ( $-0.146$  [0.06],  $P = .02$ ). Using the 1 to 5 scale, ratings of URiM residents were 0.10 points lower in MK, 0.13 points lower in SBP, 0.09 points lower in PBLI, and 0.10 points lower in PROF compared with ratings of non-URiM, non-White residents.

### Influence of gender

The interaction of resident race/ethnicity with resident gender or PGY was not significant in any of the 6 core competencies (see Supplemental Digital Appendix 5 at <http://links.lww.com/ACADMED/B273>). There was an interaction between resident race/ethnicity and faculty gender in the PROF competency (see Figure 3 and Supplemental Digital Appendix 6 at <http://links.lww.com/ACADMED/B273>). In the PROF competency, men faculty rated non-URiM residents higher than they did URiM residents, and there was a greater difference in scores between URiM residents and non-URiM residents with men faculty compared with women faculty (difference in adjusted standardized scores between URiM and non-URiM residents, mean [SE], for men faculty  $-0.199$  [0.06] vs women faculty  $-0.014$  [0.07],  $P = .013$ ). Using the 1 to 5 rating scale, men faculty rated URiM residents 0.13 points lower in PROF than non-URiM residents, whereas women faculty rated URiM residents 0.01 points higher than non-URiM residents in PROF.

### Discussion

In this multisite study exploring the association between resident race/ethnicity and clinical performance assessment scores in graduate medical education, we found that resident race/ethnicity was a factor in assessment, with URiM residents receiving lower scores compared with non-URiM residents. The overall difference in competency scores between URiM and non-URiM residents was small.

While comparable data of disparities by race/ethnicity in graduate medical education are lacking, an overall small but significant effect was similarly seen in a study comparing scores by resident gender using this same cohort.<sup>33</sup> Our findings align with studies of assessments in undergraduate medical education, which found that race/ethnicity was negatively associated with URiM student assessment, including clerkship grades and narrative comments on clerkship evaluations.<sup>13,16</sup> In addition, a study of Medical Student Performance Evaluations found differences in the descriptive language used, such that Black

students were more likely to be described as “competent” and White students more likely to be described using “standout” descriptors.<sup>12</sup>

In our study, scores for URiM residents were lower than those for both White and non-URiM, non-White residents. Other studies have reported differences in clerkship grades and honor society membership favoring White medical students over both underrepresented and not underrepresented minority students, even after adjusting for United States Medical Licensing Examination Step 1 scores.<sup>12,13</sup>

We found that resident gender was not significantly associated with differences in assessment scores between URiM and non-URiM residents. There has been limited study of the intersecting effects of race/ethnicity and gender in medical education.<sup>7,16</sup> Evidence, including prior work in this same cohort, shows significant gender-based differences in assessment metrics linked to time in training or PGY.<sup>33,38,39</sup> The small number of URiM residents in our current study may have limited our ability to discern differences across multiple variables. Simply put, our findings may reflect an inability to detect a difference rather than the absence of a difference. Further research is needed to explore the interaction of resident race/ethnicity and gender in assessment metrics.

Given our findings, we must consider the potential sources of these differences in assessment scores. These differences may reflect a combination of factors, including the cumulative effects of a noninclusive learning environment on trainees, racial bias (conscious or unconscious) in faculty assessment of URiM residents, and structural inequities in assessment measures. We explore each of these in more detail below.

### **Inequities in experiences with the learning environment**

The differences we observed in assessment scores by resident race/ethnicity may reflect inequities in trainees’ experience within the learning environment. Evidence suggests that URiM residents regularly experience microaggressions and bias during training and these experiences present challenges in their professional role.<sup>9,10,40</sup> A recent study of resident physicians noted that Black, Latinx, and Asian residents reported more frequent experiences with biased behavior and a sense of futility in responding to these episodes.<sup>11</sup> Negative experiences from microaggressions to overt racism can trigger heightened stress and physiological arousal, which affect behavior and working memory capacity.<sup>41</sup> This may impact residents’ ability to effectively demonstrate competency in performance-oriented situations, as occurs during faculty observation of trainees.

### **Bias in faculty assessment**

The differences we observed in assessment scores by resident race/ethnicity may also reflect bias in how faculty assess learner performance. We noted a significant difference in how men and women faculty assessed URiM residents in the PROF competency, which may be due to faculty bias. Evidence suggests that URiM residents and medical students experience racial bias during their training<sup>7,8</sup> and that physician faculty have biases.<sup>42,43</sup> A study of 2,535 physicians using the Implicit Association Test found men physicians displayed more implicit White preference than women physicians.<sup>42</sup> Similarly, a study examining bias



in medical school admission committee members also using the Implicit Association Test showed that men and those who were faculty members had the largest bias measures.<sup>43</sup>

Although we did not include faculty race/ethnicity in our study, according to the Association of American Medical Colleges, the majority of full-time men faculty at U.S. medical schools in 2018 identified with races/ethnicities not underrepresented in medicine.<sup>44</sup> Given this, we speculate that in-group favoritism may play a role in our findings that ratings of non-URiM residents from men faculty were higher than those in all other resident–faculty pairing. In-group favoritism, in which people demonstrate preference for others from similar social groups, may manifest as overvaluing the efforts of individuals in non-URiM groups while conversely devaluing the efforts of those in URiM groups.

The difference we observed in rating patterns between men and women faculty in the PROF competency is notable. There is concern that professionalism may serve to sustain the values and norms of the social groups in the majority in medicine.<sup>45,46</sup> As a domain to be assessed, professionalism may facilitate excess scrutiny of the behaviors, mannerisms, and appearance of learners from minority groups.<sup>45–47</sup> Evidence suggests that URiM residents' experience with the hidden curriculum around professionalism often includes the implication that certain aspects of racial/ethnic identity, such as dress, hair, and speech, lack professionalism.<sup>9</sup>

### **Structural inequities in assessment**

Finally, the differences we observed in assessment scores by resident race/ethnicity may reflect structural inequities in assessment measures. We noted the greatest difference in scores between URiM and non-URiM residents in the SBP competency, which involves working effectively and coordinating care in various health care systems and interprofessional teams.<sup>30</sup> The SBP competency is known to be subjective and difficult to assess, which may enable faculty bias in the assessment of this competency.<sup>48</sup> As the SBP competency involves interaction with the health care system, its assessment, in part, reflects the actions and reactions of the health care system.<sup>48,49</sup> Differences in SBP scores may therefore reflect structural bias in how the health care system interacts with URiM residents.

### **Implications and future research**

Importantly, we must consider the implications of these differences in assessment scores for both trainees and training programs. Evidence suggests that even small differences in performance assessment scores can have a cascade effect and compound disparities in subsequent outcomes.<sup>23</sup> Most notably, assessment in competency-based medical education has implications for resident readiness to practice.<sup>50</sup> Differences in assessment may also impact other outcomes, such as receiving awards, access to and caliber of fellowship training opportunities, and achieving leadership positions such as chief resident.<sup>51,52</sup>

In addition, disparities in assessment may have detrimental effects on learner engagement with the profession. Negative experiences related to race/ethnicity in the learning environment, including inequities in assessment, may negatively impact learners' mental health and well-being and erode their sense of altruism, empathy, and enthusiasm for the

profession.<sup>53–56</sup> Ultimately, these compounded effects may result in fewer URiM faculty at academic institutions and further reinforce the disparities we observed in this study.

While this study is exploratory in nature, our results hint at larger intrinsic or structural inequities in graduate medical education and point to a critical need to study and promote equitable assessment in medical education.<sup>57,58</sup> Specifically, further study of racial/ethnic disparities in assessment and the impact on learners is needed, including how systems of assessment support intrinsic equity and how to address equity in faculty assessment training.<sup>57</sup>

In our study, we incorporated baseline IM-ITE percentile rank as a measure of baseline medical knowledge. While we adjusted for this variable in our analysis, we did note a difference between URiM and non-URiM residents that has not been reported previously. While IM-ITE results correlate with board certification pass rates, evidence supporting the use of the exam as a predictor of clinical performance is lacking.<sup>59,60</sup> Further research is needed to understand disparities in IM-ITE results and the impact on residents.

These findings should be interpreted in the context of the limitations of our study. First, our study is exploratory in nature and further research is needed to confirm these findings. We did not adjust the level of significance for multiple comparisons to enable comparison with future work and because of concern for Type II error or failing to detect true disparities associated with race/ethnicity.<sup>37</sup> Our results should be interpreted in this light. The use of resident self-reported race/ethnicity information obtained from residency applications limits our ability to discern disparities within and between racial and ethnic groups and may not adequately capture the experience of those belonging to multiple racial and ethnic groups.<sup>61,62</sup> We were unable to explore the impact of faculty race/ethnicity on resident assessment scores. Further research is needed into the effects of faculty race/ethnicity, age, and rank on disparities in assessment scores.

While we examined the interaction between resident race/ethnicity and gender, we were not able to assess the effects of other factors that may intersect with racial/ethnic inequities, such as nationality, socioeconomic class, sexual orientation, or disability status. Small numbers of URiM residents limited our ability to assess for interactions between multiple variables. We used gender designations as determined by participants' professional gender identity, and our results do not capture the experiences of those identifying as genderqueer or nonbinary. Finally, our study focused on internal medicine residency training programs at academic medical centers, which may limit the generalizability of our findings across fields. A larger-scale study of the interaction between resident race/ethnicity and gender in assessment metrics is a planned next step.

Our study provides novel evidence of disparities in assessment associated with resident race/ethnicity in graduate medical education. While attention has focused on recruiting a diverse workforce, effort is also needed to achieve equity within medical education.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments:

The authors acknowledge the following individuals for reviewing earlier versions of this report: J. Sawalla Guseh II, MD, Massachusetts General Hospital; Taison Bell, MD, University of Virginia; and Francois Rollin, MD, Emory University School of Medicine.

## Funding/Support:

This project was supported by the Association of American Medical Colleges Southern Group on Educational Affairs. Creation of this dataset was supported in part by the Josiah Macy Jr. Foundation President's Grant to Robin Klein.

## References

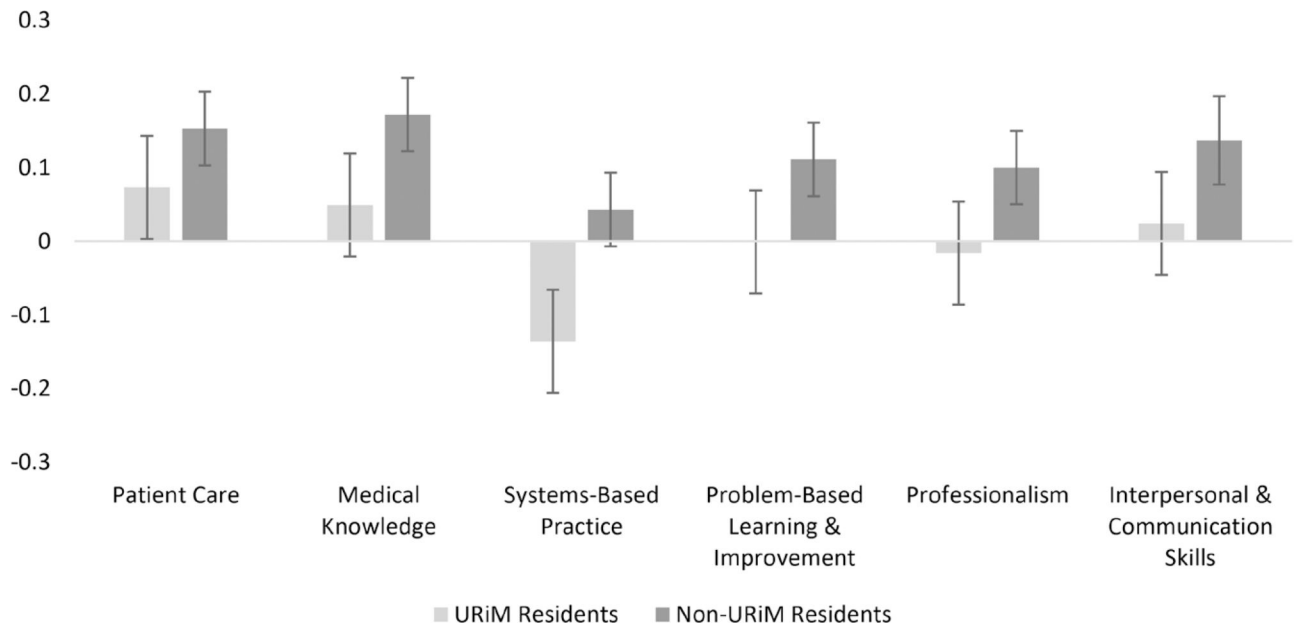
1. Egener BE, Mason DJ, McDonald WJ, et al. The charter on professionalism for health care organizations. *Acad Med.* 2017;92:1091–1099. [PubMed: 28079726]
2. Association of American Medical Colleges. AAMC Statement on Gender Equity. <https://www.aamc.org/what-we-do/equity-diversity-inclusion/aamc-statement-gender-equity>. Published January 2020. Accessed April 13, 2022.
3. Fang D, Moy E, Colburn L, Hurley J. Racial and ethnic disparities in faculty promotion in academic medicine. *JAMA.* 2000;284:1085–1092. [PubMed: 10974686]
4. Ly DP, Seabury SA, Jena AB. Differences in incomes of physicians in the United States by race and sex: Observational study. *BMJ.* 2016;353:i2923. [PubMed: 27268490]
5. Khan MS, Lakha F, Tan MMJ, et al. More talk than action: Gender and ethnic diversity in leading public health universities. *Lancet.* 2019;393:594–600. [PubMed: 30739695]
6. Crenshaw K Demarginalizing the intersection of race and sex: A Black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *Univ Chic Leg Forum.* 1989;140:139–167.
7. Hill KA, Samuels EA, Gross CP, et al. Assessment of the prevalence of medical student mistreatment by sex, race/ethnicity, and sexual orientation. *JAMA Intern Med.* 2020;180:653–665. [PubMed: 32091540]
8. Fnais N, Soobiah C, Chen MH, et al. Harassment and discrimination in medical training: A systematic review and meta-analysis. *Acad Med.* 2014;89:817–827. [PubMed: 24667512]
9. Osseo-Asare A, Balasuriya L, Huot SJ, et al. Minority resident physicians' views on the role of race/ethnicity in their training experiences in the workplace. *JAMA Netw Open.* 2018;1:e182723. [PubMed: 30646179]
10. Bullock JL, Lockspeiser T, Del Pino-Jones A, Richards R, Teherani A, Hauer KE. They don't see a lot of people my color: A mixed methods study of racial/ethnic stereotype threat among medical students on core clerkships. *Acad Med.* 2020;95(11 suppl):S58–S66.
11. de Bourmont SS, Burra A, Nouri SS, et al. Resident physician experiences with and responses to biased patients. *JAMA Netw Open.* 2020;3:e2021769. [PubMed: 33226429]
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:e0181659. [PubMed: 28792940]
13. Low D, Pollack SW, Liao ZC, et al. Racial/ethnic disparities in clinical grading in medical school. *Teach Learn Med.* 2019;31:487–496. [PubMed: 31032666]
14. Campos-Outcalt D, Rutala PJ, Witzke DB, Fulginiti JV. Performances of underrepresented-minority students at the University of Arizona College of Medicine, 1987–1991. *Acad Med.* 1994;69:577–582. [PubMed: 8018270]
15. Reteguiz J, Davidow AL, Miller M, Johanson WG Jr. Clerkship timing and disparity in performance of racial-ethnic minorities in the medicine clerkship. *J Natl Med Assoc.* 2002;94:779–788. [PubMed: 12392041]
16. Rojek AE, Khanna R, Yim JW, et al. Differences in narrative language in evaluations of medical students by gender and under-represented minority status. *J Gen Intern Med.* 2019;34:684–691. [PubMed: 30993609]

17. 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;99:1138–1150. [PubMed: 17987918]
18. Lee KB, Vaishnavi SN, Lau SK, Andriole DA, Jeffe DB. Cultural competency in medical education: Demographic differences associated with medical student communication styles and clinical clerkship feedback. *J Natl Med Assoc.* 2009;101:116–126. [PubMed: 19378627]
19. Stegers-Jager KM, Steyerberg EW, Cohen-Schotanus J, Themmen AP. Ethnic disparities in undergraduate pre-clinical and clinical performance. *Med Educ.* 2012;46:575–585. [PubMed: 22626049]
20. Woolf K, Potts HW, McManus IC. Ethnicity and academic performance in UK trained doctors and medical students: Systematic review and meta-analysis. *BMJ.* 2011;342:d901. [PubMed: 21385802]
21. Woolf K, Rich A, Viney R, Needleman S, Griffin A. Perceived causes of differential attainment in UK postgraduate medical training: A national qualitative study. *BMJ Open.* 2016;6:e013429.
22. Bullock JL, Lai CJ, Lockspeiser T, et al. In pursuit of honors: A multi-institutional study of students’ perceptions of clerkship evaluation and grading. *Acad Med.* 2019;94(11 suppl):S48–S56. [PubMed: 31365406]
23. Teherani A, Hauer KE, Fernandez A, King TE Jr, Lucey C. 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;93:1286–1292. [PubMed: 29923892]
24. 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;177:659–665. [PubMed: 28264091]
25. Wijesekera TP, Kim M, Moore EZ, Sorenson O, Ross DA. All other things being equal: Exploring racial and gender disparities in medical school Honor Society induction. *Acad Med.* 2019;94:562–569. [PubMed: 30234509]
26. Grimm LJ, Redmond RA, Campbell JC, Rosette AS. Gender and racial bias in radiology residency letters of recommendation. *J Am Coll Radiol.* 2020;17:64–71. [PubMed: 31494103]
27. Powers A, Gerull KM, Rothman R, Klein SA, Wright RW, Dy CJ. Race- and gender-based differences in descriptions of applicants in the letters of recommendation for orthopaedic surgery residency. *JB JS Open Access.* 2020;5:e20.00023–e20.00023.
28. Harden RM, Crosby JR, Davis MH. AMEE guide no. 14: Outcome based education: Part 1—An introduction to outcome-based education. *Med Teach.* 1999;21:7–14.
29. Frank JR, Mungroo R, Ahmad Y, Wang M, De Rossi S, Horsley T. Toward a definition of competency-based education in medicine: A systematic review of published definitions. *Med Teach.* 2010;32:631–637. [PubMed: 20662573]
30. Accreditation Council for Graduate Medical Education. Internal Medicine Milestones. <https://www.acgme.org/Portals/0/PDFs/Milestones/InternalMedicineMilestones.pdf>. Revised November 2020. Accessed April 13, 2022.
31. Association of American Medical Colleges. ERAS for medical schools. <https://www.aamc.org/services/eras-for-institutions/medical-schools>. Accessed April 13, 2022.
32. Association of American Medical Colleges. Underrepresented in medicine definition. <https://www.aamc.org/what-we-do/diversity-inclusion/underrepresented-in-medicine>. Accessed April 13, 2022.
33. Klein R, Ufere NN, Rao SR, et al. ; Gender Equity in Medicine workgroup. Association of gender with learner assessment in graduate medical education. *JAMA Netw Open.* 2020;3:e2010888. [PubMed: 32672831]
34. Yudkowsky R, Park YS, Downing SM, eds. *Assessment in Health Professions Education.* 2nd ed. New York, NY: Routledge; 2020.
35. Murad MH, Wang Z, Chu H, Lin L. When continuous outcomes are measured using different scales: Guide for meta-analysis and interpretation. *BMJ.* 2019;364:k4817. [PubMed: 30670455]
36. Austin PC, Goel V, van Walraven C. An introduction to multilevel regression models. *Can J Public Health.* 2001;92:150–154. [PubMed: 11338155]

37. Althouse AD. Adjust for multiple comparisons? It's not that simple. *Ann Thorac Surg.* 2016;101:1644–1645. [PubMed: 27106412]
38. Klein R, Julian KA, Snyder ED, et al. ; From the Gender Equity in Medicine (GEM) workgroup. Gender bias in resident assessment in graduate medical education: Review of the literature. *J Gen Intern Med.* 2019;34:712–719. [PubMed: 30993611]
39. 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;177:651–657. [PubMed: 28264090]
40. Orom H, Semalulu T, Underwood W III. The social and learning environments experienced by underrepresented minority medical students: A narrative review. *Acad Med.* 2013;88:1765–1777. [PubMed: 24072111]
41. Burgess DJ, Warren J, Phelan S, Dovidio J, Van Ryn M. Stereotype threat and health disparities: What medical educators and future physicians need to know. *J Gen Intern Med.* 2010;25:169–177. [PubMed: 19898908]
42. Sabin J, Nosek BA, Greenwald A, Rivara FP. Physicians' implicit and explicit attitudes about race by MD race, ethnicity, and gender. *J Health Care Poor Underserved.* 2009;20:896–913. [PubMed: 19648715]
43. Capers Q 4th, Clinchot D, McDougle L, Greenwald AG. Implicit racial bias in medical school admissions. *Acad Med.* 2017;92:365–369. [PubMed: 27680316]
44. Association of American Medical Colleges. Diversity in medicine: Facts and figures 2019. Figure 16. Percentage of full-time U.S. medical school faculty by sex and race/ethnicity, 2018. <https://www.aamc.org/data-reports/workforce/interactive-data/figure-16-percentage-full-time-us-medical-school-faculty-sex-and-race/ethnicity-2018>. Accessed April 13, 2022.
45. Frye V, Camacho-Rivera M, Salas-Ramirez K, et al. Professionalism: The wrong tool to solve the right problem? *Acad Med.* 2020;95:860–863. [PubMed: 32134778]
46. Wyatt TR, Balmer D, Rockich-Winston N, Chow CJ, Richards J, Zaidi Z. “Whispers and shadows”: A critical review of the professional identity literature with respect to minority physicians. *Med Educ.* 2021;55:148–158. [PubMed: 33448459]
47. Lee JH. The weaponization of medical professionalism. *Acad Med.* 2017;92:579–580.
48. Lurie SJ, Mooney CJ, Lyness JM. Measurement of the general competencies of the Accreditation Council For Graduate Medical Education: A systematic review. *Acad Med.* 2009;84:301–309. [PubMed: 19240434]
49. Li JT, Stoll DA, Smith JE, Lin JJ, Swing SR. Graduates' perceptions of their clinical competencies in allergy and immunology: Results of a survey. *Acad Med.* 2003;78:933–938. [PubMed: 14507628]
50. Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR. The role of assessment in competency-based medical education. *Med Teach.* 2010;32:676–682. [PubMed: 20662580]
51. Santhosh L, Babik JM. Trends in racial and ethnic diversity in internal medicine subspecialty fellowships from 2006 to 2018. *JAMA Netw Open.* 2020;3:e1920482. [PubMed: 32031642]
52. Klein R, Law K, Koch J. Gender representation matters: Intervention to solicit medical resident input to enable equity in leadership in graduate medical education. *Acad Med.* 2020;95(12 suppl):S93–S97. [PubMed: 32889942]
53. Hardeman RR, Perry SP, Phelan SM, Przedworski JM, Burgess DJ, van Ryn M. Racial identity and mental well-being: The experience of African American medical students, a report from the medical student CHANGE study. *J Racial Ethn Health Disparities.* 2016;3:250–258. [PubMed: 27271066]
54. Hardeman RR, Przedworski JM, Burke SE, et al. Mental well-being in first year medical students: A comparison by race and gender: A report from the medical student CHANGE study. *J Racial Ethn Health Disparities.* 2015;2:403–413. [PubMed: 26413458]
55. Banos JH, Noah JP, Harada CN. Predictors of student engagement in learning communities. *J Med Educ Curric Dev.* 2019;6:2382120519840330. [PubMed: 31025001]

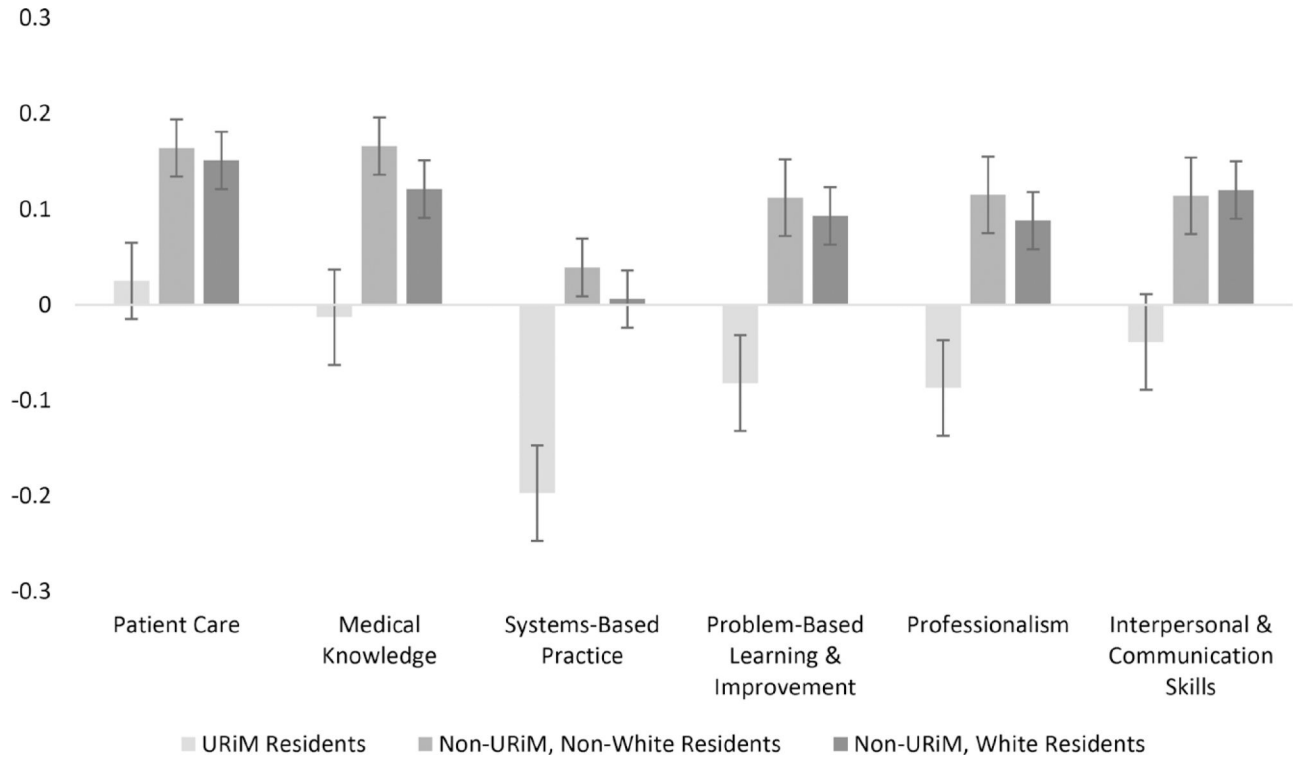
56. Pradhan A, Buery-Joyner SD, Page-Ramsey S, et al. To the point: Undergraduate medical education learner mistreatment issues on the learning environment in the United States. *Am J Obstet Gynecol.* 2019;221:377–382. [PubMed: 31029660]
57. Colbert CY, French JC, Herring ME, Dannefer EF. Fairness: The hidden challenge for competency-based postgraduate medical education programs. *Perspect Med Educ.* 2017;6:347–355. [PubMed: 28516341]
58. Lucey CR, Hauer KE, Boatright D, Fernandez A. Medical education’s wicked problem: Achieving equity in assessment for medical learners. *Acad Med.* 2020;95(12 suppl):S98–S108. [PubMed: 32889943]
59. Schwartz RW, Donnelly MB, Sloan DA, Johnson SB, Strodel WE. The relationship between faculty ward evaluations, OSCE, and ABSITE as measures of surgical intern performance. *Am J Surg.* 1995;169:414–417. [PubMed: 7694980]
60. Babbott SF, Beasley BW, Hinchey KT, Blotzer JW, Holmboe ES. The predictive validity of the internal medicine in-training examination. *Am J Med.* 2007;120:735–740. [PubMed: 17679136]
61. Ross PT, Hart-Johnson T, Santen SA, Zaidi NLB. Considerations for using race and ethnicity as quantitative variables in medical education research. *Perspect Med Educ.* 2020;9:318–323. [PubMed: 32789666]
62. Khunti K, Routen A, Pareek M, Treweek S, Platt L. The language of ethnicity. *BMJ.* 2020;371:m4493. [PubMed: 33229395]



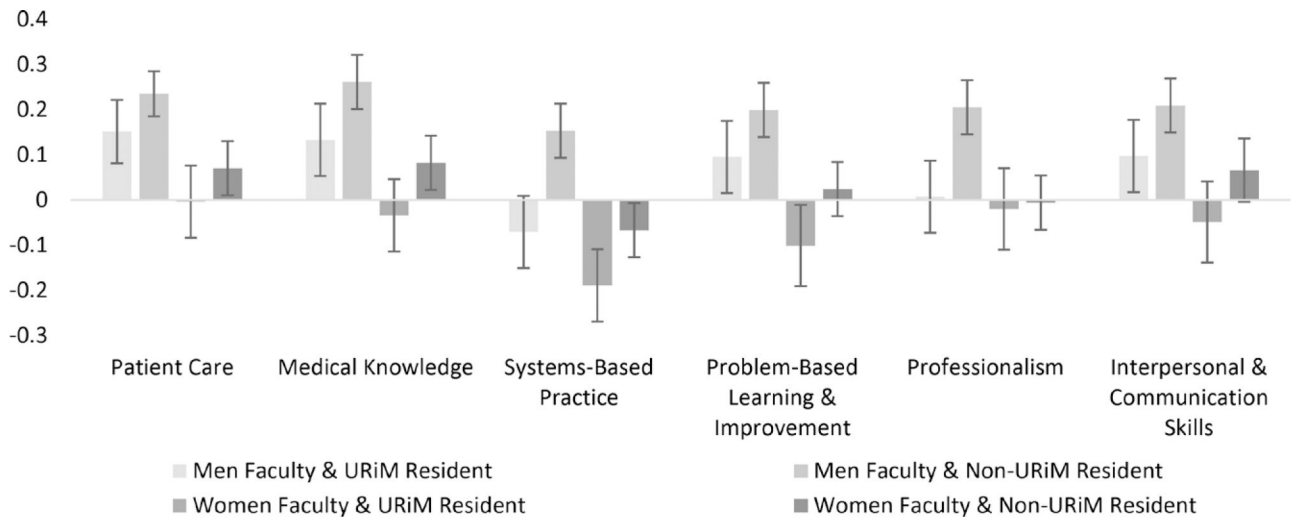


**Figure 1.**

Mean adjusted standardized core competency scores for underrepresented in medicine (URiM) and not underrepresented in medicine (non-URiM) residents in a study of the association between resident race/ethnicity and assessment scores, 2016–2017. Mean adjusted standardized scores, standard errors, and *P* values were obtained from cross-classified random-intercept mixed models adjusted for resident gender, postgraduate year, and baseline Internal Medicine In-Training Examination percentile rank; rotation time of year (July–September, October–December, January–March, April–June) and setting (university, Veterans Administration, community or public hospital); and faculty gender, academic rank (assistant professor/instructor/ chief resident, associate professor, professor, no rank/clinical associate), and specialty (general medicine, hospital medicine, subspecialty).



**Figure 2.** Mean adjusted standardized core competency scores by resident race/ethnicity in a study of the association between resident race/ethnicity and assessment scores, 2016–2017. Mean adjusted standardized scores, standard errors, and *P* values were obtained from cross-classified random-intercept mixed models adjusted for resident gender, postgraduate year, and baseline Internal Medicine In-Training Examination percentile rank; rotation time of year (July–September, October–December, January–March, April–June) and setting (university, Veterans Administration, community or public hospital); and faculty gender, academic rank (assistant professor/instructor/chief resident, associate professor, professor, no rank/clinical associate), and specialty (general medicine, hospital medicine, subspecialty). Abbreviations: URiM, underrepresented in medicine; non-URiM, not underrepresented in medicine.



**Figure 3.**

Mean adjusted standardized core competency scores for underrepresented in medicine (URiM) and not underrepresented in medicine (non-URiM) residents by faculty gender in a study of the association between resident race/ethnicity and assessment scores, 2016–2017. Mean adjusted standardized scores, standard errors, and *P* values were obtained from cross-classified random-intercept mixed models adjusted for resident gender, postgraduate year, and baseline Internal Medicine In-Training Examination percentile rank; rotation time of year (July–September, October–December, January–March, April–June) and setting (university, Veterans Administration, community or public hospital); and faculty academic rank (assistant professor/instructor/chief resident, associate professor, professor, no rank/clinical associate) and specialty (general medicine, hospital medicine, subspecialty).

**Table 1**

Characteristics of Resident and Faculty Participants in a Study of the Association Between Resident Race/Ethnicity and Assessment Scores, 2016–2017

Characteristic	No. (%)
<b>No. residents</b>	703
<b>Resident URiM designation</b>	
URiM	94 (13.4)
Non-URiM	609 (86.6)
<b>Resident gender</b>	
Men	387 (55.0)
Women	316 (45.0)
<b>Resident postgraduate year</b>	
PGY-1	269 (38.3)
PGY-2	226 (32.1)
PGY-3	208 (29.6)
<b>No. faculty</b>	605
<b>Faculty gender</b>	
Men	318 (52.6)
Women	287 (47.4)
<b>Faculty academic rank</b>	
Professor	111 (18.3)
Associate professor	115 (19.0)
Assistant professor or instructor	323 (53.4)
Chief resident	30 (5.0)
No rank or clinical associate	26 (4.3)
<b>Faculty specialty</b>	
General medicine	239 (39.5)
Hospital medicine	223 (36.9)
Subspecialty	143 (23.6)
<b>Faculty educational role</b>	
Program director	8(1.3)
Associate program director	35 (5.8)
Chief resident	31 (5.1)

Abbreviations: URiM, underrepresented in medicine; non-URiM, not underrepresented in medicine; PGY, postgraduate year.