UCLA
UCLA Previously Published Works

Title

Association of Gender and Race/Ethnicity with Internal Medicine In-Training Examination Performance in Graduate Medical Education

Permalink

https://escholarship.org/uc/item/81r68597

Journal

Journal of General Internal Medicine, 37(9)

ISSN

0884-8734

Authors

Klein, Robin
Koch, Jennifer
Snyder, Erin D
et al.

Publication Date

2022-07-01

DOI

10.1007/s11606-022-07597-z

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Association of Gender and Race/Ethnicity with Internal Medicine In-Training Examination Performance in Graduate Medical Education

Robin Klein, MD, MEHP[1], Jennifer Koch, MD[2], Erin D. Snyder, MD[3], Anna Volerman, MD[4], Wendy Simon, MD[5], Simerjot K. Jassal, MD[6], Dominique Cosco, MD[7], Anne Cioletti, MD[8], Nneka N. Ufere, MD[9], Sherri-Ann M. Burnett-Bowie, MD[9], Kerri Palamara, MD[9], Sarah Schaeffer, MD[10], Katherine A. Julian, MD[10], and Vanessa Thompson, MD[10]

[1]Department of Medicine, Division of General Internal Medicine and Geriatrics, Emory University School of Medicine, 49 Jesse Hill Jr Dr, Atlanta, GA, USA; [2]Department of Medicine, University of Louisville, Louisville, KY, USA; [3]Department of Medicine, University of Alabama Birmingham School of Medicine, Birmingham, AL, USA; [4]Departments of Medicine and Pediatrics, University of Chicago, Chicago, IL, USA; [5]Department of Medicine, University of California, Los Angeles, Los Angeles, USA; [6]Department of Medicine, VA San Diego Healthcare System, University of California, San Diego, San Diego, USA; [7]Department of Medicine, Washington University St. Louis, St. Louis, USA; [8]Department of Medicine, University of Utah, Salt Lake City, USA; [9]Department of Medicine, Massachusetts General Hospital, Boston, MA, USA; [10]Department of Medicine, University of California, San Francisco, San Francisco, CA, USA.

BACKGROUND: Disparities in objective assessments in graduate medical education such as the In-Training Examination (ITE) that disadvantage women and those self-identifying with race/ethnicities underrepresented in medicine (URiM) are of concern.

OBJECTIVE: Examine ITE trends longitudinally across post-graduate year (PGY) with gender and race/ethnicity.

DESIGN: Longitudinal analysis of resident ITE metrics at 7 internal medicine residency programs, 2014–2019. ITE trends across PGY of women and URiM residents compared to non-URiM men assessed via ANOVA. Those with ITE scores associated with less than 90% probability of passing the American Board of Internal Medicine certification exam (ABIM-CE) were identified and odds of being identified as at-risk between groups were assessed with chi square.

PARTICIPANTS: A total of 689 IM residents, including 330 women and URiM residents (48%).

MAIN MEASURES: ITE score

KEY RESULTS: There was a significant difference in ITE score across PGY for women and URiM residents compared to non-URiM men ($F_{(2,\ 1321)}$ 4.46, p=0.011). Adjusting for program, calendar year, and baseline ITE, women and URiM residents had smaller ITE score gains (adjusted mean change in score between PGY1 and PGY3 (se), non-URiM men 13.1 (0.25) vs women and URiM residents 11.4 (0.28), p<0.001). Women and URiM residents had greater odds of being at potential risk for not passing the ABIM-CE (OR 1.75, 95% CI 1.10 to 2.78) with greatest odds in PGY3 (OR 3.13, 95% CI 1.54 to 6.37).

CONCLUSION: Differences in ITE over training were associated with resident gender and race/ethnicity. Women and URiM residents had smaller ITE score gains across PGY translating into greater odds of potentially being seen as at-risk for not passing the ABIM-CE. Differences in ITE over training may reflect differences in experiences of women and URiM residents during training and may lead to further disparities.

Received August 13, 2021
Accepted March 30, 2022
Published online June 16, 2022

J Gen Intern Med 37(9):2194–9
DOI: 10.1007/s11606-022-07597-z
© The Author(s), under exclusive licence to Society of General Internal Medicine 2022

INTRODUCTION

Evidence suggests there are disparities in learner assessment in medical education that disadvantage women and those identifying with race/ethnicities underrepresented in medicine (URiM).[1–9] Most notably, this includes differences in ratings and language used in assessments associated with gender and race/ethnicity.[6–8]

Less attention has focused on disparities in objective assessments of learners including the In-Training Examination (ITE). A standardized test of medical knowledge used in internal medicine (IM) residency programs in the USA, the ITE is designed to provide residents and programs with an objective assessment of medical knowledge and to identify educational needs of both programs and learners.[10,11]

While studies primarily focused on clinical performance assessments have noted differences in ITE scores associated with gender and race/ethnicity,[8,9,12] dedicated study is lacking. This study aims to explore trends in ITE over training by resident gender and race/ethnicity.

The In-Training Examination

Administered annually, the IM ITE is modeled after the American Board of Internal Medicine's certification exam (ABIM-CE).[10] Questions are written at the expected

proficiency level of a post-graduate year (PGY) 2 resident and cover key content areas in the field. ITE performance metrics, including score and percentile rank, are reported to both the resident and program. ITE score is the percent of questions answered correctly while percentile rank is a relative measure of ITE performance comparing a resident's score with scores of all IM residents at the same PGY level nationwide.[10]

Intended to assess medical knowledge, ITE scores generally increase annually as residents gain knowledge during training.[13,14] ITE performance has been shown to correlate with other measures of medical knowledge including medical knowledge milestones and ABIM-CE performance.[12–21] The majority of IM training programs administer the ITE to residents in all 3 years of training to assess resident medical knowledge.[21]

Multiple studies have demonstrated an association between ITE and ABIM-CE performance.[12–20] While ITE performance in any PGY is associated with ABIM-CE performance, ITE score in PGY3 has higher predictive value for passing the ABIM-CE compared to earlier tests.[16,20] Given ITE scores generally increase over training, residents must score higher on each subsequent administration to maintain the same estimated probability of passing the ABIM-CE.[20] A national study established probabilities of passing the ABIM-CE for ITE scores in each year of training so that ITE scores of 52% in PGY1, 60% in PGY2, and 65% in PGY3 were associated with approximately a 90% probability passing the ABIM-CE.[20]

METHODS

We conducted a longitudinal analysis of ITE performance metrics of residents at seven US IM residency training programs from 2014 to 2019: University of Alabama (Birmingham), University of Chicago, University of Louisville, University of Texas (Austin), University of California San Diego, University of California Los Angeles, and Washington University.

Participants included IM residents who started and completed training between 2014 and 2019 at each site in our study. Residents on specialized tracks outside of the 3 years traditionally allotted for IM training, such as ABIM Research Pathway or Medicine-Pediatrics, were excluded.

We collected resident gender and self-reported race/ethnicity information and ITE score and percentile rank for each post-graduate year of training. Gender designations were determined by participants' professional gender identity. Self-reported race/ethnicity information obtained from residency applications (Electronic Residency Application Service) was used to determine those who identified with race/ethnicities underrepresented in medicine.[22] We used the Association of American Medical Colleges (AAMC)'s definition of URiM to include African American and/or Black, Hispanic/Latinx, Native American (American Indian, Alaska Native, and Native Hawaiian), and Pacific Islander.[23] Other race/ethnicities were categorized as not underrepresented in medicine (non-URiM).

Resident characteristics and ITE metrics were extracted from education management systems at each site. Data was de-identified and analysis was performed on de-identified data in aggregate.

We examined patterns in ITE across PGY by resident gender and race/ethnicity. First, we explored trends in ITE score across PGY by resident gender and URiM designation separately. After examining trends of individual demographic groups, we grouped women and URiM residents together and examined ITE trends of women and URiM residents compared to with their non-URiM men resident peers. This approach was used for multiple reasons. First, when looking at ITE trends with resident gender and race/ethnicity separately in this dataset, we found similar patterns in both groups. Second, evidence suggests disparities in ITE impacting both women and URiM residents.[8,9,24] Finally, this approach shifts focus from the ITE performance of individual demographic groups, especially those with small numbers of residents.

We explored ITE score trends at the resident and program levels. At the resident level, a repeated-measure ANOVA was performed to evaluate differences in overall trend in ITE across repeated measures (i.e., post-graduate year) between groups.[25] Specifically, we used repeated-measure ANOVA with an interaction term for PGY and groups to assess for differences in ITE across PGY between groups and controlled for program and calendar year the ITE was taken.

Then, we analyzed gain scores between groups. A commonly used index of performance in educational research, gain score is the change in score between two testing events.[25] In this case, gain score is the change in ITE between PGY1 and PGY3. We calculated change in ITE (i.e., gain score) for each participant and then calculated the mean change in ITE for groups. We assessed for differences in mean change in ITE between groups and controlled for program, calendar year of the ITE, and baseline ITE score, defined as the PGY1 ITE, to account for score differences at baseline. Focusing on gain score or change in ITE provides a more intuitive understanding of how the ITE differs over time between groups.

At the program level, we examined the association between gain scores of groups and the proportion of women and URiM residents within each program. For each program, we computed the difference in gain scores between groups as follows:

$$\text{Difference in change in ITE} = \text{Change in ITE for non} - \text{URiM men residents} - \text{Change in ITE for women and URiM residents}$$

For the difference in change in ITE, a positive value indicates greater ITE gains for non-URiM men, a negative value indicates greater ITE gains for women and URiM residents, and a value of 0 indicates no difference between groups. We

assessed the correlation between the difference in gain scores of a program and the proportion of women and URiM residents within that program using Pearson correlation.

Finally, we assessed the gender and race/ethnicity of residents at potential risk for not passing the ABIM-CE based on ITE score. We used ITE scores to identify those at potential risk for not passing the ABIM-CE, defined as those with an ITE score associated with a less than 90% probability of passing the ABIM-CE. Using ITE scores and their associated probabilities of passing the ABIM-CE derived from the literature, we set ITE score thresholds for those at risk (PGY1 ITE score ≤ 52%, PGY2 ITE score ≤ 60%, and PGY3 ITE score ≤ 65%).[20] After identifying those at risk, we assessed the odds of being identified as at risk between groups using chi-square.

We report ANOVA results including the degrees of freedom, *F* statistic for between groups comparison, and *p* value and for correlation; we report chi-square results including degrees of freedom, sample size, chi-square statistic, and *p* value.

Institutional Review Boards at each institution reviewed and exempted the study protocol.

RESULTS

Data was collected on 689 residents, including 286 (41.5%) women and 75 (10.9%) URiM residents (Table 1). Taken together, this included 330 women and URiM residents (47.9%) and 356 non-URiM men residents (51.9%). Overall, ITE scores increased (mean change in ITE score +12.4%) and percentile rank decreased (mean change in percentile rank −2.7) over residents' 3 years of training.

Looking at trends by gender and race/ethnicity separately, similar trends in ITE were seen with women compared to men and URiM residents compared to non-URiM residents (Table 2). Adjusting for program, calendar year, and baseline ITE, women had smaller gains in ITE score (adjusted mean change in score between PGY1 and PGY3, men vs women 12.8 vs 11.7) and a greater decline in percentile rank than men (adjusted mean change in percentile rank, −0.9 vs −5.3). A similar pattern was seen in ITE metrics of URiM residents including smaller gains in ITE score (adjusted mean change in score, non-URiM vs URiM residents, 12.5 vs 10.6) and a greater decrease in percentile rank (mean change in percentile rank, non-URiM vs URiM residents, −2.4 vs −5.0).

Table 1 Characteristics of Residents and ITE Performance Metrics

	N (%)	Range across programs
Number of residents, *N*	689	44 to 142
Resident gender		
Men, *n* (%)	400 (58.5%)	46.9 to 73.6%
Women, *n* (%)	286 (41.5%)	26.4 to 53.1%
Resident race/ethnicity		
Non-URiM residents	614 (89.1%)	78.3 to 95.8%
URiM residents	75 (10.9%)	4.2 to 21.7%
Resident gender and race/ethnicity		
Non-UriM, men residents, *n* (%)	359 (51.9%)	35.4 to 70.8%
Women and URiM residents, *n* (%)	330 (47.9%)	29.2 to 64.6%
Change in ITE score*, mean (se)	12.41 (0.20)	
Change in ITE percentile rank*, mean (se)	−2.7 (0.69)	

Abbreviations: URiM, underrepresented in medicine; Non-URiM, not underrepresented in medicine; IM-ITE, Internal Medicine In-Training Examination; PGY, post-graduate year;
**Change in ITE is the difference in ITE metric between PGY1 and PGY3*

Resident-Level Trends in ITE

Considering women and URiM residents together, there were significant differences in ITE performance across post-graduate year for women and URiM residents compared to non-URiM men residents in both ITE score ($F_{(2,\ 1321)}$ 4.46, p 0.011) and ITE percentile rank ($F_{(2,\ 1321)}$ 7.23, p 0.001) after controlling for program and date of ITE.

Adjusting for program, calendar year of the ITE, and baseline ITE performance, women and URiM residents had smaller ITE score gains over training than non-URiM men residents (adjusted mean change in score between PGY1 and PGY3 (se), non-URiM men residents 13.1 (0.25) vs women and URiM residents 11.4 (0.28), $p < 0.001$) and greater decline in ITE percentile rank (adjusted mean change in percentile rank between PGY1 and PGY3 (se), non-URiM men residents 0.0 (0.93) vs women and URiM residents −5.8 (1.02), $p <$ 0.001) than their non-URiM men resident peers.

Program-Level Trends in ITE

In all programs, the difference in gain scores of women and URiM residents and non-URiM men residents indicated greater ITE gains for non-URiM men residents. The correlation between the difference in change in ITE score over training between groups and the proportion of women and URiM residents in a program approached but did not meet the standard threshold for significance (r −0.68, p 0.09).

Identifying At-Risk Residents

Residents at potential risk for not passing the ABIM-CE (less than 90% probability) were identified using ITE score thresholds (PGY1 ITE score ≤ 52%, PGY2 ITE score ≤ 60%, and PGY3 ITE score ≤ 65%). See Table 3. Using ITE from any PGY year, 85 residents (12%) met criteria for being at risk for not passing the ABIM based on ITE score.

Overall, there was a significant difference in proportion of women and URiM residents identified as at risk for not passing the ABIM-CE compared to non-URiM men resident (X^2(1, 685) =5.04, p 0.02). Using scores in any PGY, women and URiM residents had greater odds of being identified as at risk compared to non-URiM, men residents (OR 1.75, 95% CI 1.10 to 2.78). Women and URiM residents had the greatest odds of being at risk in PGY3 (OR 3.13, 95% CI 1.54 to 6.37).

Table 2 Change in ITE Metrics over Training by Resident Gender and Race/Ethnicity

	Change in ITE score*, unadjusted mean	Change in ITE score*, adjusted mean† (se)	*P* value††	Change in ITE rank*, unadjusted mean (se)	Change in ITE rank*, adjusted mean† (se)	*P* value††
Resident gender						
Men	12.70	12.8 (0.24)	0.005	−1.2	−0.9 (0.89)	0.001
Women	12.99	11.7 (0.30)		−4.9	−5.3 (1.09)	
Resident race/ethnicity						
Non-URiM residents	12.53	12.5 (0.20)	0.001	−2.4	−2.2 (0.75)	0.03
URiM residents	11.63	10.6 (0.56)		−5.0	−6.8 (2.07)	
Resident gender and race/ethnicity						
Non-URiM men residents	12.93	13.1 (0.25)	<0.001	−0.4	0.0 (0.93)	<0.001
Women and URiM residents	11.82	11.4 (0.28)		−5.3	−5.8 (1.02)	

Abbreviations: URiM, underrepresented in medicine; Non-URiM, not underrepresented in medicine; ITE, In-Training Examination
**Change in ITE is difference in ITE metric between PGY1 and PGY3*
† Means adjusted for program, calendar year and baseline ITE (ITE from PGY1)
†† P value of difference in change in ITE metric between groups adjusting for program, calendar year and baseline ITE from PGY1

DISCUSSION

In this multi-site, longitudinal study, we found significant differences in ITE trend across PGY associated with resident gender and race/ethnicity. Specifically, women and URiM residents had smaller ITE score gains over training than their peers and this translated into greater odds of being seen as potentially at risk for not passing the ABIM-CE than their non-URiM men resident peers.

Our findings are consistent with prior studies noting small but significant differences in ITE performance associated with resident race/ethnicity and gender.[8,9,12,24] Importantly, our findings indicate a difference in the overall trajectory of ITE scores over the course of training when controlling for differences in baseline ITE performance. This is consistent with a single-institution study in orthopedics which found a decline in ITE percentile rank during training for women and URiM residents despite no difference in baseline United States Medical Licensing Examination (USMLE) Step 1 scores.[24] Although race/ethnicity was not included, a national study of the medical knowledge milestones noted significant differences in ITE scores between men and women residents in PGY2 and PGY3.[12]

These disparities in ITE trends over training may reflect a difference in the educational experiences of women and URiM residents during training. Growing evidence suggests differences in the training experience of women and URiM learners compared to their peers. Women and URiM medical students more often report experiencing mistreatment, microaggressions, bias, and discrimination related to gender and race/ethnicity.[26–28] Women and URiM learners report that these negative experiences related to race/ethnicity and gender impact their learning and behaviors.[29–32]

The culture of learning environments has been linked to learners' academic performance on standardized examinations.[33] Microaggressions, mistreatment, and biased behavior can create a hostile and invalidating learning environment that interferes with the ability to focus, learn, work collaboratively, and develop trust, all of which can impair performance, satisfaction, and engagement with their learning environment.[28,34,35]

Disparities in ITE performance during training may be a potential marker of less inclusive learning environments. We found the correlation between the proportion of women and URiM residents in a program and the difference in ITE performance over training between women and URiM residents compared to their peers approached but did not meet the standard threshold for significance. While this must be interpreted with caution, this relationship, that as the proportion of women and URiM residents in a program increased there was less of a difference in ITE gains between women and URiM residents compared to their peers, is interesting. Viewing the

Table 3 At-Risk for Not Passing the ABIM-CE Based on ITE Score by Resident Gender and Race/Ethnicity

	ITE score threshold*	Odds ratio†, 95% CI		*P* value††
		Non-URiM men residents	Women and URiM residents	
Residents at-risk for not passing the ABIM-CE (less than 90% probability) by ITE score				
PGY1	ITE score ≤ 52%	1.00	1.70 (0.96 to 3.02)	0.07
PGY2	ITE score ≤ 60%	1.00	1.76 (0.94 to 3.31)	0.08
PGY3	ITE score ≤ 65%	1.00	3.13 (1.54 to 6.37)	0.001
Overall	Any year ITE score threshold	1.00	1.75 (1.10 to 2.78)	0.017

Abbreviations: ITE, In-Training Examination; ABIM-CE, American Board of Internal Medicine Certification Examination; URiM, underrepresented in medicine; Non-URiM, not underrepresented in medicine
**ITE score threshold used to identify those with less than 90% probability of passing the ABIM-CE*
†Odds ratio for being identified as at risk for not passing the ABIM-CE based on ITE score
††P value via chi-square test

proportion of women and URiM residents as a surrogate marker of a diverse learning environment, a potential relationship with ITE warrants further study.

Diversity in the learning environment has beneficial impact on learners as it exposes learners to a broader view of what is possible for their future and what is supported in that workplace culture. Experience with diversity in learning environments has been shown to affect learner's own implicit bias, critical thinking, and sense of self-efficacy.[36–39] Given its importance, accessible metrics to assess the impact of diversity efforts are needed. Further study is needed to explore potential of ITE performance over time as an indicator of disparate learning environments.

Importantly, we must consider the implications of these small but significant differences in ITE metrics over training. To illustrate the potential effects of these small differences in ITE metrics, we used ITE scores to identify those at potential risk for not passing the ABIM-CE. We found that women and URiM residents had greater odds of being identified as at-risk overall and this was most notable in PGY3.

Programs use ITE performance to identify both high- and low-achieving residents. Most IM programs use ITE thresholds to identify residents at risk for not passing the ABIM-CE[21] and implement interventions based on ITE. This includes requiring participation in knowledge remediation programs and implementing accountability systems of incentives and consequences based on ITE performance.[40–45] Small differences in ITE may mean women and URiM residents are more likely to be identified as needing remediation and less likely to be identified for accolades and opportunities, such as chief resident.

While remediation for residents in need is important, it is not without consequence. Residents associate remediation with negative stigma and feelings of anxiety, shame, and guilt.[46] Our findings suggest that underperformance on the ITE may not simply be an individual resident problem requiring an individual resident intervention. Rather, interventions targeting ITE performance may benefit from including both resident-level remediation and careful consideration of differences in the educational experiences of women and URiM residents.

Finally, differences in ITE performance during training may negatively impact learners in other, more subtle ways. Small differences in ITE score translate into larger declines in ITE percentile rank. As percentile rank indicates performance relative to peers, the decline percentile rank may serve to reinforce to residents and program leaders an implicit bias that women and URiM residents are lacking compared to their peers.

We used ITE score to identify those at potential risk for not passing the ABIM-CE. While the majority of IM program directors report using ITE percentile rank to identify at-risk residents[21], the often cited threshold of 35th percentile is based on early research.[13,17] ITE score is a more robust measure to identify those at risk as percentile rank is a relative measure that varies with changes in performance of the comparison peer group.[20]

Our study must be considered in light of its limitations. Although we saw similar trends in ITE performance for women and URiM residents, it is important to note that women and URiM learners face distinct inequities during training. The low number of URiM residents in our sample was a limitation. Limitations of data sources in turn limited our ability to explore the intersecting effects of gender and race/ethnicity and the ITE of those who identify as genderqueer or nonbinary and those belonging to multiple racial and ethnic groups. Other social identity factors not included in our analysis may play a role. Generalizability may be limited for ITEs in fields outside of IM.

This multi-site, longitudinal study found significant differences in ITE across PGY associated with resident gender and race/ethnicity to the disadvantage of women and URiM residents. Differences in ITE scores over the course of training may reflect inequities in the learning environment and lead to further disparities including being identified as in need of remediation. Given the importance of equitable learning environments in medical education, further study is needed to explore disparities in ITE performance during training and the impact on learners.

Corresponding Author: *Robin Klein, MD, MEHP; Department of Medicine, Division of General Internal Medicine and Geriatrics, Emory University School of Medicine, 49 Jesse Hill Jr Dr, Atlanta, GA 30303, USA (e-mail: Rklein3@emory.edu).*

Declarations:

Conflict of Interest: *The authors report no financial conflicts of interest.*

REFERENCES

1. Axelson RD, Solow CM, Ferguson KJ, Cohen MB. Assessing implicit gender bias in medical student performance evaluations. Evaluation & the health professions. 2010 Sep;33(3):365-85.
2. 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 Aug 9;12(8): e0181659.
3. Rojek AE, Khanna R, Yim JW, Gardner R, Lisker S, Hauer KE, Lucey C, Sarkar U. Differences in narrative language in evaluations of medical students by gender and under-represented minority status. Journal of general internal medicine. 2019 May 15;34(5):684-91.
4. Gerull KM, Loe M, Seiler K, McAllister J, Salles A. Assessing gender bias in qualitative evaluations of surgical residents. The American Journal of Surgery. 2019 Feb 1;217(2):306-13.
5. Brewer A, Osborne M, Mueller AS, O'Connor DM, Dayal A, Arora VM. Who Gets the Benefit of the Doubt? Performance Evaluations, Medical Errors, and the Production of Gender Inequality in Emergency Medical Education. American Sociological Review. 2020;85(2):247-270.
6. 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 May 1;177(5):651-657.
7. Mueller AS, Jenkins TM, Osborne M, Dayal A, O'Connor DM, Arora VM. Gender Differences in Attending Physicians' Feedback to Residents: A Qualitative Analysis. J Grad Med Educ. 2017 Oct;9(5):577-585.

8. Klein R, Ufere NN, Rao SR, Koch J, Volerman A, Snyder ED, Schaeffer S, Thompson V, Warner AS, Julian KA, Palamara K. Association of Gender with Learner Assessment in Graduate Medical Education. JAMA network open. 2020 Jul 1;3(7): e2010888.
9. Klein R, Ufere NN, Rao SR, Koch J, Volerman A, Snyder ED, Schaeffer S, Thompson V, Warner AS, Julian KA, Palamara K. Association of Race with Learner Assessment in Graduate Medical Education. Acad Med (in press).
10. American College of Physicians. Internal Medicine-In Training Examination. Available at https://www.acponline.org/featured-products/medical-educator-resources/im-ite. Accessed Jan 2021.
11. Garibaldi RA, Trontell MC, Waxman H, Holbrook JH, Kanya DT, Khoshbin S, Thompson J, Casey M, Subhiyah RG, Davidoff F. The in-training examination in internal medicine. Annals of internal medicine. 1994 Jul 15;121(2):117-23.
12. Hauer KE, Jurich D, Vandergrift J, Lipner RS, McDonald FS, Yamazaki K, Chick D, McAllister K, Holmboe ES. Gender Differences in Milestone Ratings and Medical Knowledge Examination Scores Among Internal Medicine Residents. Academic Medicine. 2021 May 25;96(6):876-84.
13. Waxman H, Braunstein G, Dantzker D, Goldberg S, Lefrak S, Lichstein E, Ratzan K, Schiffman F. Performance on the internal medicine second-year residency in-training examination predicts the outcome of the ABIM certifying examination. J Gen Intern Med. 1994 Dec;9(12):692-4.
14. 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(8):735–40.
15. McDonald FS, Zeger SL, Kolars JC. Associations between United States medical licensing examination (USMLE) and internal medicine in-training examination (IM-ITE) scores. Journal of general internal medicine. 2008 Jul 1;23(7):1016-9.
16. Rayamajhi S, Dhakal P, Wang L, Rai MP, Shrotriya S. Do USMLE steps, and ITE score predict the American Board of Internal Medicine Certifying Exam results? BMC Medical Education. 2020 Dec;20(1):1-8.
17. Grossman RS, Fincher RM, Layne RD, Seelig CB, Berkowitz LR, Levine MA. Validity of the in-training examination for predicting American Board of Internal Medicine certifying examination scores. Journal of general internal medicine. 1992 Jan 1;7(1):63-7.
18. Sisson SD, Casagrande SS, Dalal D, Yeh HC. Associations between quality indicators of internal medicine residency training programs. BMC medical education. 2011 Dec;11(1):1-6.
19. Kay C, Jackson JL, Frank M. The relationship between internal medicine residency graduate performance on the ABIM certifying examination, yearly in-service training examinations, and the USMLE Step 1 examination. Academic Medicine. 2015 Jan 1;90(1):100-4.
20. McDonald FS, Jurich D, Duhigg LM, Paniagua M, Chick D, Wells M, Williams A, Alguire P. Correlations Between the USMLE Step Examinations, American College of Physicians In-Training Examination, and ABIM Internal Medicine Certification Examination. Acad Med. 2020 Sep;95(9):1388-1395.
21. Willett LL, Halvorsen AJ, Adams M, Chacko KM, Chaudhry S, McDonald FS, Oxentenko AS, Swenson SL, Zaas A, Arora VM. Factors associated with declining residency program pass rates on the ABIM certification examination. The American journal of medicine. 2016 Jul 1;129(7):759-65.
22. Association of American Medical Colleges. ERAS for Medical Schools. Available at https://www.aamc.org/services/eras-for-institutions/medical-schools. Accessed July 2020.
23. Association of American Medical Colleges. Underrepresented in Medicine Definition. Available at https://www.aamc.org/what-we-do/diversity-inclusion/underrepresented-in-medicine. Accessed July 2020.
24. Foster N, Price M, Bettger JP, Goodwin CR, Erickson M. Objective Test Scores Throughout Orthopedic Surgery Residency Suggest Disparities in Training Experience. Journal of Surgical Education. 2021 Jan 14.
25. Frey, B. (2018). The SAGE encyclopedia of educational research, measurement, and evaluation (Vols. 1-4). Thousand Oaks,, CA: SAGE Publications, Inc.
26. Hill KA, Samuels EA, Gross CP, Desai MM, Zelin NS, Latimore D, Huot SJ, Cramer LD, Wong AH, Boatright D. Assessment of the prevalence of medical student mistreatment by sex, race/ethnicity, and sexual orientation. JAMA internal medicine. 2020 May 1;180(5):653-65.
27. Hu YY, Ellis RJ, Hewitt DB, Yang AD, Cheung EO, Moskowitz JT, Potts III JR, Buyske J, Hoyt DB, Nasca TJ, Bilimoria KY. Discrimination, abuse, harassment, and burnout in surgical residency training. New England Journal of Medicine. 2019 Oct 31;381(18):1741-52.
28. Anderson N, Lett E, Asabor EN, Hernandez AL, Nguemeni Tiako MJ, Johnson C, Montenegro RE, Rizzo TM, Latimore D, Nunez-Smith M, Boatright D. The Association of Microaggressions with Depressive Symptoms and Institutional Satisfaction Among a National Cohort of Medical Students. J Gen Intern Med. 2021 Apr 30. https://doi.org/10.1007/s11606-021-06786-6. Epub ahead of print.
29. Berwick S, Calev H, Matthews A, Mukhopadhyay A, Poole B, Talan J, Hayes MM, Smith CC. Mistaken identity: Frequency and effects of gender-based professional misidentification of resident physicians. Academic Medicine. 2021;96(6): 869-75.
30. Myers SP, Hill KA, Nicholson KJ, Neal MD, Hamm ME, Switzer GE, Hausmann LR, Hamad GG, Rosengart MR, Littleton EB. A qualitative study of gender differences in the experiences of general surgery trainees. Journal of Surgical Research. 2018; 228:127-34.
31. Osseo-Asare A, Balasuriya L, Huot SJ, Keene D, Berg D, Nunez-Smith M, Genao I, Latimore D, Boatright D. Minority resident physicians' views on the role of race/ethnicity in their training experiences in the workplace. JAMA network open. 2018 Sep 7;1(5): e182723-.
32. 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. Academic Medicine. 2020 Nov 1;95(11S): S58-66.
33. Wayne SJ, Fortner SA, Kitzes JA, Timm C, Kalishman S. Cause, or effect? The relationship between student perception of the medical school learning environment and academic performance on USMLE Step 1. Medical teacher. 2013 May 1;35(5):376-80.
34. Zurbrügg L, Miner KN. Gender, sexual orientation, and workplace incivility: Who is most targeted and who is most harmed? Frontiers in psychology. 2016 May 2; 7:565.
35. Sue DW. Microaggressions in everyday life: Race, gender, and sexual orientation. John Wiley & Sons; 2010 Feb 9.
36. Bowman NA. College diversity experiences and cognitive development: A meta-analysis. Review of Educational Research. 2010 Mar;80(1):4-33.
37. Hurtado S. Linking Diversity and Educational Purpose: How Diversity Affects the Classroom Environment and Student Development.
38. Laird TF. College students' experiences with diversity and their effects on academic self-confidence, social agency, and disposition toward critical thinking. Research in higher education. 2005 Jun;46(4):365-87.
39. van Ryn M, Hardeman R, Phelan SM, Burgess DJ, Dovidio JF, Herrin J, Burke SE, Nelson DB, Perry S, Yeazel M, Przedworski JM. Medical School Experiences Associated with Change in Implicit Racial Bias Among 3547 Students: A Medical Student CHANGES Study Report. J Gen Intern Med. 2015 Dec;30(12):1748-56. https://doi.org/10.1007/s11606-015-3447-7. Epub 2015 Jul 1.
40. Shokar GS. The effects of an educational intervention for the "at-risk" residents to improve their scores on the in-training exam. Fam Med. 2003; 35(6):414–417.
41. Kim RH, Tan TW. Interventions that affect resident performance on the American Board of Surgery In-Training Examination: a systematic review. J Surg Educ. 2015; 7(3):418–429.
42. Kosir MA, Fuller L, Tyburski J, Berant L, Yu M. The Kolb learning cycle in American Board of Surgery In-Training Exam remediation: the Accelerated Clinical Education in Surgery course. Am J Surg. 2008;196(5):657–662.
43. Borman KR. Does academic intervention impact ABS qualifying examination results? Curr Surg. 2006;63(6):367–372.
44. Harthun NL, Schirmer BD, Sanfey H. Remediation of low ABSITE scores. Curr Surg. 2005;62(5):539–542.
45. Ferrell BT, Tankersley WE, Morris CD. Using an Accountability Program to Improve Psychiatry Resident Scores on In-Service Examinations. Journal of graduate medical education. 2015; 7(4), 555–559.
46. Krzyzaniak SM, Kaplan B, Lucas D, Bradley E, Wolf SJ. Unheard Voices: A Qualitative Study of Resident Perspectives on Remediation. Journal of Graduate Medical Education. 2021 Aug;13(4):507-14.

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