Title

A Randomized Study to Assess the Effect of Including the Graduate Record Examinations (GRE) Results on Reviewers Scores for Underrepresented Minorities

Journal

American Journal of Epidemiology, 190(9)

ISSN

0002-9262

Authors

Dang, Kristina V
Rerolle, Francois
Ackley, Sarah F
et al.

Publication Date

2021-09-01

DOI

10.1093/aje/kwab075

Peer reviewed

American Journal of Epidemiology

Vol. 190, No. 9
https://doi.org/10.1093/aje/kwab075
Advance Access publication:
March 18, 2021

Original Contribution

A Randomized Study to Assess the Effect of Including the Graduate Record Examinations Results on Reviewer Scores for Underrepresented Minorities

Kristina V. Dang, Francois Rerolle, Sarah F. Ackley, Amanda M. Irish, Kala M. Mehta, Inez Bailey, Elizabeth Fair, Cecily Miller, Kirsten Bibbins-Domingo, Eva Wong-Moy, M. Maria Glymour, and Meghan D. Morris*

* Correspondence to Dr. Meghan D. Morris, Department of Epidemiology and Biostatistics and Institute for Global Health Sciences, School of Medicine, University of California San Francisco, 550 16th Street, San Francisco, CA 94158 (e-mail: meghan.morris@ucsf.edu).

Initially submitted August 18, 2020; accepted for publication March 15, 2021.

Whether requiring Graduate Record Examinations (GRE) results for doctoral applicants affects the diversity of admitted cohorts remains uncertain. This study randomized applications to 2 population-health doctoral programs at the University of California San Francisco to assess whether masking reviewers to applicant GRE results differentially affects reviewers' scores for underrepresented minority (URM) applicants from 2018–2020. Applications with GRE results and those without were randomly assigned to reviewers to designate scores for each copy (1–10, 1 being best). URM was defined as self-identification as African American/Black, Filipino, Hmong, Vietnamese, Hispanic/Latinx, Native American/Alaska Native, or Native Hawaiian/Other Pacific Islander. We used linear mixed models with random effects for the applicant and fixed effects for each reviewer to evaluate the effect of masking the GRE results on the overall application score and whether this effect differed by URM status. Reviewer scores did not significantly differ for unmasked versus masked applications among non-URM applicants ($\beta = 0.15$; 95% CI: −0.03, 0.33) or URM applicants ($\beta = 0.02$, 95% CI: −0.49, 0.54). We did not find evidence that removing GREs differentially affected URM compared with non-URM students (β for interaction = −0.13, 95% CI: −0.55, 0.29). Within these doctoral programs, results indicate that GRE scores neither harm nor help URM applicants.

diversity; education; graduate admissions; Graduate Record Examinations (GRE); randomized study; underrepresented minority (URM)

Abbreviations: ETS, Epidemiology and Translational Science Program; GPA, grade point average; GRE, Graduate Record Examinations; SD, standard deviation; UCSF, University of California San Francisco; URM, underrepresented minority.

Diversity in higher education benefits individual students, institutions, and society, yet remains an unachieved goal (1–3). Currently there is concern that requiring standardized tests as part of graduate school admissions requirements might create barriers for underrepresented minority (URM) applicants, and many graduate programs are considering the elimination of the Graduate Record Examinations (GRE) General Test from their application requirements (4–6). A major motivation for this is the concern that requiring the GRE disproportionately harms applicants from URM groups, leading to lower admission rates for URMs (7, 8). The premise that requiring GRE scores for doctoral (PhD) program applications differentially affects applicants from URM groups has not been rigorously evaluated, and there are potential benefits to the GRE score that might offset these perceived harms. Applicants from URM groups, on average, report lower grade point averages (GPAs) than those from non-URM groups (9, 10), and test scores might provide additional information about the applicants for the reviewer. Many graduate programs embrace holistic review practices to evaluate the whole applicant, not only empirical data like GPA or standardized test scores (11), and the

lower average scores on the GRE might be counterbalanced by other considerations when evaluating applications from URM candidates (12).

Increasing the diversity of the biomedical research workforce is a high national priority (13). If requiring the GREs is an impediment to efforts to recruit and enroll URM scientists, this provides a strong argument for dropping the GRE requirement. Dropping this indicator might, however, harm URM candidate evaluations and have an impact on overall admission diversity. In the absence of GRE scores, reviewers might use other criteria on which URM candidates are even more disadvantaged, such as prestige of the undergraduate institution or letters of recommendation elicited from unpaid internships with prestigious researchers. Substantial evidence shows that implicit bias against racial/ethnic minorities is common (14, 15), and these implicit biases might be most relevant when objective information is not available. Removing the GRE might thus increase the adverse effect of implicit racism on graduate applicant decisions.

Given these competing theoretical possibilities, it is imperative to rigorously analyze the impact of eliminating the GRE from graduate applications. We assessed this in the setting of 2 doctoral programs (epidemiology and global health) by randomly assigning reviewers to evaluate applications with or without GRE results.

METHODS

We examined the effect of GRE results on reviewer scores by randomizing applications that included GRE results (unmasked applications) and applications with GRE results removed (masked applications) to graduate application reviewers. The study was conducted at the University of California San Francisco (UCSF) in the Department of Epidemiology and Biostatistics, from December 2018 until February 2020 (spanning 2 application cycles). Applications were from 2 doctoral programs in population health science. The Global Health Sciences program, a social and population sciences program, accepts applications every 2 years and contributed 1 year of applications for this study (2018/2019 cycle). The Epidemiology and Translational Science (ETS) program, a basic and biomedical sciences program, accepts applications annually. More details on the programs are provided in Web Appendix 1 and Web Tables 1 and 2. This study used data from completed applications for the ETS doctoral program in 2018/2019 ($n = 75$) and 2019/2020 ($n = 87$) and the Global Health Sciences doctoral program in 2018/2019 ($n = 36$).

Complete applications submitted to Global Health Sciences or ETS during the 2018/2019 cycle or to the ETS program during the 2019/2020 cycle were eligible for review. Complete applications included the following information: 1) demographic factors: birth city/country, age, sex, gender, sexual orientation, citizenship, race/ethnicity, highest level of each parent's education, disability, disadvantaged background, California high school attendance, historically black college attendance, China Scholarship Council participant, UCSF Summer Research Programs participant, military service, and medically underserved community resident; 2) academic training: bachelor's and graduate institution major and dates of attendance, GPA, and grades; 3) test scores: test date and test score as percentile for GRE (quantitative reasoning, verbal reasoning, analytical writing), Test of English as a Foreign Language (TOEFL), Medical College Admission Test (MCAT), or Dental Admission Test (DAT), as applicable (scores were self-reported and validated with the Educational Testing Service); 4) applicant profile: personal history statement, research experience summary, research interests, publications and presentations, resume/curriculum vitae, transcripts; 5) letters of recommendation: letters of recommendation along with response data evaluating the applicant's capacity for independent thinking, research potential, interpersonal interactions, maturity, and overall rating measured as top 1%, 5%, 10%, 25%, or 50% or <50%. Applications were submitted through an online electronic platform managed by UCSF's graduate affairs office.

Every application was reviewed and scored independently by 4 randomly assigned admissions committee members. Two reviewers read and scored the unmasked application and 2 different reviewers read and scored the masked application (Figure 1). Masked applications had redacted GRE results from the test score section and anywhere else in the application test scores were referenced (e.g., letters of recommendation, personal statements). Reviews completed per reviewer varied depending on their program and availability during each application cycle. Reviewers scored a minimum of 36 applications (19 unmasked and 17 masked) and a maximum of 113 applications (51 unmasked and 62 masked). Each reviewer received separate secure digital folders; one with unmasked applications (all GRE results and references included) and the other with masked applications (redacted GRE results) from different applicants. To reduce possible bias due to reviewer fatigue, half of the reviewers were randomized to read and score applications in the GRE-unmasked folder first, and the other half of reviewers to read and score applications in the GRE-masked folder first.

Our primary outcome was the reviewer's overall application score, which ranged from 1 (most favorable) to 10 (least favorable). Secondary outcomes were reviewer scores in 5 specific domains: 1) research experience, 2) academic training, 3) letters of recommendation, 4) level of UCSF support, and 5) research potential (Web Figure 1, available at https://doi.org/10.1093/aje/kwab075). The reviewer's overall score is the primary metric used when selecting applications for interview, where applicants are further assessed for program selection. Further, the overall score is the most proximal and likely most sensitive to reviewers evaluating the GRE in our application process.

Primary analyses included 3 kinds of variables: 1) a binary variable for unmasked versus masked applications; 2) a binary variable for whether the application was from a member of an URM status group; and 3) fixed effects representing the 14 different reviewers. URM group identification was based on self-reported information from 2 open-field formatted questions ("Racial category" and "Describe your background") and was defined following the UCSF definition (African American/Black, Hispanic/Latinx, Native American/Alaska Native, Native Hawaiian/Other Pacific Islander,

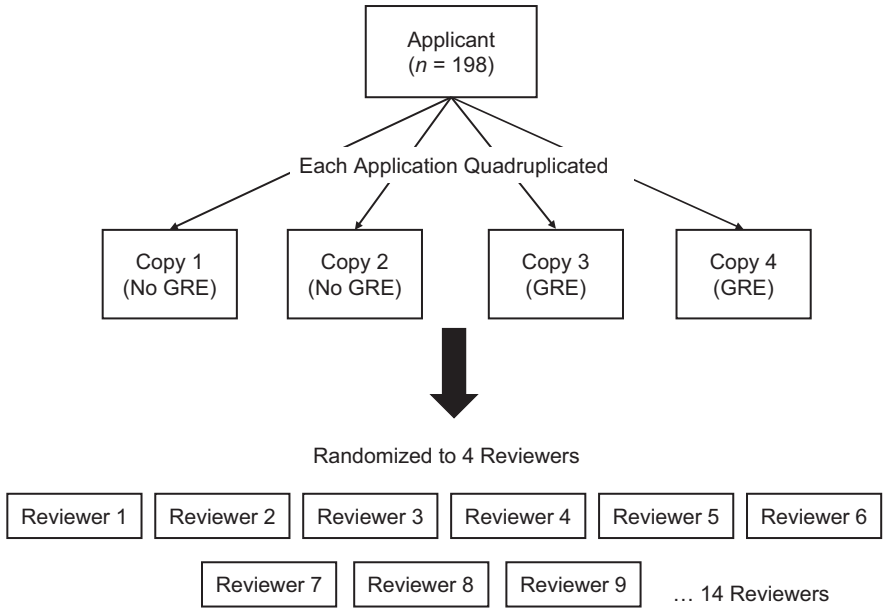

Figure 1. Data collection flow diagram for an assessment of the impact of including Graduate Record Examinations (GRE) scores in applications to doctoral programs, California, 2018–2020. We extracted an overall score per copy (4 scores per applicant), 2 scores with GRE masked and 2 scores with GRE unmasked.

Asian: Filipino, Hmong, Vietnamese, or multiple categories including at least one of the above) (16).

We used linear mixed models to evaluate the effect on reviewer overall scores of unmasked versus masked applications and whether this effect was modified by URM status. The linear mixed model included fixed effects for GRE masking, URM status, and reviewers and random effects for each applicant because there were 4 evaluation scores for each application. In sensitivity analyses, we included undergraduate GPA as a covariate in the models, to evaluate the possibility that the effect of viewing the GRE was modified by undergraduate GPA.

Upon review by the UCSF Institutional Review Board, this research was determined to be exempt from human subjects review (approval number 19-27197). All identifiable data are stored on a password-protected computer in the possession of the principal investigator.

RESULTS

Descriptive characteristics of the 198 applications in the sample (800 total reviews due to some reviewers not evaluating all assigned applications and 2 applications being reviewed by all reviewers from ETS in the 2018/2019 cycle) are presented in Table 1. Overall, there were 159 (80%) non-URM applicants, and 39 (20%) URM applicants. Average GRE percentiles and parental education were higher for non-URM applicants than URM applicants, but undergraduate and master's GPAs were similar.

The reviewer pool comprised 14 individuals who served on one or more of the admissions committees, including 8 (57%) full professors, 2 (14%) associate professors, 1 (7%) assistant professor, and 3 (21%) graduate affairs specialists. All admissions committee members had been employed by UCSF for at least 5 years. Of the 14 reviewers, 9 (64%) were female and 5 (36%) were male; 9 (64%) identified as White, 2 (14%) identified as Asian, and 3 (21%) identified as Black or biracial. On average, committee members had previously served for 3.6 (standard deviation (SD), 2.3) years on their committees. The applicant random effect accounted for 39% of the variance in reviewers' scores.

The average overall score was similar for URM applications (3.72; SD, 1.57) and non-URM applications (3.70; SD, 1.94). For GRE-masked applications the average overall score was 3.63 (SD, 1.84) and for GRE-unmasked applications, 3.77 (SD, 1.94). Among URM applications, when the GRE result was unmasked, reviewers scored applications 0.02 points worse than applications where the GRE result was masked ($\beta = 0.02$, 95% confidence interval (CI): −0.36, 0.40). Among non-URM applications, when unmasking the GRE result, reviewers scored applications 0.15 points worse than when the GRE results were masked ($\beta = -0.15$; 95% CI: −0.68, 0.37). When we assessed the interaction between GRE unmasking and URM status there was little evidence that unmasking the GRE result differentially affected applications for URMs (P for interaction of URM status and GRE unmasking = 0.56; β for interaction = −0.13, 95% CI: −0.55, 0.29, Table 2). This association was close to the null with confidence intervals including both small harms to URM applicants and moderate advantages to URM applicants.

The direction and magnitude of effect estimates with each of our secondary outcomes (research experience, academic

Table 1. Characteristics of Applicants to Doctoral Programs for Epidemiology and Translational Science and for Global Health Sciences, University of California, San Francisco, 2018/2019 and 2019/2020 Cycles

Characteristic	Total (*n* = 198)	%	Mean (SD)	URM (*n* = 39)	%	Mean (SD)	Non-URM (*n* = 159)	%	Mean (SD)
Age, years			27.9 (5.4)			29.7 (6.5)			28.6 (5.1)
Birth sex female	156	78.8		26	66.7		130	81.8	
GPA[a], bachelor's degree			3.5 (0.5)			3.3 (0.4)			3.5 (0.5)
GPA, master's degree			3.8 (0.3)			3.8 (0.3)			3.8 (0.2)
Foreign born	117	59.1		19	48.7		98	61.6	
US citizen	102	51.5		25	64.1		77	48.4	
Hispanic	20	10.1		20	51.3		0	0.0	
Race									
White	61	30.8		7	17.9		54	34.0	
Black	15	7.6		15	38.5		0	0.0	
Asian	93	47.0		2	5.1		91	57.2	
Other	29	14.6		15	38.5		14	8.8	
Highest parental education									
Some high school	7	3.5		3	7.7		4	2.5	
High school diploma or GED	17	8.6		5	12.8		12	7.5	
Some college	16	8.1		4	10.3		12	7.5	
Bachelor's degree	58	29.3		12	30.8		46	28.9	
At least some graduate school	88	44.4		13	33.3		75	47.2	
Did not report	12	6.1		2	5.1		10	6.3	
Disadvantaged background	41	20.7		18	46.2		23	14.5	
GRE self-reported percentiles									
Verbal reasoning			69.4 (22.6)			62.3 (23.8)			71.1 (22.1)
Analytical writing			60.5 (25.8)			53.6 (25.8)			62.1 (25.6)
Quantitative reasoning			68.4 (24.6)			43.7 (20.5)			74.1 (22.0)

Abbreviations: GED, General Educational Development; GPA, grade point average; GRE, Graduate Record Examinations; SD, standard deviation; URM, underrepresented minority.

[a] GPA was not scaled or adjusted, and no distinction was made for schools that scored on a 4-point versus a 5-point scale.

Table 2. Mixed Model Estimates for the Effect of Graduate Record Examinations Status and Underrepresented Minority Status on Overall Score for Applicants to Doctoral Programs for Epidemiology and Translational Science and for Global Health Sciences, University of California, San Francisco, 2018/2019 and 2019/2020 Cycles

Covariate[a]	Mean Difference	95% CI
Intercept	4.11	3.79, 4.42
GRE unmasked	0.15	−0.03, 0.33
URM	−0.02	−0.54, 0.49
GRE unmasked × URM	−0.13	−0.55, 0.29

Abbreviations: CI, confidence interval; GRE, Graduate Record Examinations; URM, underrepresented minority.
[a] Model adjusted for reviewer (fixed effects) with random effects for applicant.

training, letter of recommendation, level of UCSF support, and epidemiologic research potential, found in Web Tables 3–7) was similar to those presented in Table 2.

In sensitivity analyses with undergraduate GPA as a predictor (Web Table 8), the effect of undergraduate GPA was not statistically significant, nor did its addition to the linear mixed model in Table 2 change the results (statistically or substantively) of the other predictors.

DISCUSSION

This randomized study evaluated whether including GRE results in applications to a doctoral program in population health science disadvantaged URM candidates compared with non-URM candidates. Little evidence was found that supports a differential effect of GRE score inclusion, either harm or benefit, to URM applicants. Unmasking the GRE resulted in slightly worse average scores for non-URM applications and an even smaller decrement for URM applicants. The net result of unmasking was a small advantage to URMs of −0.13 (or 7% of a standard deviation). Our findings indicate that the use of GRE most likely had little to no effect on URM overall scores, although with our sample size, we could not rule out small harms.

Our randomized design provides much stronger evidence than previously available to understand the effect of GRE on URMs' admissions scores (17–21). By randomly assigning applications to include versus exclude the GRE and having multiple GRE-masked and GRE-unmasked reviews for each application, we were able to estimate the effect of viewing the GRE on admissions evaluations. This study design compared each applicant with themselves, with the difference being GRE status. Additionally, we are able to explicitly test the effect of viewing the GRE on overall score by URM applicants. Although URM applicants averaged lower GRE scores, reviewers apparently considered other aspects of URM candidate applications to outweigh the worse GRE scores, because the association of GRE percentiles (a decile increase) on reviewers' scores was less than 0.29 points, the upper bound for the main effect of interest (data not shown). Reviewers were encouraged to adopt holistic review practices and provided an evaluation framework that considered an applicant's experience in addition to test scores and GPA, which might have reduced the impact of lower GRE scores on final evaluations. We found that in the context of holistic review, large adverse effects to URMs were unlikely to result from removing the GRE in our graduate admissions process. Given the limited access to testing centers due to the COVID-19 pandemic, we hope our findings will be useful to graduate programs as they consider how to handle potentially large missingness of GRE test scores as part of the larger admissions process. We plan to examine the effects of COVID-19 on graduate admissions in a follow-up qualitative study. Further, the review committee's mix of race/ethnicity, tenure, and influence on the admissions process confers a perspective that might reduce the reliance on GRE scores (22). While we observed reviewer effects that were adjusted for in our analysis using fixed effects, the study was not set up to examine reviewer differences in GRE and score by URM. Reviewer scores from initial application review are a major but not the only consideration in selecting candidates for interviews or program admission. Interview invitations are also influenced by considerations unrelated to the applicant, such as mentors' availability. We focused here on the outcome we considered most likely to be detectably influenced by GRE masking (reviewer scores), but future research assessing more distal outcomes is likely to be important. In particular, if final decision-making about admissions is influenced by subjective considerations, masking the GRE might have an important impact on the final decisions.

Conversely, the study has some notable limitations. It is based in a single public university, so external generalizability should be considered when applying our results to another program. Because several graduate programs at UCSF had already dropped the GRE requirement for the 2018–2019 application cycle, and many others had declared the GRE requirement optional, these programs could not help us answer the posited hypotheses. We therefore included graduate programs in 2 population health research that had not yet dropped the GRE requirement, for a total of 198 unique applications (applicants are allowed to apply to only one UCSF graduate program) over 2 application cycles (2018/2019 and 2019/2020). While our sample size is modest, it delivered an informative confidence interval, with a point estimate close to the null: The upper bound of our confidence interval—0.29 or approximately

15% of a standard deviation in reviewer scores—indicated that if there is any differential adverse effect of using the GRE on URM applicants it is likely to be small. Our sample includes 2 relatively small programs, with a modest number of individual reviewers, and might not generalize to larger programs or different disciplines. Although the upper bound of our confidence interval indicates that a large adverse effect of using the GRE is unlikely, more evidence on this question from larger programs with heterogeneous characteristics with respect to faculty expertise, funding structures, and training priorities would be valuable. Further, the applicant random effect accounted for 39% and the reviewer accounted for 37% of the variance in reviewer scores; a sizeable amount of variance is still unaccounted for, which might be due to random noise or could reflect that different reviewers prioritize different characteristics in applications. Given the current discussion of GRE, reviewers might have evaluated GRE scores differently during these 2 review cycles than they would have in a period when GREs were receiving less scrutiny. Additionally, we cannot rule out the possibility that reviewers used other information about race/ethnicity to inform their decisions in ways that are not generalizable.

The debate about inclusion of GRE scores in graduate student applications has centered on 2 claims: that the GRE differentially disadvantages URM applicants in the review process, reducing diversity in accepted cohorts, and that the GRE does not predict outcomes among admitted students (17–20, 23). Prior evaluations of the GRE as a predictor of admitted student outcomes analyze only those students who were admitted to their graduate programs, creating the potential for selection bias. In particular, without the GRE, application review will depend on a range of other factors (e.g., undergraduate institutions, letters of recommendation, prior research experiences), and the ways in which those components are evaluated is likely to vary by discipline and evolve over time. Thus, in a contemporary context, where diversity is a broadly accepted goal of biomedical research training and population health programs, and academia at large, evidence on the impact of the GRE is needed to guide decision making.

In this randomized study, we found little statistical evidence that including GRE scores in admissions decisions affects URM applicants when applying to our graduate programs. Although this study has a small sample, the confidence interval suggests large harms are unlikely and many other factors had much greater effects on outcomes.

ACKNOWLEDGMENTS

Author affiliations: Department of Epidemiology and Biostatistics, School of Medicine, University of California, San Francisco, San Francisco, California, United States (Kristina V. Dang, Francois Rerolle, Sarah F. Ackley, Amanda M. Irish, Kala M. Mehta, Inez Bailey, Cecily Miller, Kirsten Bibbins-Domingo, Eva Wong-Moy, M. Maria Glymour, Meghan D. Morris); and Institute for Global Health Sciences, University of California, San Francisco, San Francisco, California, United States (Inez Bailey, Elizabeth Fair, Cecily Miller, and Meghan D. Morris).

This work was funded by the National Institutes of Health (grant 1UL1MD009608-01), UCSF Department of Epidemiology and Biostatistics core support fund, and UCSF School of Medicine John Watson Scholar Fund (M.D.M.).

We thank UCSF Graduate Division for feedback on the objective and design of the study by members of the Department of Epidemiology and Biostatistics Diversity Committee. Our study is one of many activities to dismantle exclusive policies that limit diversity in graduate training. To ensure that applicants of the 2018–2020 cycles were not harmed through the study, the admissions committee selected candidates for interview and made admissions decisions based on unmasked application data. We appreciate the countless hours our admission committee members provide each year reviewing applications and the larger Department of Epidemiology and Biostatistics members' commitment to training the next generation of public health researchers. Authors are members of the admissions committee and have a vested interest in the success of the program. A deidentified limited data set reflecting data analyzed during the study is available upon request to the corresponding author.

This work was presented at the Health Disparities Research Symposium, October 11, 2019, San Francisco, California, and the Society of Epidemiologic Research Conference, December 16–18, 2020, online.

Conflict of interest: none declared.

REFERENCES

1. Rudenstein NL. Student Diversity and Higher Learning. In: Orfield G, ed. *Diversity Challenged: Evidence on the Impact of Affirmative Action*. Cambridge, MA: Harvard Education Publishing Group; 2001:31–48.
2. Milem JF. The Educational Benefits of Diversity: Evidence from Multiple Sectors. In: Chang MJ, Witt D, Jones J, et al., eds. *Compelling Interest: Examining the Evidence on Racial Dynamics in Higher Education*. Stanford, CA: Stanford University Press; 2003:126–169.
3. Ward KM, Zarate ME. The influence of campus racial climate on graduate student attitudes about the benefits of diversity. *Rev High Educ*. 2015;38(4):589–617.
4. Langin K. A wave of graduate programs drops the GRE application requirement. *Science*; 2019. https://www.sciencemag.org/careers/2019/05/wave-graduate-programs-drop-gre-application-requirement. Accessed July 6, 2019.
5. Cahn PS. Do health professions graduate programs increase diversity by not requiring the graduate record examination for admission? *J Allied Health*. 2015;44(1):51–56.
6. University of California San Francisco. Graduate Division: GRE Announcement. https://graduate.ucsf.edu/gre-announcement. Accessed October 14, 2019.
7. Langin K. Ph.D. programs drop standardized exam. *Science*. 2019;364(6443):816.

8. Posselt JR. Toward inclusive excellence in graduate education: constructing merit and diversity in PhD admissions. *American Journal of Education*. 2014;120(4): 481–514.
9. Miller C, Stassun K. A test that fails. *Nature*. 2014; 510(7504):303–304.
10. Croizet J-C, Dutrevis M. Socioeconomic status and intelligence: why test scores do not equal merit. *J Poverty*. 2015;8(3):91–107.
11. Educational Testing Service. *A Snapshot of the Individuals Who Took the GRE Revised General Test August 2011–June 2014*. Princeton, NJ: Educational Testing Service; 2014. https://www.ets.org/s/gre/pdf/snapshot_test_taker_data_2014.pdf. Accessed December 19, 2019.
12. Bleske-Rechek A, Browne K. Trends in GRE scores and graduate enrollments by gender and ethnicity. *Dermatol Int*. 2014;46:25–34.
13. Department of Health and Human Services. *The Secretary's Advisory Committee on National Health Promotion and Disease Presention Objectives for 2020.* Washington, DC: Department of Health and Human Services; 2008. https://www.healthypeople.gov/2010/hp2020/advisory/PhaseI/PhaseI.pdf. Accessed August 16, 2020.
14. Walton GM, Spencer SJ. Latent ability grades and test scores systematically underestimate the intellectual ability of negatively stereotyped students. *Psychol Sci*. 2009;20(9): 1132–1139.
15. Godsil RD, Tropp LR, Goff PA, et al. *Addressing Implicit Bias, Racial Anxiety and Stereotype Threat in Education and Healthcare.* Washington, DC: Perception Institute; 2014. https://equity.ucla.edu/wp-content/uploads/2019/12/Science-of-Equality-Vol.-1-Perception-Institute-2014.pdf. Accessed August 3, 2020.
16. University of California San Francisco. URM definition. https://diversity.ucsf.edu/URM-definition. Accessed July 14, 2020.
17. Weiner OD. How should we be selecting our graduate students? *Mol Biol Cell*. 2014;25(4):429–430.
18. Hall JD, O'Connell AB, Cook JG. Predictors of student productivity in biomedical graduate school applications. *PLoS One*. 2017;12(1):1–14.
19. Sealy L, Saunders C, Blume J, et al. The GRE over the entire range of scores lacks predictive ability for PhD outcomes in the biomedical sciences. *PLoS One*. 2019;14(3): 1–17.
20. Moneta-Koehler L, Brown AM, Petrie KA, et al. The limitations of the GRE in predicting success in biomedical graduate school. *PLoS One*. 2017;12(1):1–17.
21. Attiyeh G, Attiyeh R. Testing for bias in graduate school admissions. *J Hum Resour*. 1997;32(3):524–548.
22. Squire DD. "It's pretty essential": a critical race counter-narrative of faculty of color understandings of diversity and equity in doctoral admissions. *The Urban Review*. 2020;S2:173–197.
23. Petersen SL, Erenrich ES, Levine DL, et al. Multi-institutional study of GRE scores as predictors of STEM PhD degree completion: GRE gets a low mark. *PLoS One*. 2018;13(10):e0206570.