

A Model for Holistic Review in Graduate Admissions That Decouples the GRE from Race, Ethnicity, and Gender

Marenda A. Wilson,^{†*} Max A. Odem,[†] Taylor Walters,[§] Anthony L. DePass,^{||} and Andrew J. Bean^{†**}

[†]Deans' Office, The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences, Houston, TX 77030; [‡]Graduate College, Rush University, Chicago, IL 60612; [§]College of Arts and Sciences, Oberlin College and Conservatory, Oberlin, OH 44074; ^{||}Department of Biology, Long Island University, Brooklyn, NY 11201; ^{**}Department of Neurobiology and Anatomy, Cell Biology and Biochemistry, McGovern Medical School at The University of Texas Health Science Center at Houston, Houston, TX 77030; [#]Department of Pediatrics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030

ABSTRACT

Graduate schools around the United States are working to improve access to science, technology, engineering, and mathematics (STEM) in a manner that reflects local and national demographics. The admissions process has been the focus of examination, as it is a potential bottleneck for entry into STEM. Standardized tests are widely used as part of the decision-making process; thus, we examined the Graduate Record Examination (GRE) in two models of applicant review: metrics-based applicant review and holistic applicant review to understand whether it affected applicant demographics at The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences. We measured the relationship between GRE scores of doctoral applicants and admissions committee scores. Metrics-based review of applicants excluded twice the number of applicants who identified as a historically underrepresented minority compared with their peers. Efforts to implement holistic applicant review resulted in an unexpected result: the GRE could be used as a tool in a manner that did not reflect its reported bias. Applicant assessments in our holistic review process were independent of gender, racial, and citizenship status. Importantly, our recommendations provide a blueprint for institutions that want to implement a data-driven approach to assess applicants in a manner that uses the GRE as part of the review process.

INTRODUCTION

In 2016, more than 1.8 million students were enrolled in certificate, master's, and doctoral graduate programs, with an annual applicant pool of ~2.2 million (Okahana and Zhou, 2017). For many of these programs, prospective students participate in the admissions process by submitting a package of information designed to allow faculty to assess each applicant's potential for success in their respective graduate programs. A typical package contains an application form that summarizes personal, demographic, and prior training information; supplemental materials; institutional verification; and normalizing data. Materials include academic transcripts, standardized test scores, curricula vitae, letters of recommendation, and essays that generally outline applicants' previous experience and describe why they want to attend graduate school. On the basis of the information provided in the application package, faculty are tasked with assessing each applicant's qualifications, suitability, and potential to succeed in a doctoral program. When successful, the graduate admissions process identifies qualified students who are admitted to programs and subsequently complete the intended degree(s). Thus, graduates contribute to science and enter the science, technology,

David Feldon, *Monitoring Editor*

Submitted Jun 26, 2018; Revised Nov 26, 2018; Accepted Nov 28, 2018

CBE Life Sci Educ March 1, 2019 18:ar7

DOI:10.1187/cbe.18-06-0103

Author contributions: M.A.W., acquisition of data, analysis and interpretation of data, and drafting and revision of the article; M.A.O., interpretation of data and review of the article; T.W., analysis of data and drafting of the article; A.L.D., analysis and interpretation of data and review of article; A.J.B., analysis and interpretation of data and drafting and revision of the article.

*Address correspondence to: Marenda A. Wilson (Marenda.Wilson-Pham@rush.edu).

© 2019 M. A. Wilson *et al.* CBE—Life Sciences Education © 2019 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

"ASCB®" and "The American Society for Cell Biology®" are registered trademarks of The American Society for Cell Biology.

engineering, and mathematics (STEM) workforce, thereby supporting university missions and the vitality of graduate programs.

Over the past few decades, institutions of higher learning and federal agencies have worked together to improve access to opportunities and representation across STEM disciplines (Valantine and Collins, 2015; Mervis, 2016). However, the process of admissions in graduate education has been insulated from many of these sweeping changes. Specifically, the lack of resources and available time of a volunteer faculty workforce to review and discuss every applicant in a large applicant pool has resulted in a process that has been slow to adopt these changes. Additionally, even when changes have been made, they may be impacted by implicit biases of faculty reviewers in which there are shared views, values, and prejudices of the dominant culture in society—one that is middle-class, male, and overwhelmingly white (in 2016, of the 517,091 full-time professors, 73% were white, 5% were African American, 4% were Hispanic, and 0.4% were Pacific Islander or Native American; U.S. Department of Education, 2017). As a result, there is significant underrepresentation in the advanced training, workforce, and leadership ranks despite the investment in a STEM workforce that is representative of national demographics (underrepresented minorities [URMs] make up 30% of the population, 8.5% of doctoral students, 4% of postdoctoral fellows, 5% of principal investigators on research grants, and 13% of the workforce; U.S. Census Bureau, 2010; National Institutes of Health, 2012; National Center for Education Statistics, 2013; Gibbs *et al.*, 2014, 2016; National Science Foundation [NSF], 2017). Thus, a blueprint for best practices in applicant review and regular assessments of the effectiveness of admissions committees in graduate admissions is greatly needed.

The use of standardized tests as a means for normalization of applicants from various undergraduate institutions and training pathways has been under scrutiny (Grossbach and Kuncel, 2011; Roush *et al.*, 2014; Wilson *et al.*, 2014; Durning *et al.*, 2015; Pacheco *et al.*, 2015; Moneta-Koehler *et al.*, 2017; Park *et al.*, 2018). The Graduate Record Examination (GRE) is a widely used standardized test for application to master's and doctoral degree STEM programs. The test consists of three sections: Verbal Reasoning, Quantitative Reasoning, and Analytical Writing. The Verbal Reasoning section measures critical analysis and the recognition of associations between words and concepts. The Quantitative Reasoning section addresses the ability to solve complex problems using basic math and data analysis. The Analytical Writing section assesses the ability to present critical analysis. However, there is significant concern that GRE results, contrary to the recommendations of its developers, are often used as a mechanism to manage applications by setting cutoff scores to enable smaller applicant pools for committee review (Posselt, 2016).

The predictive validity of the GRE has been studied extensively, and its utility in graduate admissions is controversial (Kuncel *et al.*, 2001; Miller and Stassun, 2014; Moneta-Koehler *et al.*, 2017). Women and URMs on average score lower on the GRE than well-represented (white and Asian-American) men (Educational Testing Services [ETS], 2014). Thus, it has been argued that use of GRE scores in graduate admissions has contributed to the underrepresentation of multiple demographic groups in professions related to the STEM disciplines (Kuncel

et al., 2001; Miller and Stassun, 2014; Posselt, 2016). To its credit, the ETS, developers of the GRE, have cautioned against strict interpretation of GRE scores for URMs, because validity studies have used small sample sizes (ETS, 2011, 2014, 2015a,b). However, in practice, this admonition has been lost by many who use the GRE in the graduate admission process (Posselt, 2016).

Admissions committees receive hundreds to thousands of applications for review in a short application review cycle. Consequently, considerable emphasis is placed on quantitative measures, such as GRE scores, to manage the application review process (Kent and McCarthy, 2016; Posselt, 2016). Some committees triage applicants who fall below arbitrary score cutoffs (Posselt, 2016). Committees are often made aware of unconscious bias and are provided frameworks for ethical goals as they relate to merit, diversity, and potential of applicants in the admissions process. However, there is some degree of score bias when selecting applicants who are perceived as most qualified (Atwood *et al.*, 2011; Posselt, 2016). This may be the result of explicit and unconscious socialization during the training and academic careers of faculty, reflecting epistemology, language, behaviors, and attitudes as expectations in their functional roles at institutions (Clark, 1989; Stichweh, 1992; Becher and Trowler, 2001; Jacobs, 2013). Over time, this can shape internalized stereotypes and preferences about others and could ultimately influence how faculty interact with and view prospective students (Milkman *et al.*, 2015; Posselt, 2016). However, despite evidence to the contrary, many faculty believe that their training as objective experts legitimizes their ability to assess applicants independently of racial, ethnic, and other social characteristics.

Homophily, or love of self, has roots in social similarity, which breeds preferences (and the strongest divides) between individuals based on likenesses in race, ethnicity, age, religion, education, occupation, and gender, generally in that order (Lazarsfeld and Merton, 1954; McPherson *et al.*, 2001). It functions by associating one's social group with superiority while associating other groups with negative feelings, and can limit one's social and professional networks in a manner that restricts the information that is received (McPherson *et al.*, 2001). In the context of admissions, it is also coupled to likeability and perceptions of risk in decision-making processes (Kanter, 1977). As a consequence, social similarities between an applicant and a faculty reviewer who is tasked with predicting the most-qualified applicants may result in an unconscious susceptibility to homophily. Consequently, homophily in graduate admissions could disproportionately advantage applicants who represent the dominant culture by impacting who faculty reviewers see as least risky, competitive for admission, a good "fit" for the graduate program, and worthy of admission.

Holistic (or whole-file) review is an emerging solution to bias, implicit and explicit, in the doctoral admissions process. It minimizes use of the triage strategy and increases consideration of other components of the application package, such as the personal statement, letters of recommendation, evidence of research participation, productivity, and traditional quantitative metrics such as grade point average (GPA) and standardized test scores. This type of review places greater emphasis on the skills and experiences that are thought to be relevant for

success in graduate school. Holistic review also minimizes dependence on quantitative metrics that may reflect a “fixed mind-set” that may lead to unfavorable outcomes in programs in which critical and analytical thinking are critical for success (Kyllonen, 2011). The goal of holistic review is to prevent a single part of the application package from disproportionate consideration in the admissions process. One of the principles by which holistic review may succeed is that it helps to point out that the strength of an applicant in one area may overcome a weakness in another area.

Evaluation plans that detail outcomes of holistic review provide insight on the benefits of the practice. At the University of Illinois in Chicago, implementation of a holistic review process at the College of Nursing significantly increased the diversity of the entering nursing student class (Scott and Zerwic, 2015). The number of URM students at the College of Nursing who were offered admission increased from 36.8 to 42.5%. However, this report lacked the statistical analyses necessary to appropriately determine causality. In another study, holistic review was assessed after 1 year at a “western medical school” (Cantwell *et al.*, 2010). Statistical comparisons were made before (2005–2008) and after (2009) holistic review with regard to admissions outcomes. While the authors determined that URMs were 2.4 times more likely to be admitted to the “western medical school” than their well-represented peers, admissions decisions could be statistically linked only to increases in interview scores, resulting in the observed changes in diversity during the 1-year period. However, the interview scores were not reliable between interviewers, which suggested a flaw in the link between interview scores and admissions outcomes. These representative studies suggest positive impacts of holistic review in increasing diversity in the admissions process, but also demonstrate the need for more rigorous statistical analyses on the impact of holistic review on admissions outcomes.

There is no clearly prescribed practice of holistic review by doctoral admissions committees. Some institutions advertise the goals of their holistic review, although it is often unclear whether these processes are consistent across schools or programs even within the same institution. Phrases such as “credentials considered include academic qualifications gauged by indicators in multiple parts of the application,” “holistic evaluation criterion include, but are not limited to, the potential for academic success,” and “the selection of students is based on an individualized, holistic review of each application, including (but not limited to) the student’s academic record, letters of recommendation, the scores on both the General GRE and GRE Subject test, the statement of purpose, personal qualities and characteristics, as well as past accomplishments and potential to succeed” describe goals of the program without elaborating on specifics of the actual process of holistic review (Kent and McCarthy, 2016).

We recently reported detailed efforts to remove barriers for URM students to enter and complete doctoral programs in the biomedical sciences (Wilson *et al.*, 2018). The goal of the work was to determine whether initiatives that were implemented by The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences (GSBS) over the past decade increased representation of historically underserved and underrepresented minorities at the graduate

school. Statistically significant increases in diversity over time in the doctoral program were the result of several initiatives centered around an overhaul of the admissions process (Wilson *et al.*, 2018). Specifically, there were significant increases in the number of male and female URM applicants who were offered an interview following the switch from an admissions process that was heavily focused on GRE scores to one in which the GRE was one of several factors considered for admissions (Wilson *et al.*, 2018). While these data suggest that efforts to increase matriculant diversity at the graduate school have been successful, the role and impact of the GRE, in light of reports as to its discriminatory impact (Moneta-Koehler *et al.*, 2017), remained unclear. Thus, we sought a deeper understanding of the impact of the GRE in a holistic admissions process to determine whether any observed influences reflect the reported biases of the test.

As a case study of holistic review, we analyzed data over a decade-long period (2007–2017) at The University of Texas MD Anderson Cancer Center UTHealth GSBS. We analyzed how a shift in the method of applicant review by the graduate school impacted our previously reported increases in the diversity of doctoral applicants following the implementation of a holistic review process at the graduate school. We present data that show 1) use of GRE scores to triage applicants significantly reduces the diversity of the applicant pool and 2) holistic review can be an effective tool to mitigate the variance of GRE scores that is observed between different populations of applicants. Further, our results provide a model for a holistic review process that considers the GRE in a manner that is independent of race, ethnicity, and gender.

METHODS

The Graduate School

The University of Texas MD Anderson Cancer Center UTHealth GSBS is the degree-granting entity of The University of Texas MD Anderson Cancer Center and The University of Texas Health Science Center at Houston. The GSBS offers three master’s programs, a medical physics PhD program, and eight biomedical sciences PhD programs in 1) biochemistry and cell biology, 2) cancer biology, 3) genetics and epigenetics, 4) immunology, 5) microbiology and infectious diseases, 6) neuroscience, 7) quantitative sciences, and 8) therapeutics and pharmacology. The graduate school has a centralized biomedical sciences admissions process in which students who are admitted to the graduate school can join any of the biomedical sciences programs.

Data Sources

All work was conducted at the GSBS Deans’ Office. The data presented were extracted from the admissions and student databases as previously described (Wilson *et al.*, 2018).

Definitions

URM student/applicant: An American citizen who self-identified as Black/African American, Native (American Indian, Native Alaskan, Native Hawaiian), Pacific Islander, or Hispanic in the application for admission.

Well-represented student/applicant: An American citizen who self-identified as white (non-Hispanic) or Asian American in the application for admission.

International student/applicant: An individual who is not a U.S. citizen or permanent resident. Racial or ethnic data are not collected from international students/applicants in the application for admission.

Domestic student/applicant: An individual who is a U.S. citizen or permanent resident. Racial and ethnic data are voluntarily collected in the application for admission.

Participants

Applicants. For determination of the number of applicants who applied to the doctoral program who were administratively triaged based on Quantitative and Verbal Reasoning GRE scores (had a score below the 50th percentile on the Quantitative or Verbal Reasoning sections of the GRE) or were reviewed for admissions between 2007–2012 and 2013–2017, data were collected from our admissions database and exported into Excel and PRISM for analyses.

- 2007–2012: applicants ($n = 2945$); applicants triaged by Quantitative and Verbal Reasoning scores ($n = 1073$); applicants reviewed ($n = 1872$).
- 2013–2017: applicants ($n = 2871$); applicants triaged by Quantitative and Verbal Reasoning scores ($n = 0$); applicants reviewed ($n = 2871$).

Students. Data on student body demographics and Verbal and Quantitative Reasoning GRE scores were collected by querying the student database. Student records that were incomplete (missing data: GRE scores, admissions committee scores) were eliminated from the study before data analysis.

- 2007–2012: Of the 528 student records that were analyzed, 96 were removed from the analyses because of a lack of GRE scores and/or admissions committee scores. Thirty-nine of these records belonged to MD/PhD applicants who were not required to take the GRE, while 57 records did not have an admissions committee score in the database. Following removal of these records from our analyses, a total of 432 records were analyzed. Additionally, 39 admissions scores were identified as outliers by statistical analysis software and removed for a final data set of 393 (see *Outliers* section).
- 2013–2017: Of the 307 student records that were analyzed, 56 were removed from the analysis because of a lack of GRE scores and/or admissions committee scores. Thirty-eight of these records belonged to MD/PhD applicants who were not required to take the GRE, while 18 records did not have an admissions committee score in the database. Following removal of these records, a total of 251 records were analyzed.

Outliers

We first tested for the presence of outliers in the samples on the basis of committee admissions scores so that the results were not skewed based on data-entry errors, scoring errors, or extremes in the scoring patterns. The automated ROUT method included in PRISM software v. 7.03 (GraphPad Software, La Jolla, CA) was used, and the false-discovery rate for outlier detection (Q) was set to 1%. Out of 432 students in the 2007–2012 sample, ROUT detected 39 outliers. Those outliers were removed before additional statistical analyses were performed.

Statistical Analyses

Statistical analyses were conducted using PRISM.

Sample. See detailed description in the *Participants* section. Linear regression analysis was used to determine whether there were differences between the admissions committee assessments (slopes) and acceptance thresholds (y-intercepts) between selected student groups. The D’Agostino and Pearson omnibus and Shapiro-Wilk normality tests were used to assess normality for the admissions scores in the 2007–2012 and 2013–2017 samples. The Spearman’s rank correlation coefficient was calculated to determine relationships between GRE scores, admissions committee scores, and other variables collected for each student (e.g., gender, race) within the two samples. Fisher’s exact tests were used to determine whether there were significant differences between applicants who were reviewed during the 2007–2012 and 2013–2017 periods. Significance was set to $p < 0.05$, and all values are reported as two-tailed.

Predictive Metrics. The input variables used in this study were the percentiles of applicants’ scores on the Quantitative and Verbal Reasoning GRE sections. The percentile on the Quantitative and Verbal Reasoning sections of the GRE were measured instead of raw scores to normalize variances that can occur between tests.

Performance Metrics. The output variable used in statistical analysis of each data set was the normalized admissions score. During the 10-year period in which scores were collected, the scale for admissions scores changed from 10–100 (100 as the highest score an applicant could receive from any one reviewer) to 1–9 (1 as the highest score an applicant could receive from any reviewer). Thus, all scores were normalized on a scale of 1–9 before analyses.

PROCEDURES

Preintervention: Metrics-Based Applicant Review

From 2007 to 2013, completed applications for admission to the graduate school were first subjected to an arbitrary cutoff based on cumulative undergraduate and graduate GPA and GRE scores. Applicants with a GPA of 3.0 or higher and GRE scores above the 50th percentile on each section of the test were immediately sent for formal review by the admissions committee. Applicants who fell below these cutoffs were labeled as tier II and administratively triaged. These applicants were not reviewed for admission to the graduate school unless faculty members or directors of individual PhD programs requested that they be “rescued” for review by the admissions committee. However, the tier I and II labels assigned to applicants were visible to the admissions committee before review and discussion.

During applicant review, each applicant was presented to the committee by a primary and secondary reviewer; this was followed by open discussion of each application. In addition to discussing the applicant’s quantitative metrics, the committee considered the applicant’s academic qualifications, research experience, and potential for success in graduate school; the sophistication of the of the personal and research statements; recommendations from research faculty; and the optional statement of adversity that discloses obstacles or disadvantages a student may have had to overcome to achieve academic success.

Following the discussion, the two reviewers presented final scores, providing a range for other committee members. A committee score was calculated from a simple average of scores provided by each committee member. The committee score informed the GSBS deans' admission decision.

Postintervention: Holistic Applicant Review

A complete description of the admissions committee and the scoring of applicants has been previously described (Wilson *et al.*, 2018 and Supplemental Material therein). Briefly, in 2013, the admissions committee altered its application review to shift the discussion away from GRE scores and focus more on academic success and noncognitive factors in the belief that they might be better predictors of long-term success in the biomedical sciences.

In this multitiered applicant review, applications for admission to the graduate school are accepted from the beginning of September until the beginning of January each year. Applications are processed as they are completed and are assigned to one of four admissions committee meetings scheduled during the months of November, December, January, and February. In this process, all applicants are reviewed without significantly delaying the admissions process and overwhelming reviewers by separating applicants for review into two tiers: tier I and tier II.

- Tier I applicants have GPAs of 3.0 or higher and GRE scores in Quantitative, Verbal, and Analytical sections that are higher than the 50th percentile. Tier I applicants are immediately sent for formal review by the admissions committee, but their tier I statuses are hidden.
- Tier II applicants have GPAs of less than 3.0 and/or GRE scores on any section of the test that are below the 50th percentile. Applicants who fall into the tier II category are then reviewed by an internal admissions committee of three assistant/associate deans who have doctoral degrees in the biomedical sciences. Members of this committee meet four times between the months of November and January. This internal review process involves review of each part of the application by all three members of the committee. Applicants are moved into the tier I group for discussion during the next formal GSBS admissions committee when at least two out of three internal review members consider their applications potentially acceptable, while keeping their "tier" status undisclosed to the admissions committee.

The admissions committee meetings proceed in a manner similar to National Institutes of Health study sections, with each applicant presented by a primary, secondary, and tertiary reviewer; this is followed by open discussion of each application. Following the discussion, the three reviewers offer their scores, which provides a range of scores for other committee members. A committee score is calculated from a simple average of scores provided by each committee member and informs the GSBS deans' admission decision.

RESULTS

Using GRE Score Cutoffs to Triage Doctoral Applicants Disproportionately Affects URM Students

Concerns have been raised regarding bias in the GRE based on demographic distribution of scores. For example, well-represented (white and Asian-American) males on average score

higher than their underrepresented peers and females across all groups (ETS, 2014; Miller and Stassun, 2014). Thus, we examined the impact of a metrics-based applicant review on doctoral applicants at the GSBS during 2007–2012, in which applicants who had below-average scores (tier II) were not reviewed by the admissions committee (see *Procedures* section). We analyzed the number applicants who applied to the doctoral program, the number of applicants who had below-average Quantitative and Verbal Reasoning GRE scores, and the number of applicants who were reviewed by the admissions committee (Table 1, 2007–2012). While 36% of the total applicant pool was triaged, the percentage of white male applicants who were triaged was 26%, while the percentage nearly doubled for Black male applicants (59.5%) and tripled for Black female applicants (75.8%). This procedure therefore reduced the consideration of African Americans from 141 to 41 (2.2% of the 1872 applications that were reviewed) in the applicant pool. Similarly, Hispanic male and female applicants were triaged at double the rates of their white male applicant peers (50 and 67%, respectively), reducing their numbers to 4% of the applicant pool that underwent admissions committee review. Overall, while 12% of all doctoral applicants belonged to a historically URM group (Black, Native, and/or Hispanic), nearly two-thirds (64%) of those applicants had below-average scores on the Quantitative and/or Verbal Reasoning sections of the GRE and were not reviewed by the admissions committee. Well-represented applicants (Asian Americans and whites) had significantly higher representation in the reviewed applications than their respective proportions in the total applicant pool based on lower than average triage rates. Further, despite having more female (~53%) than male applicants (~47.4%) in the total applicant pool during this time period, we observed gender impacts that, across most racial groups, resulted in a decreased representation of female applicants compared with male applicants (49.6 and 50.4% following triage, respectively). These findings suggest that heavy use of the GRE can limit the accessibility of STEM graduate education for historically underrepresented and underserved groups.

An Admissions Committee Can Mitigate GRE Score Variances between Demographic Groups

We have reported significant increases in the diversity of our doctoral applicant and student body demographics between 2004 and 2017 following the identification and removal of barriers that prevent entry of URM students into graduate school (Wilson *et al.*, 2018). We found that moving away from a metrics-based admissions process resulted in significant increases in the admissions of URM students in a manner in which there were no significant changes in GRE scores over time. However, we did not analyze whether the changes that we observed were a result of 1) discontinuing the process of administrative triage based on GRE cutoffs, 2) disparities between URM and non-URM applicant review, or 3) a combination of both. Thus, we hypothesized that a committee in which GRE scores were at the center of applicant review would correlate with committee scores and disproportionately impact URM applicants. In this instance, an analysis of applicants who accepted the offer of admission (see *Students* subsection in *Methods*) would reveal a correlation between admissions committee scores and GRE scores during 2007–2012, and these correlations would be

TABLE 1. The number and percentage of applications for admission to the graduate school that were received, triaged by Quantitative (Q) and Verbal (V) GRE scores, and reviewed by the admissions committee from 2007–2012 and 2013–2017 by race, ethnicity, gender, and citizenship status

Application status ^a	2007–2012 (Preintervention)				2013–2017 (Postintervention)			
	Numbers		Percentages		Numbers		Percentages	
	Received	Triaged by GRE (Q,V) scores < 50th percentile	Reviewed with GRE (Q,V) scores ≥ 50th percentile	Triaged by GRE (Q,V) scores < 50th percentile	Reviewed and reviewed	Offers of admission	Reviewed and reviewed	Offers of admission
Total applicants	2945	1073	1872	100	2871	541	100.0	100.0
Total male applicants	1396	453	943	47.4	1292	209	45.0	38.6
Asian-American males	97	34	63	3.3	89	17	3.1	3.1
White males	399	104	295	13.5	367	91	12.8	16.8
Black males	42	25	17	1.4	31	6	1.1	1.1
AI, AN, NH, or PI males	18	9	9	0.6	11	3	0.4	0.6
Hispanic males	79	40	39	2.7	65	13	2.3	2.4
International males	782	256	526	26.6	756	81	26.3	15.0
Total female applicants	1549	620	929	52.6	1579	332	55.0	61.4
Asian-American females	127	46	81	4.3	122	31	4.2	5.7
White females	414	133	281	14.1	452	154	15.7	28.5
Black females	99	75	24	3.4	72	14	2.5	2.6
AI, AN, NH, or PI females	7	3	4	0.2	14	6	0.5	1.1
Hispanic females	109	74	35	3.7	108	28	3.8	5.2
International females	794	299	495	27.0	860	109	30.0	20.1

Fisher's exact test ^b	Reviewed		
	Preintervention	Postintervention	Significance, p
Total URM applicants, %	15 (128/848)	23 (301/1331)	<0.0001
Male URM applicants, %	15 (65/423)	19 (107/563)	0.1497
Female URM applicants, %	15 (63/425)	25 (194/768)	<0.0001
International applicants, %	45 (1021/1869)	55 (1616/2947)	0.9054

^aAI, American Indian; AN, Alaskan Native; NH, Native Hawaiian; PI, Pacific Islander. A subset of students in this study identified as belonging to more than one racial group. Students who identify as Hispanic could also select one of the four categories for racial identity.

^bNumber of applicants who were reviewed at the GSBS. Fractions in parentheses represent the number of students/total applicants. International applicants were not included in totals for comparisons involving URM applicants. Fisher's exact test was used to compare applicant proportions pre- vs. postintervention.

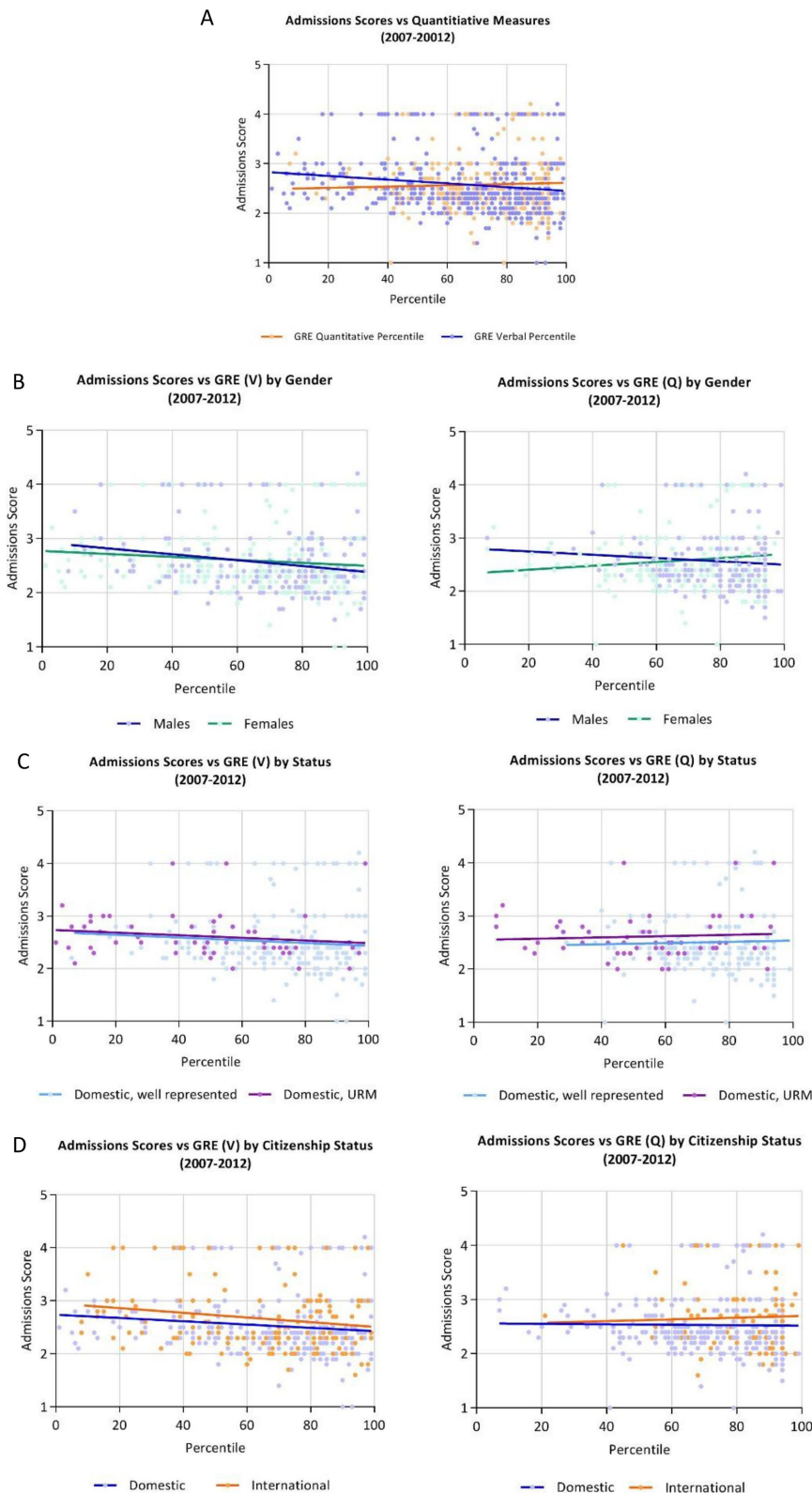


FIGURE 1. Admissions scores in metrics-based applicant review are correlated to Verbal Reasoning GRE scores. Committee admissions scores plotted according to the Quantitative and Verbal GRE scores (A). The mean Quantitative GRE score was 74.44 with an SD of 18.01, and the mean Verbal GRE score was 65.67 with an SD of 24.2. The relationship between admissions and GRE scores based on applicant gender (B), URM status (C), and citizenship status (D).

impacted following changes to applicant review (2013–2017). To clarify which of the three possibilities might be responsible for the changes we observed in the entering student body, we used scatterplot analyses to determine whether there were any correlations between applicants’ Quantitative Reasoning and/or Verbal Reasoning GRE scores (x-axis) and the assessments of those applicants by the committee (Figure 1, admissions scores on y-axes; see *Statistical Analyses* section). Briefly, linear regression analyses were added to the scatterplots to visualize trends between Quantitative Reasoning (orange trend line) and Verbal Reasoning (blue trend line) scores (Figure 1A). There were three possible outcomes of these analyses: a horizontal trend line, a positively sloped trend line, and a negatively sloped trend line. A horizontal trend line suggests that there is no relationship between the GRE scores and admissions committee scores. A negatively sloped trend line suggests that as an applicant’s GRE scores increase, the ranking given to that applicant by the admissions committee also increases. A positively sloped trend line suggests that as an applicant’s GRE scores increase, he or she is more likely to be ranked negatively by the admissions committee.

Statistical analyses of the data showed a horizontal trend line between Quantitative Reasoning GRE scores and admissions committee scores, and a negatively sloped trend line between Verbal Reasoning GRE scores and admissions committee scores (Figure 1A). These data suggest that applicants were assessed in a manner that was dependent on Verbal Reasoning GRE scores, but not Quantitative Reasoning scores. Further analyses of these data revealed that there were no statistically significant differences (or differences in the slopes of the trend lines) between admissions committee assessments of male and female (Figure 1B, left panel), URM and well-represented (Figure 1C, left panel), or domestic and international students (Figure 1C, right panel). Table 2 summarizes the statistical analyses that were conducted on each data set. Confirming our trend line analyses, there was a correlation between admissions committee scores of applicants and Verbal Reasoning GRE scores (Table 2) that was independent of gender (Figure 1B), race and ethnicity (Figure 1C), or citizenship status (Figure 1D).

While we observed no differences between the rate (i.e., changes in slopes) at

TABLE 2. Summary of statistical analyses of preintervention data (2007–2012) in Figure 1^a

	Number	Correlation between GRE and admissions score		Linear regression			
		Quantitative P value (Spearman's rank correlation coefficient)	Verbal P value (Spearman's rank correlation coefficient)	Difference in slope (assessment)		Difference in elevation (threshold)	
				Quantitative p value	Verbal p value	Quantitative p value	Verbal p value
All students	393	0.7879 (–0.0136)	<0.0001 (–0.1957)	0.0152		N/A	
Males	164	0.1933 (–0.1021)	0.0082 (–0.2057)	0.0623	0.2855	0.5480	0.7350
Females	229	0.3755 (0.0588)	0.0048 (–0.1856)				
Well represented	198	0.2729 (–0.0783)	0.0564 (–0.1358)	0.9887	0.9799	0.1784	0.6259
URM	55	0.9831 (–0.0029)	0.0213 (–0.3100)				
Domestic	250	0.0727 (–0.1137)	0.0006 (–0.2167)	0.6433	0.5965	0.0373	0.0405
International	143	0.3924 (0.0721)	0.0479 (–0.1657)				

^aSee *Statistical Analyses* section for a complete description of the statistical tests that were used on this data set. Analyses in which $p < 0.05$ are highlighted in red.

which applicants are assessed, we next determined whether there were differences in the thresholds of assessments (y -intercepts of trend lines) between groups (e.g., implicit bias). For example, it is possible for two groups to be assessed at the same rate (slope) and yet have vastly different thresholds by which they are assessed (y -intercept). Thus, there are two possible outcomes: trend lines that overlap and have similar slopes and y -intercepts, or parallel lines that have the same slopes but different y -intercepts. The former example would suggest that there were no differences in the rates or standards by which applicants were assessed. The latter example suggests that, while there were no differences in the rates of applicant assessment, the standards by which they were assessed were different. Thus, to determine whether there were differences between the thresholds of applicant assessment, we analyzed whether there were differences in y -intercepts between groups. As a positive control, we analyzed differences between domestic and international applicants, because international applicants were reviewed by a separate admissions committee during this time period. This committee consisted of faculty members with expertise in cultural competency, record of training international students, interpretation of TOEFL scores, and an understanding of the rankings of international institutions (Table 2, Difference in elevation columns). While we did not observe any statistically significant differences in standards/thresholds (y -intercept) between males and females or URMs and the well represented, we did observe significant differences in the standards/thresholds between international and domestic applicants. The differences in the y -intercepts of trend lines in Figure 1D suggest that, while there are no differences in the assessments/slopes of the lines between these groups, the standard/threshold was higher for international students. These data suggest that the admissions committee was able to mitigate the differences in GRE scores in a manner that is independent of GRE biases in the domestic applicant pool. Importantly, it demonstrates that the lack of diversity that we reported for the entering student body in the preintervention data set was not the result of disparities in applicant review, but of a metrics-based applicant review process (Wilson *et al.*, 2018). Thus, increases in diversity that we previously reported following changes to our admissions process were the result of switching from a metrics-based review process to a holistic review of all applicants.

A Multitiered Holistic Applicant Review Process Increases Diversity of the Applicant Pool without Increasing the Workload of the Admissions Committee

In 2013, the applicant review process at the GSBS was altered such that all applicants were reviewed by an admissions committee to address the concern that use of the GRE as an initial screen for graduate admissions could unintentionally select against URM and international applicants. The process was also modified to ensure that all applicants were reviewed and discussed in a manner that was not focused on the GRE, but instead on an applicant's letters of recommendation, evidence and quality of research experiences, research productivity, and personal statement. To determine whether holistic review of all applicants increased the diversity of the pool of applicants, we repeated our analysis of application review outcomes of the doctoral applicant pool based on race/ethnicity, gender, and citizenship status (Table 1, 2013–2017). Following implementation of the holistic review process, we observed statistically significant increases in the number of URM applicants who were reviewed by the admissions committee (Table 1, 2007–2012 vs. 2013–2017, Fisher's exact test, Reviewed, Total URM applicants). The percentage of females, Blacks, Hispanics, and Native females increased in the pool of applicants who were reviewed. This increase in the diversity of the applicant pool reviewed by the committee corresponds with our previous finding that there were increases in the diversity of the entering student body during this same time period (Wilson *et al.*, 2018) and may reflect the diminished role of the GRE in the review process.

Overcommitted faculty may be tempted to use GRE scores as a method to reduce their workload. Thus, to prevent significant increases in the number of applicants reviewed by the admissions committee members, an internal admissions committee was created (see *Procedures* section) to review applicants who had below-average GRE scores and GPAs of less than 3.0. The admissions committee had no knowledge of prior review by the internal admissions committee so as not to positively or negatively influence applicant scoring.

To determine whether review of applicants with below-average GRE scores changed how the admissions committee scored applicants, we measured whether there were differences in admissions scores by race/ethnicity and gender

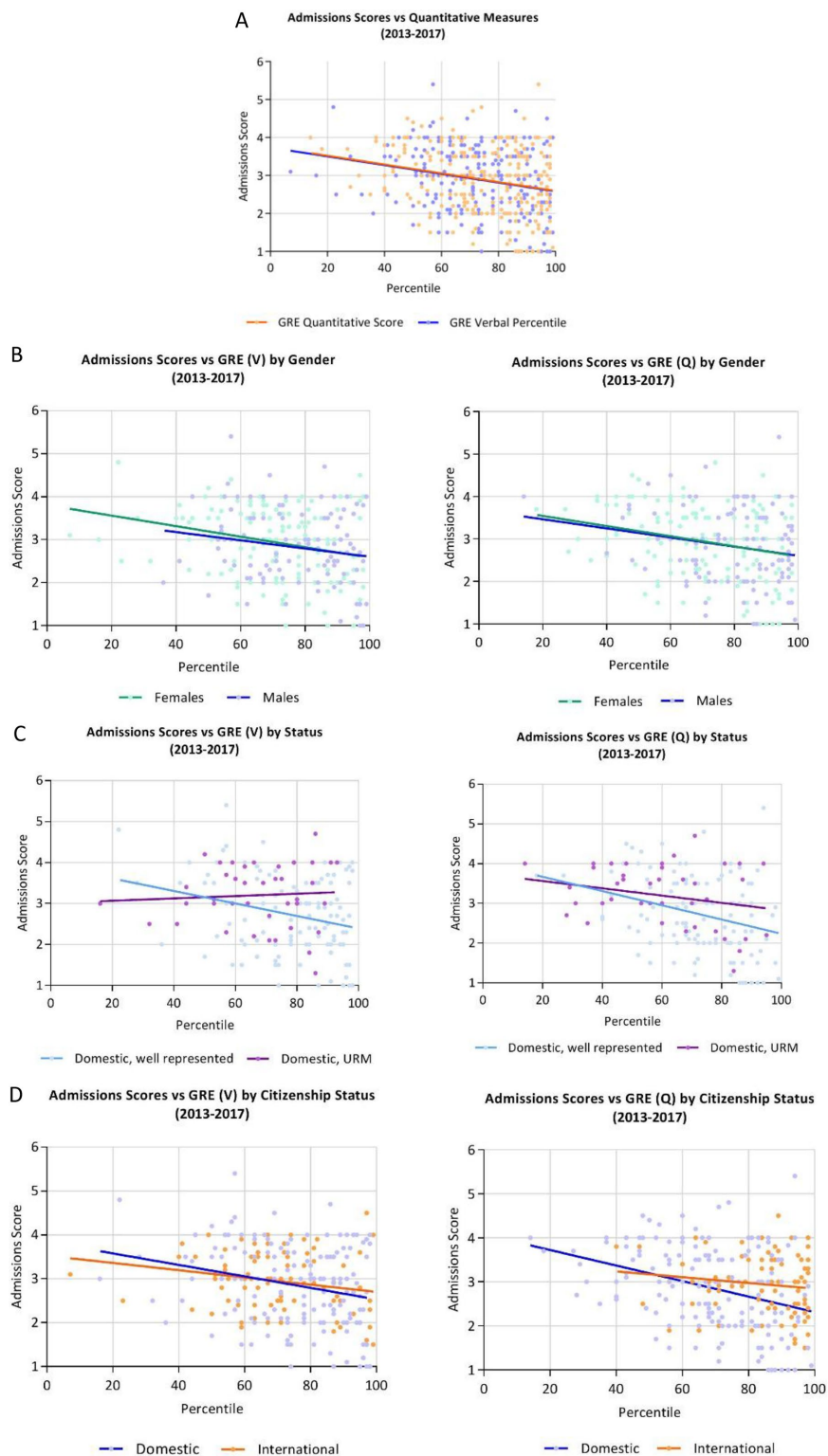


FIGURE 2. Admissions scores in holistic applicant review are correlated to Verbal and Quantitative Reasoning GRE scores. Committee admissions scores plotted according to the Quantitative and Verbal GRE scores (A). The mean Quantitative Reasoning percentile was 74.69 with an SD of 17.98, and the mean Verbal Reasoning percentile was 73.43 with an SD of 17.63. The relationship between admissions and GRE scores based on applicant gender (B), URM status (C), and citizenship status (D).

(Figure 2). As in our previous analyses, linear regression analyses were used to visualize trends between Quantitative Reasoning (orange trend line) and Verbal Reasoning (blue trend line) scores (Figure 2A) to determine whether we observed one of three possible outcomes of these analyses: a horizontal trend line (no relationship between the GRE scores and admissions committee scores), a positively sloped trend line (as an applicant’s GRE scores increase, the admissions committee scores him or her less favorably), or a negatively sloped trend line (as an applicant’s GRE scores increase, the admissions committee scores him or her more favorably). Importantly, the concept of multitiered review seemed to diminish the use of the GRE as a method of triage by committee members, as we observed no differences in the assessments of males and females (Figure 2B) and URMs and well-represented applicants (Figure 2C). Additionally, combining the review of domestic and international students resulted in no differences in the assessment of this control data set (Figure 2D, international vs. domestic applicants). Table 3 summarizes the statistical analyses that were conducted on each data set.

Surprisingly, while there was an improvement in the committee’s consistency in assessments between groups (Figure 2A vs. Figure 1A and Table 3), there were noticeable differences between the correlations and thresholds of Verbal and Quantitative Reasoning GRE scores and admissions scores between URMs and well-represented applicants and domestic and international applicants (Table 3). These data suggest that the slight shifts in how the committee assesses diverse and international students may be a result of a push from the institution to increase the diversity of the student body.

CONCLUSIONS AND RECOMMENDATIONS

Graduate school admissions decisions have profound effects on students, institutions, and the greater STEM landscape. Thus, data-driven analyses of the admissions process are of paramount importance to the broader community. To our knowledge, this is the first complete analyses of the GRE and its influence on the scores of a doctoral admissions committee. We have observed that while an applicant’s GRE scores correlate with scores received from the admissions committee during the

TABLE 3. Summary of statistical analyses of postintervention data (2013–2017) in Figure 2^a

	Number	Correlation between GRE and admissions score		Linear regression			
		Quantitative <i>p</i> value (Spearman's rank correlation coefficient)	Verbal <i>p</i> value (Spearman's rank correlation coefficient)	Difference in slope (assessment)		Difference in elevation (threshold)	
				Quantitative <i>p</i> value	Verbal <i>p</i> value	Quantitative <i>p</i> value	Verbal <i>p</i> value
All students	251	0.0003 (–0.2260)	0.0002 (–0.2348)	0.9866		0.8458	
Males	104	0.1112 (–0.1571)	0.0535 (–0.1899)	0.8518	0.6766	0.9252	0.6766
Females	147	0.0045 (–0.2330)	0.0024 (–0.2482)				
Well represented	118	0.0009 (–0.3012)	0.0109 (–0.2335)	0.2987	0.0593	0.0940	0.0283
Underrepresented minority	43	0.2564 (–0.1769)	0.6053 (0.0811)				
Domestic	160	<0.0001 (–0.3468)	0.0024 (–0.0239)	0.1399	0.4493	0.0080	0.7566
International	91	0.2477 (–0.1232)	0.0543 (–0.2024)				

^aSee *Statistical Analyses* section for a complete description of the statistical tests that were used on this data set. Analyses in which $p < 0.05$ are highlighted in red.

process of holistic review (Figures 1 and 2), these assessments were independent of race, ethnicity, and gender (Tables 2 and 3). This finding was surprising, as our metrics-based applicant review data support multiple reports that GRE scores are biased against historically URMs and women in STEM (Table 1, 2007–2012). Thus, the finding that our admissions committee assessed applicants in a manner that was correlated with GRE scores, but independent of the demographics of the applicant pool, suggests that a well-informed admissions committee, that has a reasonable workload, can mitigate the negative effects of GRE score bias that we observed (Table 1, 2007–2012) and that are reported in the literature. However, our findings do not diminish the challenges of using a standardized test with small samples sizes for validity studies of URMs to normalize a diverse pool of applicants in which scores are often linked to demographic factors such as race, ethnicity, socioeconomic class, and citizenship status.

During the process of metrics-based applicant review, we report that more than half of URM doctoral applicants were removed from the applicant review pool based on below-average scores on the Verbal and Quantitative Reasoning sections of the test. However, this is a conservative estimate, because our analysis does not take Analytical Writing GRE scores into consideration due to their subjective nature and relatively lower significance in admissions committee discussions. Thus, our data provide a starting point for additional analyses and changes to admissions processes in which a multitiered approach to applicant review uses the GRE in an appropriate and fair context. Further, contrary to concerns raised by scholars in the biomedical sciences, use of the GREs can provide a method that may assist faculty and administrators in extracting meaning from quantitative metrics provided by an applicant to graduate school.

This work supports a number of studies, working group reports, books, and articles that call for significant changes in admissions in higher education. However, despite the data, little has changed to improve the admissions process. One could argue that the process has gotten worse based on comparisons between the number of doctoral students or recipients over time relative to their demographics in the United States (Antonio, 2002; U.S. Census Bureau, 2010; Heggeness *et al.*,

2016; NSF, 2017). The slow adoption of change may be due to resistance by faculty who are unaware of their biases. Their inability to see themselves as gatekeepers of doctoral attainment but, rather, as “scientific” reviewers of applicants, can be problematic for reforming admissions (Posselt, 2016; MacLachlan, 2017). However, these blind spots, if not addressed, disadvantage applicants who do not remind reviewers of themselves in regard to education, identity, experiences, and social standing. Further, only ~70% of graduate applicants are reviewed by a centralized graduate admissions committee (MacLachlan, 2017). This creates an additional level of complexity, by which ~30% of graduate applicants could be reviewed by drastically different standards across disciplines, departments, and/or programs within the same institution.

Thus, we propose the following recommendations for a centralized, holistic review process of applicants in graduate education:

Definition of “holistic” review in graduate admissions. The admissions process in graduate education is a laborious process tailored to meet the needs of institutions, programs, and/or departments. It often requires the help of faculty members to identify applicants who are most likely to succeed, and thus requires the ability to consistently judge each applicant in the context of the applicant pool. Thus, we supply a blueprint for holistic review in the graduate admissions processes by providing a standard for analytical assessment of existing practices and a core set of data-driven approaches for reasonable and practical guidance. Institutions seeking to implement data-driven approaches with regard to the admissions process should leverage the institutional mission, values, and goals to clearly outline the process of holistic review across departments, programs, and disciplines. This strategy should involve key stakeholders in the process, such as institutional leaders, program directors, and admissions committee members, to ensure that there is transparency, consensus, and buy-in, which in the long term will ensure success of any changes made.

A method to overcome time as barrier to holistically reviewing all applicants. Institutions receive many applicants for admission to graduate school each year. While institutional leaders agree

that a holistic review of all applicants is the best method to identify the most qualified applicants, 58% of admissions personnel and faculty report time as the rate-limiting factor in applicant review (Kent and McCarthy, 2016). Review requires assessment of an applicant's academic performance, contributions to science, potential to contribute to the research mission of the program, and commitment to his or her educational success. Thus, our description of a multitiered system of applicant review can be used to help overcome the use of GRE scores by the admissions committee and staff by reducing their workloads while allowing each applicant an opportunity to be reviewed for admission. While programmatic knowledge and an understanding of the admissions process are essential, members of the internal review committee should hold doctoral degrees in a STEM discipline and serve as program/institutional leaders. Ideally, members of this committee should be willing to serve at least two terms, and terms should be staggered between members to ensure that membership does not change at once, which ensures preservation of committee practices, fresh perspectives over time, and diversity of thought. This practice also makes it difficult for any members to dominate the discussion or overly influence the committee over time.

The GRE as a tool in holistic review. While many institutions claim that their holistic applicant review mitigates the controversy over the validity of GRE scores, it remains unclear whether a committee's admissions decision is linked to an applicant's GRE scores. Specifically, statistical analyses that determine whether applicants' admissions scores are correlated with their GRE scores are not available or are unclear. Thus, we provide an alternative use for the GRE in the process of admissions in a manner that assists with the assessment of applications during the admissions process.

LIMITATIONS OF THIS STUDY

While the work presented provides the framework for standardizing holistic review in the graduate admissions process, it is limited in its interpretation for the following reasons.

Reliability. We were limited in our ability to analyze the reliability of admissions committee scores. However, given the number of different committee members (between 10 and 15) who assess applications each year, the significant correlations between the scores of the committee and GRE (Q) scores indicate that committee assessments of applicants have not changed over time.

Score compression. The admissions committee score is a relative method for ranking applicants in the applicant pool. At the graduate school, the admissions committee is instructed to use the full range of scores when assessing applicants to improve the ability to compare applicants across different admissions committee meetings, but this does not always happen. Thus, it is possible that the lack of correlation between GRE scores and admissions committee scores in some instances was not statistically significant because of compression of the score range by reviewers.

Highly contextualized data set. Although our data set represents all doctoral students who applied and matriculated at the graduate school during the course of this study, the size of the

sample is relatively small compared with the national population of doctoral students. Given this limitation, caution should be exercised when applying the recommendations in this article to larger graduate programs and groups of doctoral students.

ACKNOWLEDGMENTS

We thank Nguyen Cao for help with querying the databases for data and Lalit Patel as well as Drs. Michelle Barton and Michael Blackburn for helpful comments.

REFERENCES

- Antonio, A. L. (2002). Faculty of color reconsidered: Reassessing contributions to scholarship. *Journal of Higher Education, 73*, 582–602.
- Atwood, K. L., Manago, A. M., & Rogers, R. F. (2011). GRE requirements and student perceptions of fictitious clinical psychology graduate programs. *Psychological Reports, 108*(2), 375–378. doi: 10.2466/07.28.PR0.108.2.375-378
- Becher, T., & Trowler, P. R. (Eds.). (2001). *Academic tribes and territories: Intellectual enquiry and the culture of disciplines* (2nd ed.). Buckingham: Society for Research into Higher Education & Open University Press.
- Cantwell, B., Canche, M., Milem, J., & Sutton, F. (2010). Do the data support the discourse? Assessing holistic review as an admissions process to promote diversity at a US medical school. In annual meeting of the Association for the Study of Higher Education, Indianapolis.
- Clark, B. R. (1989). The academic life: Small worlds, different worlds. *Educational Researcher, 18*(5), 4–8. doi: 10.2307/1176126
- Durning, S. J., Dong, T., Hemmer, P. A., Gilliland, W. R., Cruess, D. F., Boulet, J. R., & Pangaro, L. N. (2015). Are commonly used premedical school or medical school measures associated with board certification? *Military Medicine, 180*(suppl 4), 18–23. doi: 10.7205/MILMED-D-14-00569
- Educational Testing Services (ETS). (2011). *Guidelines for the Use of GRE Scores*. Princeton, NJ. Retrieved June 25, 2018, from www.ets.org/gre/institutions/scores/
- ETS. (2014). *A snapshot of the individuals who took the GRE revised general test*. Princeton, NJ. Retrieved June 25, 2018, from www.ets.org/s/gre/pdf/snapshot_test_taker_data_2014.pdf
- ETS. (2015a). *Graduate Record Examination: Guide to the use of scores, 2015–2016*. Princeton, NJ. Retrieved June 25, 2018, from www.ets.org/s/gre/pdf/gre_guide.pdf
- ETS. (2015b). *Identify the best applicants for your next incoming class: Avoid these five common mistakes when using GRE scores*. Retrieved June 25, 2018, from www.ets.org/gre/bestpractices
- Gibbs, K. D., Basson, J., Xierali, I. M., & Broniatowski, D. A. (2016). Decoupling of the minority PhD talent pool and assistant professor hiring in medical school basic science departments in the US. *Elife, 5*, e21393. doi: 10.7554/eLife.21393
- Gibbs, K. D., McGready, J., Bennett, J. C., & Griffin, K. (2014). Biomedical science Ph.D. career interest patterns by race/ethnicity and gender. *PLoS ONE, 9*(12), e114736. doi: 10.1371/journal.pone.0114736
- Grossbach, A., & Kuncel, N. R. (2011). The predictive validity of nursing admission measures for performance on the National Council Licensure Examination: A meta-analysis. *Journal of Professional Nursing, 27*(2), 124–128. doi: 10.1016/j.profnurs.2010.09.010
- Heggeness, M. L., Evans, L., Pohlhaus, J. R., & Mills, S. L. (2016). Measuring diversity of the National Institutes of Health-funded workforce. *Academic Medicine, 91*(8), 1164–1172. doi: 10.1097/ACM.0000000000001209
- Jacobs, J. A. (2013). *In defense of disciplines: Interdisciplinarity and specialization in the research university*. Chicago: University of Chicago Press.
- Kanter, R. M. (1977). *Men and women of the corporation*. New York: Basic Books.
- Kent, J. D., & McCarthy, M. T. (2016). *Holistic review in graduate admissions: A report from the council of graduate schools*. Washington, DC: Council of Graduate Schools.
- Kuncel, N. R., Ones, D. S., & Hezlett, S. A. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: Implications for graduate student selection and performance. *Psychological Bulletin, 127*(1), 162–181.

- Kyllonen, P. (2011). *The case for noncognitive constructs and other background variables in graduate education (ETS GRE Board research report ETS GREB-00-11)*. Retrieved June 25, 2018, from www.ets.org/research/policy_research_reports/publications/report/2011/ined
- Lazarsfeld, P., & Merton, R. (1954). Friendship as a social process: A substantive and methodological analysis. In Berger, M., Abel, T., & Page, C. H. (Eds.), *Freedom and control in modern society* (pp. 18–66). New York: Van Nostrand.
- MacLachlan, A. J. (2017, February). *Preservation of educational inequality in doctoral education: Tacit knowledge, implicit bias and university faculty (UC Berkeley CSHE, 1.17, January 2017)*. Berkeley: Center for Studies in Higher Education, University of California–Berkeley. Retrieved June 25, 2018, from <https://escholarship.org/uc/item/5zv6c3nj>
- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1), 415–444. doi: 10.1146/annurev.soc.27.1.415
- Mervis, J. (2016). Scientific workforce. NSF makes a new bid to boost diversity. *Science*, 351(6277), 1017. doi: 10.1126/science.351.6277.1017
- Milkman, K. L., Akinola, M., & Chugh, D. (2015). What happens before? A field experiment exploring how pay and representation differentially shape bias on the pathway into organizations. *Journal of Applied Psychology*, 100(6), 1678–1712. doi: 10.1037/apl0000022
- Miller, C., & Stassun, K. (2014). A test that fails. *Nature*, 510, 303–304.
- Moneta-Koehler, L., Brown, A. M., Petrie, K. A., Evans, B. J., & Chalkley, R. (2017). The limitations of the GRE in predicting success in biomedical graduate school. *PLoS ONE*, 12(1), e0166742. doi: 10.1371/journal.pone.0166742
- National Center for Education Statistics. (2013). *Integrated postsecondary education data system*. Washington, DC.
- National Institutes of Health. (2012). *Draft report of the Advisory Committee to the Director Working Group on Diversity in the Biomedical Research Workforce*. Bethesda, MD.
- National Science Foundation. (2017). *Women, minorities, and persons with disabilities in science and engineering: 2017 (Special report NSF 17-1310)*. Arlington, VA.
- Okahana, H., & Zhou, E. (2017). *Graduate enrollment and degrees: 2006 to 2016*. Washington, DC: Council of Graduate Schools.
- Pacheco, W. I., Noel, R. J., Porter, J. T., & Appleyard, C. B. (2015). Beyond the GRE: Using a composite score to predict the success of Puerto Rican students in a biomedical PhD program. *CBE—Life Sciences Education*, 14(2), ar13. doi: 10.1187/cbe.14-11-0216
- Park, H. Y., Berkowitz, O., Symes, K., & Dasgupta, S. (2018). The art and science of selecting graduate students in the biomedical sciences: Performance in doctoral study of the foundational sciences. *PLoS ONE*, 13(4), e0193901. doi: 10.1371/journal.pone.0193901
- Posselt, J. R. (2016). *Inside graduate admissions: Merit, diversity, and faculty gatekeeping*. Cambridge, MA: Harvard University Press.
- Roush, J. K., Rush, B. R., White, B. J., & Wilkerson, M. J. (2014). Correlation of pre-veterinary admissions criteria, intra-professional curriculum measures, AVMA-COE professional competency scores, and the NAVLE. *Journal of Veterinary Medical Education*, 41(1), 19–26. doi: 10.3138/jvme.0613-087R1
- Scott, L. D., & Zerwic, J. (2015). Holistic review in admissions: A strategy to diversify the nursing workforce. *Nursing Outlook*, 63(4), 488–495. doi: <https://doi.org/10.1016/j.outlook.2015.01.001>
- Stichweh, R. (1992). The sociology of scientific disciplines: On the genesis and stability of the disciplinary structure of modern science. *Science in Context*, 5, 3–15. doi: 10.1017/S0269889700001071
- U.S. Census Bureau. (2010, March 24). 2010 census shows America's diversity. Newsroom Archives. Retrieved June 25, 2018, from www.census.gov/newsroom/releases/archives/2010_census/cb11-cn125.html
- U.S. Department of Education. (2017). Table 315.20. Integrated Postsecondary Education Data System (IPEDS), Spring 2014, Spring 2016, and Spring 2017 Human Resources component, Fall Staff section. National Center for Education Statistics
- Valantine, H. A., & Collins, F. S. (2015). National Institutes of Health addresses the science of diversity. *Proceedings of the National Academy of Sciences USA*, 112(40), 12240–12242. doi: 10.1073/pnas.1515612112
- Wilson, M. A., DePass, A., & Bean, A. J. (2018). Institutional interventions that remove barriers to recruit and retain diverse biomedical PhD students. *CBE—Life Sciences Education*, 17(2), ar27. doi: 10.1187/cbe.17-09-0210
- Wilson, M. B., Sedlacek, W. E., & Lowery, B. L. (2014). An approach to using noncognitive variables in dental school admissions. *Journal of Dental Education*, 78(4), 567–574.