Journal of Women and Minorities in Science and Engineering 25(4):283–306 (2019)

METRICS FIRST, DIVERSITY LATER? MAKING THE SHORT LIST AND GETTING ADMITTED TO PHYSICS PhD PROGRAMS

Julie R. Posselt,[1,*] *Theresa E. Hernandez,*[1]
Geraldine L. Cochran,[2] *& Casey W. Miller*[3]

[1]*University of Southern California, Rossier School of Education, Waite Phillips Hall 602G, 3470 Trousdale Parkway, Los Angeles, California 90089, USA*

[2]*Rutgers University, Office of STEM Education, Department of Physics and Astronomy, 136 Frelinghuysen Rd, Piscataway, NJ 08854*

[3]*Rochester Institute of Technology, 1104 Thomas Gosnell Hall, RIT College of Science, 84 Lomb Memorial Drive, Rochester, New York 14623, USA*

*Address all correspondence to: Julie Posselt, University of Southern California, Rossier School of Education, Waite Phillips Hall 602G, 3470 Trousdale Parkway, Los Angeles, CA 90089; Tel.: 970-310-4975, E-mail: posselt@rossier.usc.edu

Recent research suggests that faculty reliance on graduate record examination (GRE) scores early in the graduate admissions review disproportionately limits access to doctoral education for women, Black, Latinx, and Native American students. However, some faculty do engage in more holistic review—and in so doing, weigh diversity—when evaluating applicants on a short list. This paper has three objectives: to test the prevalence of this two-stage review process; to examine whether the factors associated with an applicant's selection to a short list differ from those associated with receiving an admission offer; and third, to assess implications of admissions preferences for equitable access. We conducted fixed-effects logistic regression using application-level data from six large, selective physics programs who participated in a project piloting means to increase diversity in physics. We found that faculty in these programs indeed placed a premium on standard evaluation metrics in the first round of review. Women are more likely than men to both make the short list and be admitted, whereas Black and Latinx applicants do not have significantly higher odds than white students of moving forward in the admissions process, all else equal. Our findings add weight to a rising tide of evidence that faculty must revisit the narrow framing they have traditionally used in the admissions process to increase diversity in their graduate programs.

KEY WORDS: *admissions, physics, graduate education, graduate record examination, faculty, test-optional admissions*

1. INTRODUCTION

Despite almost 70 years of predictive validity analyses of graduate admissions exams (e.g., Cureton et al., 1949; Marston, 1971; Sternberg and Williams, 1997; Kuncel et al., 2010), we have surprisingly little quantitative evidence about the relative importance of academic and social background characteristics to an applicant's odds of being admitted to graduate programs. The most comprehensive analysis is now 20 years old: Attiyeh and Attiyeh (1997) found that the three strongest predictors of admission to graduate programs in five fields were high graduate record examination (GRE) scores, high college grades, and attending a selective college. Due to group-level differences in selective college enrollment and GRE scores, those preferences systematically privilege applicants from groups that are already overrepresented in doctoral education—white, male, and wealthy applicants.

Committees enact a preference for applicants with very high GRE scores through several methods, one of the most common being the use of cutoff scores (Potvin et al., 2017). Miller and Stassun (2014) documented that the typical cutoff score used in physics excludes 25% of Asian American applicants, 50% of white applicants, 75% of Mexican American applicants, and almost all African American applicants. Posselt (2016) found, through ethnographic research of doctoral admissions in nine fields including physics and astrophysics, that professors created a short list largely by filtering applicant pools on the basis of GRE scores and grades, the latter of which they contextualized by curricular rigor and college reputation. However, applications that made the short list were subjected to holistic review ("consideration of a broad range of candidate qualities including noncognitive or personal attributes" [Kent and McCarthy, 2016, p. iii]), including contributions to diversity. This pattern was exemplified by one participant comment: "First you have to be above a bar, then we can ask the diversity question." The objectives of the current study were to examine the following research questions about access to physics for applicants from underrepresented groups:

1. In physics PhD programs aiming to increase the enrollment of students from underrepresented groups, are the academic and demographic characteristics associated with being shortlisted for admission different from those associated with receiving an offer of admission?
2. How are GRE scores associated with the probabilities of being shortlisted and admitted for women, Black, and Latinx* applicants?
3. In programs that do not require the physics GRE subject test (PGRE) for admission, do faculty place greater weight on applicants' undergraduate GPA?

We use applicant-level administrative data from six large, selective physics programs who are partners in the NSF-INCLUDES funded Inclusive Graduate Education Network to examine student pathways through the admissions process. We assess the relative strength of various academic and demographic characteristics as predictors of two outcomes: (1) being shortlisted and (2) being admitted.

2. BACKGROUND

Graduate admissions practices in physics is a research topic of importance to researchers, disciplinary faculty, and leadership alike. Physics is a large, powerful field among the disciplines with deep racial and gender inequalities, and its strong-paradigm culture and assumptions that inherent "brilliance" underlies performance make it especially resistant to change (Leslie et al., 2015; Prescod-Weinstein, 2017; Rosa and Menash, 2016;

*Native Americans and Southeast Asians are also underrepresented in physics; however, we were unable to conduct statistical analysis of access for these groups due to very small numbers and/or lack of disaggregated data in our sample.

Blue et al., 2018; Tuttle, 2017). Therefore, understanding and dismantling barriers to equitable access in physics graduate education and scientific careers may come with important lessons for other disciplines that similarly struggle with such inequalities. While there have been a handful of newsletters and opinion pieces touching on graduate admissions in physics, particularly with respect to diversity issues (e.g., Ripin, 1996; Glanz, 1996; Georgi, 2000; Ehrlich, 2007), the first significant analyses of admissions practices in physics (among other disciplines) was conducted by Posselt (2014, 2015, 2016), from an ethnography focused on how faculty conceptualize merit throughout the admissions process and its implications for diversity in graduate education. She found that faculty focus on diversity as a factor in admissions only after applying extremely high standards, using quantitative metrics to rapidly reduce the number of applications that are fully reviewed. Potvin et al. (2017) provided evidence corroborating this trend in physics, through a national survey of PhD program directors and similar leaders in physics departments. Obtaining responses from nearly all programs, they found that between one third and one half of programs apply GRE cutoff scores, or give scores uncontrolled weightings (which may unduly affect the ultimate decision to admit or reject a student). Furthermore, these faculty held applicants' GRE Quantitative and GRE Physics subject test in as high esteem as a student's entire portfolio of undergraduate courses; only letters of recommendation and undergraduate grades in physics and math courses were more important.

The consequence of applying an initial, GRE-based filter is the disproportionate exclusion of women, underrepresented ethnic/racial minorities (i.e., Black, Latinx, Native American), and US citizen applicants. These groups have mean GRE scores that are significantly lower than their comparison groups, as noted by Miller (2013) and Miller and Stassun (2014). A typical cutoff in physics for the GRE-Quantitative is about the 60th percentile among test-takers identifying as wanting to attend graduate school in the physical sciences: 40% of women score above this bar, compared to 60% of men; less than 25% of Mexican Americans and African Americans score above this bar, compared to more than 60% of Asian Americans and 50% of white Americans.

Faculty justify their reliance on the GRE as an initial filter with several rationales, including convenience in winnowing the pool and their belief that scores signal intelligence and belonging, which they view as critical to success in the field (Posselt, 2016). Scherr et al. (2017) went deeper into this pattern within physics, studying the prevalence of fixed mindsets (viewing intelligence as an inherent capacity or ability) and growth mindsets (viewing intelligence in terms of acquired knowledge and effort) about intelligence among faculty who chaired physics graduate admissions committees (Dweck, 2000). They concluded that while both mindsets appeared within nearly all participants to some extent, the majority dominantly exhibited one or the other mindset. Half exhibited growth-dominant mindset, a quarter exhibited fixed-dominant mindset, and the remainder had no dominant mindset. The implications of this are consistent with what was found by Posselt: the genius narrative associated with fixed mindsets manifests itself in physics as a (misplaced) belief among many professors that metrics such as grades and GRE scores signal innate intelligence, and therefore who is destined to succeed in physics graduate school, and that these metrics are thus justifiable as an initial filter. A major study of the validity of common admissions criteria in 25 large physics PhD programs did not find evidence to support these beliefs (Miller et al., 2019). When so many faculty chairing physics PhD admissions committees conduct admissions in ways that are consistent with such beliefs (Potvin et al., 2017), significant barriers to improving diversity in physics are to be expected.

2.1 Theoretical Perspectives and Hypotheses

Judgments of admissibility depend upon available information about applicants and the frames (i.e., "schemata of interpretation" [Goffman, 1974, p. 21]) with which admissions committees interpret that information. We pair perspectives about framing from sociology and behavioral economics, particularly as they have been applied to admissions decision-making behavior, to understand how and why the same criteria may be weighted differently throughout the admissions process.

Sociologies of undergraduate admissions suggest that frames shape decision-making at multiple levels. Macrolevel institutional admissions goals and mesolevel approaches to application review together shape microlevel interpretations of specific information about applicants. At the macro level, colleges may view the aims of admissions as not only identifying the "best" students but also correcting historic injustices against individuals from underrepresented groups, thereby preserving the legitimacy of admissions contests (Grodsky, 2007). At the meso level, Stevens (2009) proposed that decision makers frame admissions work as evaluative storytelling. Grades and test scores are insufficient to distinguish among applicants in today's competitive admissions environment; therefore, after "coarse sorting" on academics, committees construct composite narratives about finalists that contextualize the minutiae of their strengths and weaknesses, highlighting potential to add to a cohort's diversity.

At the graduate level, some faculty may similarly support diversity and equity related efforts in principle; however, the tendency to emphasize academic metrics first and consider the diversity contributions only of those who make the short list suggests that correcting inequities is not their primary priority (Posselt, 2016). Many professors also frame admissions as investment, which inclines them to a risk aversion mesoframe for initial file review (Posselt, 2016). When looking at GRE scores, they may display tendencies toward narrow framing (i.e., approaching assessments of risk without attention to context) that behavioral economics studies have identified among gamblers (Barberis et al., 2006) and investors (Kumar and Lim, 2006). We therefore hypothesize the following:

H1a: GRE scores will have positive, significant relationships with the odds of being placed on a short list.
H1b: Women, Black, and Latinx applicants will have higher odds of admission than men and white applicants, controlling for academic characteristics and being on the short list.

Research indicates that reliance on GRE scores disproportionately excludes women, Black, and Latinx students, but that faculty are aware of inequities and construct diversity as a secondary obligation of admissions. We therefore hypothesize the following:

> H2a: GRE scores will not be associated with higher odds of admission, controlling for student demographic characteristics and being on the short list.
> H2b: In a model that controls for GRE scores, women will have higher odds of being placed on the short list, relative to men, than in a model without GRE scores.
> H2c: In a model that controls for GRE scores, Black and Latinx applicants will have higher odds of being placed on the short list, relative to white applicants, than in a model without GRE scores.

Finally, we recognize the growing use of test-optional admissions and limited evidence of how it changes the weight attributed to other criteria. Therefore, we consider whether programs in our sample that do not require the PGRE may substitute this academic information by placing greater weight on undergraduate academic performance. We therefore hypothesize the following:

> H3: In programs that do not require the PGRE, associations between undergraduate GPA and odds of admission will be stronger.

When engaging in education research and reform related to equity, it is important to provide an operational definition and a conceptual model of equity (Rodriguez et al., 2012), as these determine the groups that are compared and how those comparisons are carried out. Among definitions of equity outlined by Lynch (2000), this study aligns most with "equity that compensates for social injustice for specific groups who have not received fair treatment." The project from which our analyses originated examines representation in physics as a result of inequitable graduate education policy and practice and is trying to counter the view that inequities are a function of inherent quality of individuals from those groups. Among the three models of equity described by Rodriguez et al. (2012), this study aligns most with the *equity of parity* model as it concerns access. In the context of admissions, equity seeking parity would seek higher shortlisting and admissions odds for groups that are disadvantaged relative to groups that have been historically more advantaged, with the goal of parity, or similar outcomes for all. In this study, equity would therefore mean that women, Black, Latinx, and other underrepresented groups have higher odds of being shortlisted and admitted (i.e., that faculty reveal a preference for these groups), in an effort to correct the longstanding disparities in opportunities and enrollment.

3. METHODS

We collected deidentified administrative data representing the admissions processes of the six physics PhD programs of the Inclusive Graduate Education Network (IGEN), an initiative of the National Science Foundation INCLUDES program. These are large and selective programs and were selected for the project on the basis of these criteria as well as expressed and demonstrated interest in changing admissions, recruitment, and/or retention practices in order to increase diversity and reduce inequities in their programs and the field. In the fall of 2017, we worked with the director of graduate studies and support staff in each department to establish a common set of academic and demographic variables that reflect information each program collects about PhD applicants. Each program was then provided a spreadsheet to catalogue every applicant's GRE scores (verbal, quantitative, writing, and [in five of six programs] physics subject test), undergraduate GPA, college attended, gender, race, and international/domestic status. Program personnel logged the status of each application at key junctures including whether an application was complete, whether the student had been placed on a short list, and whether the person was admitted. Data from each program were then merged into a common database.

3.1 Dependent Variables

We examined relationships of individual applicants' academic and demographic characteristics with two dichotomous outcomes: (1) whether or not a student was placed on a short list of finalists for admission and (2) whether or not the applicant received an admissions offer as of April 15. We did not distinguish applicants who may have received an offer of admission after having been placed on a wait list.

3.2 Independent Variables

Our models include four sets of demographic and academic characteristics: (1) demographics: consisting of nationality (international or domestic), gender (male or female),[†] and race/ethnicity; (2) undergraduate GPA (standardized to a 4.0 scale); (3) GRE general test scores: including verbal, quantitative, and writing; and (4) physics GRE subject test scores. When predicting admission, we also controlled for whether the student was on the short list, to account for the separate selection process that determines the short list versus when final admissions decisions are made.

Program data combined United States citizenship and permanent resident status into one category of "domestic." All GRE scores are reported in their raw scaled form, which range from 130 to 170 for verbal and quantitative and 1–6 for the analytical writing. PGRE was added separately due to one program not utilizing it.

Race/ethnicity categories were standardized across the programs and consist of white, Asian/Asian American, Black/African American, Latinx, multiracial/ethnic, Native American, and race/ethnicity unknown. Three race/ethnicity categories required standardization: Latinx, multiracial/ethnic, and race unknown. Four programs provided data with Hispanic or Latinx as an option, while one provided more specific categories

† We acknowledge that gender is not a binary (Blue et al., 2016), but the dichotomous way that these PhD programs collected gender data during admissions did not allow for other identities. This pattern of data collection leads to the oversimplification of gender and contributes to the dearth of information and research available on gender nonconforming physicists.

within the Latinx or Hispanic umbrella, which were aggregated into the Latinx category.‡ Five programs provided racial/ethnic data with a category for multiple or two or more races and one program provided data with multiple categories per some applicants. In order to standardize this data, applications that selected more than one racial/ethnic identity were coded to the "multiracial/ethnic" category and were not counted in the specific categories selected to avoid duplication. All six programs had some applications where no race/ethnicity was selected, and one program also included the category other; these comprise the race/ethnicity unknown category.

3.3 Data Analysis

The dichotomous structure of our dependent variables compelled logistic regression analyses. Separate models for short list and admission permitted us to assess the relative strength of the same criteria at two key decision points, allowing for greater nuance in claims about what matters to admission and when. Within each of these models, we added variables in four blocks, as described above: demographics, GPA, GRE general test scores, and GRE physics test scores. For the model predicting admissions, we also added short list as a covariate. Because the applications are clustered by the program to which they were received, we included fixed effects for program in both models. To simplify interpretation, we report results in terms of odds ratios.

3.4 Predicting Shortlisting

To determine the predictive probabilities of demographics, undergraduate GPA, GRE general test scores, and the GRE physics test score on an application being shortlisted, we specified the following model:

$$Y_{ij} = \beta_{0ij} + Demographics\ vars_{ij} + GPA_{ij} + GRE\_general_{ij} + PGRE_{ij} + \phi_j . \qquad (1)$$

In Eq. (1), Y_{ij} is an indicator variable that equals 1 when application $_i$ submitted to program $_j$ is placed on the short list and zero otherwise. β_{0ij} refers to the constant. $Demographics_{ij}$ is a vector of indicator variables for personal characteristics (domestic = 1; female = 1; and Asian/Asian American = 1, Black/African American = 1, Latinx = 1, multiracial/ethnic = 1, Native American = 1, and race/ethnicity unknown = 1), GPA_{ij} is undergraduate GPA, $GRE\_general_{ij}$ is a vector of GRE scores (verbal [VGRE], quantitative [QGRE], and analytical writing [WGRE]), $PGRE_{ij}$ is the physics GRE subject score, and ϕ_j is a vector of program-specific indicator variables (excluding one program as the reference category) for program fixed effects. Because applications are clustered

‡ One program provided data that included Hispanic identity separate from race, resulting in applicants who selected Hispanic to also select racial categories of white, Black, Asian, and none. While it is impossible to know the intent of the applicants' selection patterns, for simplicity and consistency only applicants who selected Hispanic and none for race were included in the Latinx category; all other applicants who selected Hispanic and a race were included in the multiracial/ethnic category. We understand that this may lead to undercounting of Latinx applicants and overcounting of multiracial/ethnic applicants from this one program.

by the program that received, reviewed, and ultimately decided to short list it or not, we included program-level fixed effects to account for within program variation and improve our model's probability estimates. Similarly, we used clustered standard errors by program to improve model precision.

3.5 Predicting Admission

To determine the predictive probabilities of demographics, undergraduate GPA, GRE general test scores, and the GRE physics test score on an applicant being admitted, we specified the following model:

$$Z_{ij} = \beta_{0ij} + Demographics_{ij} + GPA_{ij} + GRE\_general_{ij} + PGRE_{ij} + Shortlist_{ij} + \phi_j . \quad (2)$$

In Eq. (2), Z_{ij} is an indicator variable that is 1 when application $_i$ submitted to program $_j$ is admitted and zero otherwise. $Shortlist_{ij}$ is an indicator variable for whether an application was placed on the short list (shortlisted = 1), and all other model parameters are the same as described for Eq. (1), including program fixed effects and clustered standard errors. Nearly all applications that were ultimately admitted were drawn from the short list that each program compiled. We therefore included short list as an independent variable in model (2) predicting admission in order to isolate the predictive probability of demographics, GPA, GRE general scores, and the PGRE in final admissions decisions.

3.6 Reliability

The sample sizes for underrepresented racial/ethnic groups (Black, Latinx, and women of all race/ethnicities) are acceptable for making inferences through regression when we aggregate across programs, but they are below standard thresholds for logistic regression (Long, 1997) for the individual programs. We therefore do not report point estimates for Black, Latinx, and/or women for models limited to the sample of applicants in individual programs.

With respect to model reliability, Wald tests ($p < 0.00$) confirmed that all the variables included in our models predicting short list and admissions improved the maximum likelihood estimation and contributed to the best fit possible given the available data. We conducted Wald tests as opposed to likelihood-ratio tests due to the nature of our data being clustered at the program level.

4. RESULTS

4.1 Descriptive Results

Table 1 displays the distribution of demographic attributes of applications within each program. The total number of applications in our sample is 2904. The number of applications programs received ranges from 139 to 765, with an average of 484 per program.

TABLE 1: Distribution of demographics attributes of applications

	Sample Means	Program 1	Program 2	Program 3	Program 4	Program 5	Program 6
Total applications	484	638	713	765	139	352	297
International	53%	48%	60%	55%	49%	53%	42%
Domestic	47%	52%	40%	45%	51%	47%	58%
Percentage of domestic and international applications combined out of total applications							
Men	83%	85%	84%	85%	74%	82%	81%
Women	16%	15%	15%	15%	26%	17%	19%
White	35%	41%	27%	38%	28%	33%	39%
Race/ethnicity unknown	29%	12%	51%	14%	51%	57%	6%
Asian/Asian American	27%	38%	12%	41%	4%	4%	45%
Multiracial/ethnic	4%	3%	8%	5%	2%	1%	2%
Latinx	3%	4%	1%	1%	9%	3%	7%
Black/African American	1%	2%	1%	1%	6%	1%	1%
Native American	< 1%	0%	< 1%	< 1%	0%	< 1%	0%
Percentage of domestic applications received out of total applications							
Men	39%	45%	35%	37%	39%	38%	45%
Women	8%	7%	5%	7%	12%	9%	13%
White	31%	34%	26%	33%	28%	33%	34%
Race/ethnicity unknown	3%	5%	2%	2%	2%	5%	3%
Asian American	5%	6%	5%	4%	4%	3%	13%
Multiracial/ethnic	3%	2%	6%	4%	2%	1%	1%
Latinx	3%	4%	< 1%	1%	9%	3%	6%
Black/African American	1%	1%	1%	1%	6%	1%	1%
Native American	< 1%	0%	< 1%	< 1%	0%	< 1%	0%

Note: Sample means, except for total applications, are weighted. Sample mean for men and women for total applications received regardless of nationality does not equal 100% due to missing observations from programs 2 and 5.

The majority of applicants in the sample were international (53%) and men (84%). Three of the six programs, however, had a majority of domestic applicants. Across the entire sample of applications, and independent of citizenship, the largest racial/ethnic group was white (35%), followed by race/ethnicity unknown (29%) and Asian/Asian American (27%). Among domestic applications only, white applicants remain the largest group (31%), followed by Asian Americans (5%) at a much smaller percentage. In addition, the race/ethnicity unknown group falls to just 3%, illustrating that the overwhelming majority of applicants in this sample whose race/ethnicity is unknown are international. In contrast, the Latinx (3%), Black/African American (1%), and Native American (< 1%) percentages are stable, indicating that the majority of applicants who self-identify with these categories are domestic.[§]

Table 2 shows the overall race and gender breakdown of applications, those that were shortlisted, and those that were offered admission. White men make up the largest group of total applicants at 29%, followed by men with race/ethnicity unknown and Asian/Asian American men, both at 23%. For women identified as white, race/ethnicity unknown, Asian/Asian American, and multiracial/ethnic, the percent of applications are in the single digits, from 6% to 1%. The percent of applications is also in the single digits for men identified as multiracial/ethnic, Latinx, Black/African American, and Native American. Latinx and Black/African American women applications are less than 1% of total applications received for the entire data set. Native American men also make up less than 1% of total applications; there were no applications from Native American women.

Next we describe the racial and gender makeup of those shortlisted and admitted. White men make up 33% of the short list, which is four percentage points greater than their share of the applicant pool. Latinx and Black men are just 3% and 2% of the short list, respectively, a one percentage point increase each from their representation in the applicant pool. White women make up 9% of the short list, which is a three percentage point increase from their share of the applicant pool. All other groups in the short list either maintain or drop in proportion relative to their share of the applicant pool. The absence of Black women in the short list indicates that all of their applications had been eliminated from consideration in this preliminary stage of consideration. Of those admitted, white men maintain the largest proportion with 29%, equal to their share of the applicant pool. Meanwhile, white women make up 12% of admits, twice their proportion of the applicant pool and attaining the only consistent increase in share from applicant pool to short list to admit. The proportion of Latinx and Black men and women admitted remained relatively stable in the single digits compared to their proportion of those shortlisted. None of the shortlisted Native American men were admitted.

§ Researchers argue that it is important to disaggregate racial/ethnic data to the extent possible in order to reveal discrepancies in popular narratives about group access and achievement (Teranishi, 2007). Asian American students represent a diverse umbrella category that is often further lumped together with international Asian students and, therefore, Asian/Asian American students are often categorized as over-represented in STEM. However, disaggregation by citizenship shows that Asian American applicants in our overall sample constitute a slightly smaller proportion of the sample than the representation of Asians in the United States at 5.4% (U.S. Census Bureau, 2016). Within each program, the percentage of applicants who are Asian American varies considerably, from 3% to 13%.

TABLE 2: Sample characteristics by gender and race

	Applications		Short List		Admissions	
	Men	**Women**	**Men**	**Women**	**Men**	**Women**
N	2420	476	753	209	469	173
White	29%	6%	33%	9%	29%	12%
Race/ethnicity unknown	23%	5%	17%	7%	15%	8%
Asian/Asian American	23%	4%	20%	5%	19%	7%
Multiracial/ethnic	4%	1%	4%	< 1%	5%	< 1%
Latinx	2%	< 1%	3%	< 1%	3%	1%
Black/African American	1%	< 1%	2%	—	2%	—
Native American	< 1%	—	< 1%	—	—	—

Note: Each cell for men and women under Applications is out of the total applications; similarly, each cell under Short List is out of the total shortlisted, and each cell under Admissions is out of the total admitted. The totals do not include observations where gender was missing. The total for each of applications, short list, and admissions does not equal 100% due to rounding to the nearest full percentage point, except when rounding would result in 0% and hide the small numbers present. A dash indicates $N = 0$.

Table 3 shows the distribution of academic metrics for applications in our sample and by program. Applications across all programs had a mean undergraduate GPA of 3.63 (SD = 0.35), with a range of 1.65–4.0. The mean QGRE score was 165 (SD = 4.78), with a range from 146 to 170. The mean VGRE score was 158 (SD = 6.53) out of 170, with a range from 135 to 170. The mean WGRE score was 3.87 (SD = 0.80), with a range of 1.5–6. Finally, the PGRE score was 843 (SD = 130.53) out of 990, with a range of 410–990. Among the programs, GPA and GRE scores were similar in both their ranges and means. The mean GPA for each program ranged from 3.52 to 3.70. Mean QGRE scores for each program ranged from 164 to 166, VGRE scores ranged from 156 to 159, and mean WGRE scores ranged from 3.71 to 4.06. Mean PGRE scores displayed more variation, from 788 to 865.

4.2 Multivariate Results

Table 4 displays the complete results of the logistic regression models estimating odds of making the short list and receiving an admissions offer. The two models predicting short list and admissions, respectively, build in blocks of covariates, starting with demographics, then adding undergraduate GPA, GRE general scores, and finally physics GRE.[¶] The admissions model also controls for short list as a covariate at each stage of the model build. We report the findings according to our hypotheses and then present predicted probabilities for given student profiles.

¶ We added PGRE separately because one program did not use it in their admissions process. Analysis excluding physics GRE encompasses observations from all six programs, whereas analysis including physics GRE only incorporates observations from the five programs that considered physics GRE scores. Where significant differences exist between outcomes in models that do and do not consider the physics GRE, these results are noted below.

TABLE 3: Distribution of academic metrics of applications

	Program	Mean	SD	Min	Max
Undergraduate GPA	1	3.70	0.28	2.00	4.00
	2	3.65	0.36	1.70	4.00
	3	3.65	0.36	1.65	4.00
	4	3.52	0.39	2.38	4.00
	5	3.56	0.38	2.00	4.00
	6	3.52	0.38	1.70	4.00
	All	3.63	0.35	1.65	4.00
Verbal GRE	1	159.2	6.4	140	170
	2	157.3	6.5	136	170
	3	157.8	6.4	136	170
	4	156.8	6.6	142	170
	5	155.8	6.6	135	170
	6	156.9	6.6	140	170
	All	157.6	6.5	135	170
Quantitative GRE	1	165.9	4.4	147	170
	2	165.8	4.7	147	170
	3	166.0	4.2	149	170
	4	163.6	5.4	150	170
	5	164.1	5.3	146	170
	6	164.1	5.6	146	170
	All	165.4	4.8	146	170
Analytical writing GRE	1	4.1	0.8	2.0	6.0
	2	3.8	0.8	2.0	6.0
	3	3.9	0.8	2.0	6.0
	4	3.9	0.8	2.0	5.5
	5	3.7	0.8	1.5	6.0
	6	3.8	0.8	2.0	6.0
	All	3.9	0.8	1.5	6.0
Physics GRE	1	864.6	111.9	510	990
	2	856.9	127.4	500	990
	3	857.4	122.1	450	990
	4	—	—	—	—
	5	792.4	143.7	420	990
	6	788.0	149.1	410	990
	All	843.3	130.5	410	990

Note: GPA on a 4.0 scale. Quantitative, verbal, and physics GREs are raw, scaled scores.

TABLE 4: Logistic regressions predicting odds of short list and admission

	Predicting Short List				Predicting Admission			
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Short list					369.5^{***}	297.8^{***}	311.7^{***}	1783.5^{***}
					(373)	(306)	(325)	(2162)
Demographics								
Domestic	2.514^{***}	2.646^{***}	1.856^{***}	2.776^{***}	0.762	0.752	0.627	0.981
	(0.519)	(0.493)	(0.261)	(0.530)	(0.279)	(0.273)	(0.300)	(0.457)
Female	1.863^{***}	1.872^{***}	1.960^{***}	3.181^{***}	2.811^{**}	2.964^{**}	3.034^{**}	4.516^{**}
	(0.189)	(0.200)	(0.224)	(0.503)	(0.929)	(1.062)	(1.100)	(2.500)
Race/ethnicity unknown	0.779	0.832	0.852	0.662^{**}	0.865	0.874	0.866	0.834
	(0.100)	(0.095)	(0.105)	(0.103)	(0.156)	(0.163)	(0.119)	(0.125)
Asian/Asian American	1.399^{**}	1.503^{**}	1.601^{***}	1.292	0.943	1.022	1.016	1.032
	(0.158)	(0.225)	(0.208)	(0.199)	(0.376)	(0.422)	(0.436)	(0.434)
Multiracial/ethnic	0.770^{*}	0.834	0.940	0.779	2.710^{**}	2.815^{***}	3.130^{***}	2.544^{***}
	(0.088)	(0.105)	(0.144)	(0.116)	(0.844)	(0.860)	(0.587)	(0.394)
Latinx	1.124	1.484	3.009^{***}	2.569^{***}	2.254	2.747	3.273^{*}	2.513
	(0.222)	(0.363)	(0.824)	(0.728)	(1.381)	(1.789)	(1.836)	(1.588)
Black/African American	1.175	1.991	3.432^{*}	3.654	3.194	4.597	7.103	7.112^{*}
	(0.696)	(1.137)	(1.689)	(2.643)	(2.868)	(4.270)	(7.843)	(6.605)
Native American	0.332	0.659	1.552	2.981	1	1	1	1
	(0.389)	(0.746)	(2.904)	(8.034)	(.)	(.)	(.)	(.)
Undergraduate GPA		6.080^{***}	4.259^{***}	2.901^{***}		3.522^{***}	2.996^{**}	2.811^{**}
		(1.368)	(1.115)	(0.892)		(1.206)	(1.056)	(1.001)

TABLE 4: (*continued*)

	Predicting Short List				Predicting Admission			
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Verbal GRE			1.044^{***}	1.034^{**}			1.026	1.016
			(0.012)	(0.012)			(0.019)	(0.020)
Quantitative GRE			1.060^{**}	0.990			1.037^{*}	0.965
			(0.020)	(0.015)			(0.017)	(0.023)
Analytical writing GRE			1.534^{***}	1.540^{**}			1.229^{**}	1.219^{*}
			(0.181)	(0.202)			(0.091)	(0.102)
Physics GRE				1.006^{***}				1.005^{***}
				(0.001)				(0.001)
Constant	0.139^{***}	0.000^{***}	0.000^{***}	0.000^{***}	0.010^{***}	0.000^{***}	0.000^{***}	0.000^{**}
	(0.020)	(0.000)	(0.000)	(0.000)	(0.009)	(0.000)	(0.000)	(0.000)
N	2896	2742	2696	2537	2892	2738	2693	2534
df_m	4	4	4	3	4	4	4	3
neg2ll	3397.0	3072.4	2879.4	2621.2	1319.0	1253.3	1213.7	1027.1
r2_p	0.0774	0.123	0.168	0.196	0.569	0.569	0.577	0.619

$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

Note: Coefficients in odds ratios; standard errors in parentheses. Race/ethnicity are in the order of greatest number of applicants to least for the full sample. All specifications include fixed effects for the program, with program dummy variables and one program dropped for reference.

df_m, model degrees of freedom; neg2ll (these are Ls for log likelihood), deviance; r2_p, pseudo R squared

4.2.1 H1a: GRE Scores Will Have Positive, Significant Relationships with Odds of Being Placed on a Short List

We examined whether GRE scores had a significant relationship with the odds of being placed on the short list. In model 3 without PGRE, we find support for hypothesis 1a. Each additional ten points on the verbal and quantitative score is associated with an increase in short list odds of 44% ($p < 0.001$) and 60% ($p < 0.01$), respectively. Each additional point on the writing score is associated with a 53% increase in the odds of being shortlisted ($p < 0.001$). However, in model 4 predicting shortlisting that includes PGRE, each ten-point increase in PGRE is associated with a 6% increase in the odds of being placed on the short list, all else equal. And after adding PGRE to the model, the quantitative GRE score is no longer a statistically significant predictor of being placed on the short list. When PGRE is available, admissions committees may not rely as heavily on the quantitative GRE score, but when PGRE is not available, the weight that might have been given to it is shifted to QGRE. Alternatively, covariance in QGRE and PGRE may help explain the difference in the QGRE's significance across these two estimations.

4.2.2 H1b: Women, Black, and Latinx Applicants Will have Higher Odds of Admission than Men and White Applicants, Controlling for Academic Characteristics and Being on the Short List

Hypothesis 1b considered whether women, Black/African American, and Latinx students had higher odds of being admitted compared to men and white applicants, respectively, controlling for GPA, GRE, and being on the short list. We find partial support for this hypothesis. Model 7 in Table 4 shows that on average, women on the short list have 3 times the odds of being admitted compared to men ($p < 0.01$) and 4.5 times the odds of being admitted when PGRE is considered and held constant. Latinx applicants on the short list have 3.3 times higher odds than white applicants of being admitted ($p < 0.05$) when controlling for GPA and GRE general scores. However, they are not more likely to be admitted when PGRE is also included and held constant (see model 8 in Table 4). By contrast, Black/African American applicants on the short list do not have statistically significant different odds of being admitted compared to white applicants when only controlling for GPA and GRE general scores (see model 7). Yet, when PGRE is added as a control in model 8, Black/African American applicants' odds are about 7 times higher than white applicants ($p < 0.05$). Results for African American students, however, should be interpreted with care, given their low representation in the sample (1%).

4.2.3 H2a: GRE Scores Will Not Be Associated with Significantly Higher Odds of Admission, Controlling for Student Demographic Characteristics and Being on the Short List

We examined whether GRE scores would remain significantly associated with offers of admission controlling for placement on the short list, demographics, and GPA. We do not find support for hypothesis 2a, suggesting that standardized test scores continue to be weighed significantly during final decisions about who on the short list will be offered admission. The exception to this overall finding is that VGRE is no longer significantly associated with offers of admission in the final stage, as supported by models 7 and 8.

Contributing to the balance of evidence against hypothesis 2a, WGRE scores are significant in predicting admission ($p < 0.01$, $p < 0.05$) in models 7 and 8, respectively. Interestingly, a similar pattern of significance for QGRE and PGRE scores to that in the short list models emerges in the models for admission. Specifically, in model 7, which does not include PGRE, QGRE is significantly associated with offers of admission ($p < 0.05$); however, QGRE is not related to admission in model 8, which controls for PGRE. In that model PGRE does predict admission ($p < 0.001$).

The pattern of QGRE predicting short list and admissions absent PGRE provides further support for the suggestion that QGRE and PGRE explain the largely overlapping variance in who gets into physics doctoral programs. Evidence for hypothesis 1a indicated that GRE scores are prominent factors in initial filtering of the pool, and the balance of the evidence against hypothesis 2a suggests that admissions committees continue to rely heavily on standardized tests in final decisions.

4.2.4 Models with and without GRE Scores

Recall: H2b: In a model that controls for GRE scores, women will have higher odds of being placed on the short list, relative to men, than in a model without GRE scores and H2c: In a model that controls for GRE scores, Black/African American and Latinx applicants will have higher odds of being placed on the short list, relative to white applicants, than in a model without GRE scores.

To model whether professors' interpretations of applicant GRE scores is disproportionately filtering out applicants from groups that are underrepresented in physics, we looked at how odds of being shortlisted change for women, Black/African American, and Latinx applicants as we shift from a model with only demographics and GPA to a model that also has GRE scores. Indeed, we do see support for hypotheses 2b and 2c, although with some mixed results depending on whether or not the model includes PGRE. To be clear about interpretation: We learn by adding new variables to a regression model how much variation in the outcome of interest is explained by group variation in the new variables. Thus, by adding GRE scores (a variable for which there is known gender and racial/ethnic group variation) to models predicting who is shortlisted, we can see the extent to which the consideration of GRE is affecting odds of women, Black, and Latinx students making the short list.

For hypothesis 2b, we find that all else equal, women's odds of being shortlisted relative to men increase slightly when QGRE, VGRE, and WGRE are added to the model and increase further to 3.2 times ($p < 0.001$) that of men's odds when PGRE is also included. Similarly, holding GRE scores constant, Latinx applicants gain a statistically significant advantage in the odds of being shortlisted relative to white applicants. Latinx applicants' odds of being shortlisted are three times higher than white applicants when controlling for the GRE general scores ($p < 0.001$). By adding GRE scores to the model, we hold constant its effect in the admissions process and thus are able to envision how other factors would relate to admissions outcomes if GRE scores had a weight of 0 in the minds of decision makers.

4.2.5 H3: Programs That Do Not Require the PGRE Will Have Stronger Associations between Undergraduate GPA and Odds of Admission

Finally, we compared models of individual programs to examine whether those that do not require physics GRE have stronger associations between GPA and admission offers than those that do require the subject test. We find support for hypothesis 3a in that the positive impact of GPA on admissions is strengthened by program 4, which does not require the physics GRE. In our model for program 4, a one-point increase in GPA (i.e., from a 3.0 to a 4.0) is associated with 20 times greater odds of being admitted ($p < 0.01$). In contrast, programs 1, 5, and 6 have no significant association between GPA and offers of admission, while a one-point increase in GPA is associated with 6.2 ($p < 0.001$) and 4 ($p < 0.01$) times greater odds of an admissions offer for programs 2 and 3, respectively.

4.3 Predicted Probabilities for Given Student Profiles

To show the consistency with which women applicants displayed an advantage in admissions probability relative to men when academics are held equal, Figs. 1 and 2 simulate women's and men's predicted probabilities of being admitted. We display how these probabilities vary according to two key quantitative metrics—physics GRE scores and undergraduate GPA—as well as what happens to these probabilities when the other metrics are simulated as average for the sample (to simulate probabilities for a "typical" applicant) vs. very high (to simulate probabilities for academically high achieving applicants).

Figure 1 displays the increasing probabilities of admission that come with increasing PGRE scores, with panel A setting other academic variables in the model (undergraduate GPA, QGRE, VGRE, and WGRE) at their sample means (i.e., to resemble a typical applicant) and panel B setting other academic variables at the sample maxima (i.e., to resemble applicants with the strongest metrics under conventional admissions). Women at the low end of this sample's PGRE range, who have mean academics in other regards, have approximately a 40% probability of admission compared to 17% probability of admission for men with the same characteristics. However, women at the low end of the PGRE range who are simulated as having very high grades and other GRE scores have a 60% probability of admission. This probability rises to 100% (within model error) if they also have other academics at the maximum for other metrics or to 90% if their other academic metrics are at the sample mean. Men with high PGRE scores and maximum scores on other academic metrics have a probability of admission in excess of 80%, still lower than that of women, suggesting either a gender preference or gender contextualization when interpreting applications. Figure 2 displays similar overall patterns, albeit

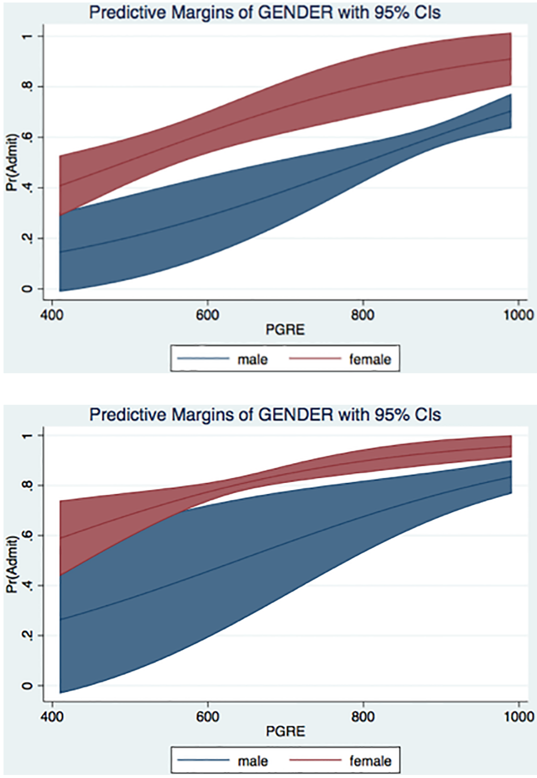

FIG. 1: Probability of admission for men and women as a function of physics GRE scores with UGPA, QGRE, VGRE, and WGRE set to (a) sample means and (b) sample maxima

across the range of UGPA rather than PGRE. Note here that the confidence intervals are much wider at the low end of UGPA, largely because few people with GPAs below a 2.5 applied to these programs.

5. DISCUSSION

The physics PhD programs in this study are part of a project aimed at greater inclusion of women and underrepresented ethnic/racial minority groups in physics. We see with relative clarity that gender may be a consideration in the admissions process. Women applicants have both greater odds of being placed on the short list and greater odds of being admitted if they are on the short list in all models. However, the same is not true for Black/African American and Latinx applicants. This implies that gender diversity is likely a consideration in admissions decisions and that committees have a process, albeit informal, for addressing inequities in gender (Posselt, 2016). The difference in results for Black/African American and Latinx applicants may indicate that admissions committees do not have a process for addressing ra-

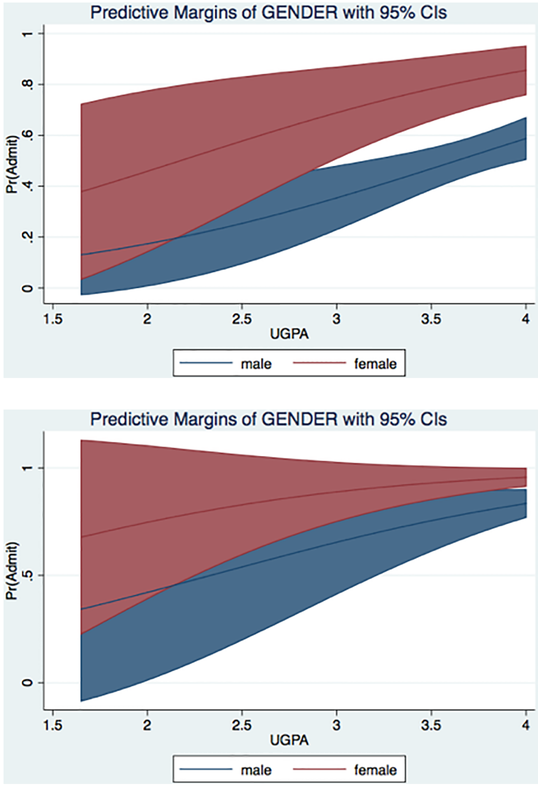

FIG. 2: Probability of admission for men and women as a function of undergraduate GPA with PGRE, QGRE, VGRE, and WGRE set to (a) sample means and (b) sample maxima

cial/ethnic diversity or that the process used is not successful. Multiracial/ethnic applicants on the short list had statistically significant higher odds of being admitted in all models (5–8). Without qualitative data on how these applicants were perceived by admissions committee members, speculation as to the reason for their higher odds is unwarranted.

When controlling for GPA and GRE general scores, the Physics subject test is a significant predictor for Latinx but not Black/African American applicants' placement on the short list. Counterintuitively, this relationship flips when also controlling for short list to determine the odds of admission for these groups. These results provide evidence consistent with previous research (Miller, 2013) that standardized tests are a barrier to graduate school access for Latinx and Black/African American students. These findings also support the need to disaggregate data by demographics where possible—in both data collection and analysis (Teranishi, 2007)—rather than grouping individuals with various identities as "underrepresented minorities."

In addition, the results indicate a somewhat ambiguous role for the physics GRE for Latinx and Black/African American applicants in the admissions process. This may be due, in part, to the small numbers of applicants from these groups; relatively small *n*'s from these groups could be a sign of a type 2 error, in which insufficient statistical power precludes us from seeing relationships that are present in reality. It could also be that we added it after the other GRE exams. Our dataset also contained one program that did not utilize the physics GRE, changing the samples analyzed by the models that included this variable, which may have also contributed to the change in significance of the physics GRE for different racial/ethnic groups across models for short list and admissions. These results point to the need for further study of admissions factors comparing programs that currently require the physics GRE with more programs that do not require or utilize this factor.

Verbal, quantitative, and writing GRE scores had significant relationships with odds of being placed on the short list when the physics GRE was not included in the model. This finding is consistent with previous research indicating that scores on standardized tests are weighed heavily in the admissions process (Potvin et al., 2017). In the model where the physics GRE was controlled along with undergraduate GPA and general GRE scores, Black/African American applicants already on the short list had greater odds of being admitted. This supports findings that the physics GRE, in particular, is a barrier to admissions for Black/African American applicants (Miller and Stassun, 2014). Since no Black/African American women made the short list, this group may face additional, unexamined barriers to graduate school in the admissions process.

6. LIMITATIONS

Our sample consisted of large, selective PhD programs in physics programs that opted into a network aimed at changing admissions, recruitment, and retention practices. These programs may be more subject to decision-making biases of large organizations, but they may also be more inclined to consider gender and race/ethnicity in their evaluation and selection processes. Regardless, they were not randomly selected, so the results are not generalizable to all types of US physics PhD programs. Similarly, readers should not generalize to all evaluations of students who identify with specific racial/ethnic or gender categories.

This project exists because of extremely low participation of some groups in physics, a challenge that presents limitations for this research as well. For example, there is a longstanding pattern of excluding indigenous groups from research studies (Jones, 2010; Popejoy and Fullerton, 2016), and indeed, participating programs did not have enough Native American applicants for us to perform statistical analysis on their outcomes. Also, interpretation of the multivariate analysis in light of the descriptive statistics reveals that women's greater odds of being shortlisted and admitted are true primarily for white and international (both from Asian and race/ethnicity unknown) women applicants due to their overrepresentation among women applicants. The same implication is true for reporting results for particular racial/ethnic groups that largely represent the outcomes for only one gender. Due to the very low number or absence of Latinx, Black/

African American, and Native American women, the results discussed are more practically applicable to men in these categories.

Analytically, we compared outcomes for women and Black/African American and Latinx applicants to male applicants and white applicants, respectively. As researchers, we created research questions and related hypotheses about whether or not the odds of being placed on the short list and odds of admissions for women, Black/African American, Latinx applicants were greater than those for white men. Although this is a standard practice, it runs the risk of implying white and male engagement in physics as a norm to which other groups should be performing.

7. IMPLICATIONS FOR FUTURE RESEARCH

The mixed results in our findings for Latinx and Black/African American applicants may be a function of their low representation in the sample of applicants. Groups underrepresented in physics face a number of barriers to applying to graduate physics programs (Cochran et al., 2017). Ambiguity and opacity about the way that applicants are typically evaluated in the admissions process (Potvin et al., 2017) are two such problems. This ambiguity is part of a vicious cycle that has long plagued research on racial equity in STEM and graduate education: Institutions' failure to provide equitable access to African, Latinx, and Native Americans has resulted in persistently low enrollment that makes rigorous, quantitative analysis of the reasons for their presence/absence and success/struggle difficult to determine. In the case of physics, African, Latinx, and Native Americans earn 9.4% of all physics bachelor's degrees, which both caps the potential for more equitable representation in graduate education and the workforce while compromising quantitative analyses of the relationships with forces that may impede their progress in this field.

To address this problem, we need data-gathering efforts across institutions in order to create larger subsamples of students from underrepresented groups. With such samples, it would be valuable to examine, for example, how much committees weigh the GRE and whether it varies for different racial/ethnic groups. Given the large fraction of international students in physics, it would also be valuable to analyze whether programs that receive large numbers of applications from international students put greater weight on the VGRE and WGRE scores, perhaps informally using these measures as proxies for English ability. We need more research, too, on the role that socioeconomic status plays in students' opportunities within the field of physics.

Future research should consider study designs that allow for inclusion of Native Americans, Southeast Asian, gender nonconforming, LGBTQ, and other rarely studied populations in STEM. For example, we also need to continue qualitative studies that, while not broadly generalizable across institutions, may exhibit theoretical generalization—that is, they may elucidate mechanisms that explain the low baccalaureate pursuit in physics among groups that are underrepresented in physics, and may uncover stimuli to their loss throughout the various transitions leading to careers in the field. Qualitative studies are especially useful in capturing phenomena among populations present in small numbers. These results point to the need for descriptive statistics disaggregated as race/ethnicity by gender even when the absolute numbers are insufficient for statistical significance testing.

8. IMPLICATIONS FOR PRACTICE

The combination of relationships we found corroborates descriptive analyses by Miller and Stassun (2014): rigid reliance on the GRE (i.e., especially when faculty are constructing the short list) in physics PhD programs may be preventing underrepresented groups from advancing in the admissions process. Our findings thus add weight to the arguments of researchers and leaders in the physics and broader graduate education communities, and increasingly by Educational Testing Service itself, that faculty must revisit the narrow framing with which they approach the admissions process. Specifically, how professors and admissions committees think about the GRE needs to be addressed; its weight as a factor should be controlled and cutoff scores must not be applied.

Implementing systematic holistic review offers one possibility for a more thoughtful and complete, but still systematic, approach to admissions decision-making. It represents a practical approach toward broader framing through contextualizing students' academic qualifications and by considering a wider set of applicant characteristics from the start of the review process. If faculty decide to include GRE scores in application review, they can control scores' weight through a few means: folding test scores into a broader judgment of academic preparation, for example, and viewing scores as one of the last pieces of information about an applicant rather than one of the first. Putting scores at the end of the file can prevent anchoring bias, in which judgments made about people based on the first things you learn about them tend to stick. Many PhD programs in STEM disciplines are removing GRE requirements or no longer even collecting scores. The impacts of test-optional graduate admissions have yet to be examined empirically.

However, the best evidence about implicit bias teaches us that reducing reliance on the GRE, eliminating it from the review process altogether, and evaluating applicants on a broader set of criteria cannot be institutions' only practical responses. Faculty should receive opportunities to learn about the ways that gender and racial biases can creep into the minds of even well-intentioned decision-makers and how to check those biases. Further, they should become attuned to the distinctive biases and systems of oppression that affect opportunities of women and people of color, generally, and the specific sets of biases and interlocking systems of oppression that affect women of color who identify with different racial and ethnic groups. It is our belief that such consciousness-raising about the distinctive challenges that women, Black, and Latinx students face will be more productive for equity and inclusion in the long term than blinding gender and racial data from applications, if STEM PhD programs wish to be intentional about increasing representation and perspectives of these groups.

As part of their participation in the Inclusive Graduate Education Network, the same PhD programs participating in this study have now also received professional development in the development of more holistic approaches to admissions. The programs responded to that training in somewhat distinct ways. However, whether they adopted formal, rubric-based analysis or simply initiated the first open conversation about their assumptions about GRE scores and the admissions process, early evidence from the IGEN external evaluator suggested such professional development opportunities and holistic review holds potential for an array of benefits. The barriers for women of all racial/ethnic groups and for people of color are multiple and interrelated (Johnson et al., 2017), demanding a multifaceted and systemic approach toward equity and inclusion. In that approach, our data suggest that professional development that spurs consciousness of relationships between academic preparation and demographics can, in time, result in more equitable outcomes.

ACKNOWLEDGMENTS

We gratefully acknowledge the chairs of the six physics departments who participated in this study and the National Science Foundation for Awards 1633275, 1649297, 1807047, and 1644885, which facilitated this work.

REFERENCES

Attiyeh, G., & Attiyeh, R. (1997). Testing for bias in graduate school admissions. *The Journal of Human Resources*, 32(3), 524–548.

Barberis, N., Huang, M., & Thaler, R. H. (2006). Individual preferences, monetary gambles, and stock market participation: A case for narrow framing. *American Economic Review*, 96(4), 1069–1090.

Blue, J., Traxler, A., & Cid, X. (2018). Gender matters. *Physics Today*, 71(3), 40–46.

Cochran, G., Hodapp, T., & Brown, E. (2017). Identifying barriers to ethnic/racial minority students' participation in graduate physics. Paper presented at Physics Education Research Conference 2017, Cincinnati, OH. Accessed March 17, 2019, from https://www.compadre.org/per/items/detail.cfm?ID=14576

Cureton, E. E., Cureton, L. W., & Bishop, R. (1949). Prediction of success in graduate study of psychology at the University of Tennessee. *American Psychologist*, 4(8), 361–362.

Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality, and development*. New York, NY: Psychology Press.

Ehrlich, R. (2007). Where are the female physicists? *CSWP Gazett*e: *The Newsletter of the Committee on the Status of Women in Physics of the American Physical Society*, 26(2), 3–5.

Georgi, H. (2000). Views from an affirmative activist. *STATUS*, 4–13.

Glanz, J. (1996). How not to pick a physicist? *Science*, 274(5288), 710–712.

Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. Cambridge, MA: Harvard University Press.

Grodsky, E. (2007). Compensatory sponsorship in higher education. *American Journal of Sociology*, 112(6), 1662–1712.

Johnson, A., Ong, M., Ko, L. T., Smith, J., & Hodari, A. (2017). Common challenges faced by women of color in physics, and actions faculty can take to minimize those challenges. *Physics Teacher*, 55(6), 356–360.

Jones, D. (2010). Psychology: A WEIRD view of human nature skews psychologists' studies. *Science*, 328, 1627.
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.
Kumar, A., & Lim, S. S. (2008). How do decision frames influence the stock investment choices of individual investors? *Management Science*, 54(6), 1052–1064.
Kuncel, N. R., Wee, S., Serafin, L., & Hezlett, S. A. (2010). The validity of the graduate record examination for master's and doctoral programs: A meta-analytic investigation. *Educational and Psychological Measurement*, 70(2), 340–352.
Leslie, S. J., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, 347(6219), 262–265.
Marston, A. R. (1971). It is time to reconsider the graduate record examination. *American Psychologist*, 26(7), 653–655.
Lynch, S. J. (2000). *Equity and science education reform*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
Miller, C. W. (2013). Admissions criteria and diversity in graduate school. *APS News*, 22(2), 8.
Miller, C. W., Zwickl, B. M., Posselt, J. R., Silvestrini, R. T., and Hodapp, T. (2019). Typical physics PhD admissions criteria limit access to underrepresented groups but fail to predict doctoral completion. *Science Advances*, 5(1), 1–8.
Miller, C., & Stassun, K. (2014). A test that fails. *Nature*, 510(7504), 303–304.
Popejoy, A. B., & Fullerton, S. M. (2016). Genomics is failing on diversity. *Nature News*, 538(7624), 161.
Posselt, J. R. (2014). Toward inclusive excellence in graduate education: Constructing merit and diversity in PhD admissions. *American Journal of Education*, 120(4), 481–514.
Posselt, J. R. (2015). Disciplinary logics in doctoral admissions: Understanding patterns of faculty evaluation. *The Journal of Higher Education*, 86(6), 807–833.
Posselt, J. R. (2016). *Inside graduate admissions: Merit, diversity, and faculty gatekeeping*. Cambridge, MA: Harvard University Press.
Potvin, G., Chari, D., & Hodapp, T. (2017). Investigating approaches to diversity in a national survey of physics doctoral degree programs: The graduate admissions landscape. *Physical Review Physics Education Research*, 13(2), 020142.
Prescod-Weinstein, C. (2017). Curiosity and the end of discrimination. *Nature Astronomy,* 1, 0145.
Ripin, B. H. (1996). Fighting the gender gap: Standardized tests are poor indicators of ability in physics. *American Physical Society News*, 5(7).
Rodriguez, I., Brewe, E., Sawtelle, V., & Kramer, L. H. (2012). Impact of equity models and statistical measures on interpretations of educational reform. *Physical Review Special Topics-Physics Education Research*, 8(2), 020103.
Rosa, K., & Mensah, F. M. (2016). Educational pathways of Black women physicists: Stories of experiencing and overcoming obstacles in life. *Physical Review Physics Education Research*, 12(2), 020113.
Scherr, R. E., Plisch, M., Gray, K. E., Potvin, G., & Hodapp, T. (2017). Fixed and growth mindsets in physics graduate admissions. *Physical Review Physics Education Research*, 13(2), 020133.
Sternberg, R. J., & Williams, W. M. (1997). Does the graduate record examination predict meaningful success in the graduate training of psychologists? A case study. *American Psychologist*, 52(6), 630–641.
Stevens, M. L. (2009). *Creating a class: College admissions and the education of elites*. Cambridge, MA: First Harvard University Press.
Teranishi, R. T. (2007). Race, ethnicity, and higher education policy: The use of critical quantitative research. *New Directions for Institutional Research*, 2007(133), 37–49.
Tuttle, S. (2017). Astronomical community: The power of being counted. *Nature Astronomy*, 1(6), 0154.
US Census Bureau. (2016). *American community survey 1-year estimates.* Washington, DC: US Census Bureau.