Editors' Suggestion

Physics Graduate Record Exam does not help applicants "stand out"

Nicholas T. Young [1,2] and Marcos D. Caballero [1,2,3,4]

[1] *Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824, USA*
[2] *Department of Computational Mathematics, Science, and Engineering, Michigan State University, East Lansing, Michigan 48824, USA*
[3] *Center for Computing in Science Education & Department of Physics, University of Oslo, N-0316 Oslo, Norway*
[4] *CREATE for STEM Institute, Michigan State University, East Lansing, Michigan 48824, USA*

(Received 24 August 2020; accepted 1 June 2021; published 23 June 2021)

One argument for keeping the physics Graduate Record Exam (GRE) is that it can help applicants who might otherwise be missed in the admissions process stand out. In this work, we evaluate whether this claim is supported by physics graduate school admissions decisions. We used admissions data from five Ph.D.-granting physics departments over a 2-year period ($N = 2537$) to see how the fraction of applicants admitted varied based on their physics GRE scores. We compared applicants with low GPAs to applicants with higher GPAs, applicants from large undergraduate universities to applicants from smaller undergraduate universities, and applicants from selective undergraduate institutions to applicants from less selective undergraduate institutions. We also performed a mediation and moderation analysis to provide statistical rigor and to better understand the previous relationships. We find that for applicants who might otherwise have been missed (e.g., have a low GPA or attended a small or less selective school), having a high physics GRE score did not seem to increase the applicant's chances of being admitted to the schools. However, having a low physics GRE score seemed to penalize otherwise competitive applicants (i.e., applicants with mid to high GPAs). Thus, our work suggests that the physics GRE does not, in fact, help applicants who might otherwise be missed stand out.

DOI: 10.1103/PhysRevPhysEducRes.17.010144

I. INTRODUCTION

While applying to graduate programs requires many components, perhaps none is as scrutinized as the Graduate Records Exam (GRE), and in physics, the physics GRE. Indeed, research into graduate admissions in physics suggests that the physics GRE is one of the most important components of the applications for determining which applicants will be admitted, based on both student and faculty perspectives [1,2] and analysis of the admissions process [3,4]. Despite its prominence in the admissions process, the physics GRE is known to be biased against women and people of color in physics [5], resulting in lower average scores compared to white and Asian males. At least one in three programs use a cutoff score [2], with 700 being a common choice [6], meaning applicants from groups already underrepresented in physics graduate programs can be further marginalized as they are less likely to achieve these scores. This is in addition to the observation that many physics students of color already see the GRE as a barrier to applying to graduate school [7–9].

Further, the physics GRE might not even be useful for determining which applicants will be successful in graduate school. For example, Miller *et al.*, suggest that the physics GRE is not useful for predicting which applicants will earn their PhDs [6]. Additionally, Levesque *et al.*, argue that using the common 50th percentile cutoff score for the physics GRE would have caused admissions committees to reject nearly 30% of students who would later receive a national prize postdoctoral fellowship, which can be viewed as a proxy for research excellence [10]. Yet despite evidence suggesting the physics GRE does not predict these typical ways of measuring "success" in graduate school and calls from the American Astronomical Society and the American Association of Physics Teachers to eliminate the physics GRE from admissions [11,12], most physics graduate programs still require applicants to submit their physics GRE scores. Currently, nearly 90% of physics and astronomy graduate programs still accept the physics GRE, with over half requiring or recommending submitting a score [11]. Of those that do not accept physics GRE scores from applicants, all of the programs are solely astronomy graduate programs or joint physics and astronomy graduate programs. While it is uncertain where removing the physics GRE affects any measure of graduate school success (e.g., completion rate), initial work by Lopez suggests that removing the physics GRE does increase the diversity of applicants [13].

Given these documented issues with the physics GRE, why do departments continue to use it? First, given that many programs are seeing a larger number of applicants, the physics GRE provides a quick way to filter the applications down to a more reasonable number for faculty review. Unlike in undergraduate admissions, graduate admissions tend to be decentralized and done at the departmental level by a faculty committee. Hence, faculty are asked to review applications in addition to their regular teaching and research duties and thus, might not have the time to read the letters of recommendation and applicant essays for every applicant.

Second, some faculty view GRE scores as measures of innate intelligence [3,14] or ability to become a Ph.D.-level scientist [5]. After all, they and other faculty likely had high GRE scores in order to be admitted to graduate school, and may exhibit a survivorship bias, believing that a high GRE score is needed to succeed. Further, physics is seen as a "brilliance-required" field, where innate intelligence is required for success [15].

A third argument, and the most interesting one in terms of the scope of this paper, is that standardized tests such as the physics GRE can help students stand out [16]. The ETS, the creator of the GRE and physics GRE, claims that subject GREs "can help you stand out from other applicants by emphasizing your knowledge and skill level in a specific area" [17]. For example, a student with an average grade point average (GPA) might be able to stand out from other applicants if they did exceptionally well on the physics GRE.

In addition, applicants from smaller universities or universities that are not known to the admissions committee might benefit from performing well on a standardized measure. For example, the ETS claims that the GRE provides a "common, objective measure to help programs compare students from different backgrounds" [18] and physics admissions committees worry that removing the GRE would limit their ability to compare applicants from different backgrounds [19]. Anecdotally, some faculty claim that a good physics GRE score could aid students from small liberal arts colleges in the admissions process [20].

We already know that GPAs are interpreted in context of the applicant's university. Posselt has shown that among more prestigious graduate programs, the applicant's GPA is viewed in the context of their undergraduate institution with high GPAs from prestigious institutions seen favorably, low GPAs from an unknown school as unfavorably, and high GPAs from unknown schools and middle GPAs from prestigious institutions in the middle [21]. Therefore, a standardized test such as the physics GRE could provide an assumed equal comparison for an admissions committee and might allow the applicant from an unknown school to stand out or have a similar chance of admission as an applicant from a more well-known school.

Finally, graduate admissions have been documented to be "risk adverse," where admissions committees select applicants most likely to complete their program [3,14]. As applicants from smaller universities may be judged based on how previously enrolled students from their university did in the program [21], a risk adverse admissions committee might be less likely to admit applicants from small universities whose students have previously struggled in their program. However, perhaps a high standardized test score could overcome these perceptions and signal that the applicant might indeed be successful in the program.

Our goal then is to focus on the third argument. Does the physics GRE help applicants "stand out" in the admissions process in practice? If that is the case, we would expect those disadvantaged in the admissions process, those who have low GPAs, attended a smaller institution, or identify as part of a group currently underrepresented in physics, to be admitted at similar rates as their more advantaged peers with similar physics GRE scores. Specifically, we ask the following:

(1) How does an applicant's physics GRE score and undergraduate GPA affect their probability of admission?
(2) How are these probabilities of admission affected by an applicant's undergraduate institution, gender, and race?

As Small points out in his critique of admissions and standardized test studies [22], multiple variables rather than just a standardized test might best explain our results and therefore, a framework that allows for substitutions and trade-offs between variables is necessary. Therefore, we ask an additional research question:

(3) How might the above relationships be accounted for through mediating and moderating relationships?

This paper is organized as follows: Sec. II provides an overview of mediation and moderation analysis. We then describe our data, how we determined what constitutes "standing out," and how we implemented mediation and moderation analysis in Sec. III. In Sec. IV, we describe our findings and in Sec. V, we use those findings to answer our research questions and explain our limitations and choices which may affect our results. Finally, we describe our future work in Sec. VI and the implications of our work for graduate admissions in physics in Sec. VII.

II. BACKGROUND

Before we can answer the third research question, it is important to describe what we mean by mediating and moderating relationships.

In a mediating relationship, two variables are only related because they are also related to some common third variable. For example, a student who played video games the night before an exam might do poorly because they stayed up playing video games too late and did not get enough sleep. Therefore, video games and doing poorly on the exam are only related due the common factor of lack of sleep. Lack of sleep is then a mediating variable.

In a moderating relationship, the strength of the relationship between two variables depends on some third variable. For example, the relationship between someone liking dogs and owning a dog likely depends on whether they are allergic to dogs. That is, we would expect someone who likes dogs but is allergic to dogs is less likely to own a dog than someone who likes dogs but is not allergic to dogs is. Being allergic to dogs is then a moderating variable.

Mathematically, suppose that some input X has an effect on output Y. We would say that some other input M mediates the relationship between X and Y if X only has an effect on Y because X has an effect on M and M has an effect on Y [23]. For a simple case, we can represent these relationships as

$$Y = i_1 + cX, \tag{1}$$

$$M = i_2 + aX, \tag{2}$$

$$Y = i_3 + c'X + bM, \tag{3}$$

where i represents the intercepts. These relationships are visually shown in Fig. 1.

Using this representation, the direct effect of X on Y is represented by c' and the indirect effect is represented by ab. The total effect is then $c' + ab$, which for a linear regression model, is equal to c. Equivalently, in the case the linear regression, the indirect effect is $c - c'$.

However, if Y is binary, linear regression is not appropriate and logistic regression should be used instead. In this case, Rijnhart *et al.*, recommend using ab as the indirect effect as their simulation studies found the ab estimate of the indirect effect exhibited less bias than the $c - c'$ estimate [24].

To determine if the indirect effect is statistically significant, a common approach is to use a Sobel test. However, simulations suggest that the Sobel test is underpowered and that bootstrapping is a good alternative [25]. Specifically, those simulations find that using the percentiles of a bootstrapped estimate of the indirect effect to estimate the confidence interval is a good compromise between avoiding type I errors while maintaining statistical power. In their approach (which has also been used in PER studies before, e.g., Ref. [26]), if ab is different than zero, then there is some degree of mediation.

More specifically, there are three cases.

(1) If $ab \neq 0$ and $c' = 0$ then M fully mediates the relationship between X and Y.

(2) If $ab \neq 0$ and $c' \neq 0$, then M partially mediates the relationship between X and Y. In that case, we can estimate the amount of mediation as the fraction of the total effect attributed to the indirect effect, $ab/(ab + c')$ [27,28].

(3) If $ab = 0$, then M does not mediate the relationship between X and Y.

This approach can also be adapted to multiple mediators and these mediators can be predictors of other mediators. An example of this serial mediation case with two mediators is shown in Fig. 2. Equations (1) to (3) can then be modified to be

$$Y = i_4 + cX, \tag{4}$$

$$M_1 = i_5 + a_1 X, \tag{5}$$

$$M_2 = i_6 + a_2 X + a_3 M_1, \tag{6}$$

$$Y = i_7 + c'X + b_1 M_1 + b_2 M_2. \tag{7}$$

In this case, there are three indirect effects. First, there are the indirect effects of the mediators individually, $a_1 b_1$ and $a_2 b_2$, and second, there is the indirect effect of the mediators together $a_1 a_3 b_2$. The total indirect effect is then $a_1 b_1 + a_2 b_2 + a_1 a_3 b_2$ [29].

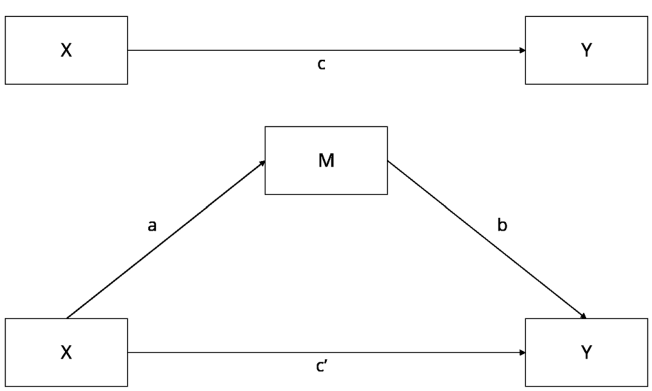

FIG. 1. Visual representation of Eqs. (1) to (3). The top graphic shows Eq. (1) while the bottom graphic shows Eqs. (2) and (3).

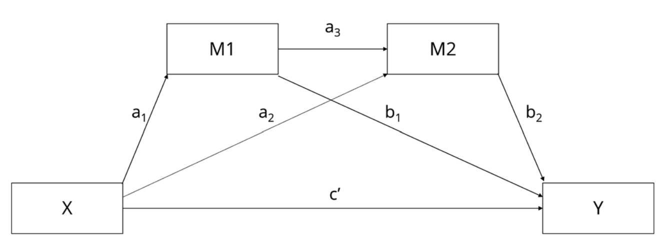

FIG. 2. Visual representation of Eqs. (5) to (7) showing serial mediation with two mediators.

More generally, for N mediators, we can generate $N+1$ equations where the first N are of the form

$$M_n = i_n + a_n X + \sum_{j=1}^{n-1} a_j M_j \tag{8}$$

and the final equation is of the form

$$Y = i_y + c'X + \sum_{j=1}^{N} b_j M_j. \tag{9}$$

So far, we have assumed that the relationship between the mediator M and the output Y does not depend on any other variables. However, it is possible that the relationship between M and Y could also depend on X or some other variable, meaning there is a conditional indirect effect (see Preacher *et al.*, [30]). In the case that the relationship between M and Y depends on X, we would say that X moderates the relationship between M and Y. Practically, this means we must add an interaction term to Eq. (3), which then becomes [30]

$$\begin{aligned} Y &= i_3 + c'X + b_1 M + b_1' XM \\ &= i_3 + c'X + (b_1 + b_1' X)M. \end{aligned} \tag{10}$$

We use the prime on b coefficients to denote an interaction coefficient for a mediator while an unprimed b coefficient is a coefficient of a mediator.

The conditional indirect effect is then $a(b_1 + b_1'X)$. If $b_1' = 0$, we would say that there is no moderation and the indirect effect is the standard ab.

In the case that there are multiple mediators, Eq. (10) can be modified to include multiple mediators and interaction terms for all pairs of variables where moderation may be of interest.

In the special case that X is binary, Eq. (10) reduces to $Y = i_{x=0} + b_1 M$ when $X = 0$ and $Y = i_{X=1} + (b_1 + b_1')M$ when $X = 1$. Therefore, to test if there is moderation, we can simply regress M on Y given $X = 0$ and again given $X = 1$ and subtract the slopes to calculate b_1' instead of including an interaction term in the model.

III. METHODS

A. Data

Data for this study come from the physics departments at five selective, research-intensive, primarily white universities. Four of these universities are public and part of the Big Ten Academic Alliance while the remaining university is a private Midwestern university. During the 2017–2018 and 2018–2019 academic years, graduate admissions committees at these five universities recorded all physics applicants' undergraduate GPA, GRE scores, undergraduate institution, and demographic information such as gender, race, and domestic status. In addition, the universities recorded whether each applicant made the shortlist, was offered admission, and whether the applicant decided to enroll. Because our study includes all applicants rather than only admitted applicants, we are unlikely to suffer from the range restrictions noted in critiques of other admissions studies (e.g., Refs. [22,31]). However, we do address a possible range restriction in Sec. V B.

Because of different requirements and admissions processes for international students and domestic students (e.g., international students need to submit a test of English proficiency), we only include domestic students in our study. We then remove any applicant for whom a physics GRE and GPA were not recorded, leaving us with 2537 applicants. While we in theory could use multiple imputations to address the data as Nissen *et al.*, recommends [32], faculty reviewing the applications do not, to our knowledge, do this and hence, we would be creating data that was not available in the admissions process. Distributions and analysis of the remaining physics GRE scores and GPAs appear in the Appendix.

As the applicant's undergraduate university does not contain meaning in itself, we needed to categorize the institutions. We chose to categorize the institutions by their size and their selectivity. We then used the number of physics bachelor's degrees awarded per year as a measure of the size of the university. We assume that universities with more graduates are more well known and hence, would likely be known to the admissions committees. In contrast, universities that produce fewer bachelor's degrees might not be known to the admissions committees and hence, might be unknown programs. It would then be these applicants from "smaller" programs who might need to "stand out." We acknowledge that some programs that produce a small number of physics bachelor's degrees each year might not be unknown to the admissions committees due to previous applicants from such schools or research collaborations or partnerships. However, there is no way in our data to know if this is the case.

To determine whether a university should be counted as a "small university" we used the undergraduate institution names to look up the number of typical physics bachelor's degrees from AIP's public degree data [33,34]. As of this writing, degree data for the 2018–2019 academic year was not available, so we used data from the 2016–2017 and 2017–2018 academic years to quantify the number of bachelor's degrees. Additionally, this would have been the most recent data available when admissions committees would have reviewed applications and many of the applicants would be represented in the data as bachelor's degree recipients

To account for the institution's prestige, we used Barron's selectivity index [35]. Barron's selectivity index is a measure based on the undergraduate acceptance rate of an institution as well as characteristics of its undergraduate incoming classes, such as mean SAT scores, high school

TABLE I. Summary of the comparisons we analyzed, which group needs to stand out and which does not, and the figure number showing the results.

Variable	Group that tends to be privileged in admissions	Group that tends not to be privileged in admissions	Results
Program size	Applicants from physics programs that rank in top 25% of programs based on yearly graduates	Applicants from physics programs that rank in bottom 75% of programs based on yearly graduates	Fig. 5
University selectivity	Applicants from universities ranked as most selective or highly selectively (Barron's value of 1 or 2)	Applicants from any other university (Barron's value of 3 or lower)	Fig. 6
Gender	Male applicants	Female applicants	Fig. 7
Race	Asian or white applicants	Black, Latinx, Multiracial, and Native applicants	Fig. 8

GPAs, and class rank. We assume selectivity is a proxy for prestige as prestigious institutions tend to have low acceptance rates and high SAT scores and GPAs from incoming students. In contrast to the AIP data, Barron's selectivity index applies to the institution as a whole rather than only the physics department.

B. Probability of admission procedure

Determining whether an applicant is more or less likely to be admitted first requires computing admissions probabilities. To do so, we grouped applicants based on their GPAs and physics GRE scores. Prior work has found that the physics GRE score and undergraduate GPA are two of the most important aspects of the applications [2–4]. Our previous work specifically found that the physics GRE score and undergraduate GPA were able to predict with 75% accuracy whether an applicant would be admitted to one public Midwestern physics graduate program.

In addition, physics is a "high consensus" discipline, meaning most programs agree on what consists of a successful applicant [3]. Therefore, despite many other components of the applications that affect whether an applicant will be admitted, we believe using the physics GRE score and undergraduate GPA provides a first-order overview of what admissions committees would use to admit applicants.

In order to ensure a reasonable number of applicants in each group to do meaningful analysis, we grouped applicants into bins based on their GPA and physics GRE score. We choose to use GPA bins 0.1 units in width and physics GRE bins 50 points in width. The GPA bins were selected to ensure that that GPAs with the same tenth digit were in a single bin. That is, 3.50 through 3.59 would be in a single bin. All GPAs were already reported on the 4.0 scale and physics GRE scores were reported using the standard 200-990 scale so we did not need to do any conversions.

We then computed the fraction of applicants in each bin who were admitted to the program they applied. As we are interested in applicants "standing out," we frame our results as whether applicants in a bin are admitted at a higher rate than the overall rate (all accepted applicants divided by all applicants). If applicants are admitted at a higher rate than the overall rate, it suggests that these applicants did in fact stand out to the admissions committee.

In our framing of standing out, we are assuming that graduate admissions operate under a deficit model. That is, due to their privilege, some applicants had better resources, opportunities, or choices available to them and as a result, may appear as better candidates for the program compared to applicants who did not have those available to them. To our knowledge, those with less privilege and/or resources are not directly compared to those with more privilege and/or resources but instead compared to an ideal applicant who often resembles someone from a more privileged background. For example, Owens *et al.* found that faculty valued advanced course knowledge and programming skills in incoming graduate students [36], which may be more characteristic of applicants coming from better resourced institutions. Using this framing, we created four groups of applicants who might or might not need to stand out, which are summarized in Table I and explained in detail below.

To take into account the size of the institution, we first used the AIP data to determine the national quartile each applicant's institution ranked in terms of all bachelor's degree recipients for each of the two years of data. Because not all institutions reported data in both years and the number of graduates could vary significantly between years, we conducted separate analyses first with the highest quartile an institution reached in the two years and second with the lowest quartile the program reached in the two years. For example, if an institution was ranked in the 3rd quartile the first year and the 4th quartile in the second year, our first analysis would use the 4th quartile and our second analysis would use the 3rd quartile. We then define the large programs as those in the 4th quartile and small programs as those in the 1st through 3rd quartiles. We address this choice in the discussion.

When using Barron's selectivity index to take into account the selectivity of the institution, we used Chetty *et al.*'s, [37] five groupings (Ivy League +, remaining most selective institutions, highly selective institutions, selective institutions, and nonselective institutions) as a guide. As there was a single applicant from a nonselective institution, selective and nonselective were grouped into a single category. Because we are interested in smaller, less-known

TABLE II. Counts of applicants by gender and race who provided both GPAs and physics GRE scores.

	Race							
Gender	Asian	Black	Latinx	Multi	Native	White	Unreported	Total
Men	247	49	99	166	4	1410	112	2087
Women	56	2	19	26	0	308	28	439
Unreported	1	0	0	1	0	5	4	11
Total	304	51	118	193	4	1723	144	2537

programs compared to larger, well-known programs, we took the selective and nonselective group to be our "less selective institution" group and institutions in the first three of Chetty *et al.*'s, categories as our "most selective institutions." This corresponds to grouping institutions with a Barron's index of 1 and 2 together as the most selective institutions and all other values together as the less selective institutions.

To understand how high physics GRE scores might help applicants identifying as part of a group currently underrepresented in physics, we compared women's admission probability to men's admission probability and applicants of color's admission probability to applicants not of color's admission probability. While it should be noted that gender is not binary [38], the data the admissions committee recorded are only in terms of the male and female binary and hence, we cannot comment on how high physics GRE scores may impact applicants of other genders.

Furthermore, given the limited number of applicants identifying as part of a racial group underrepresented in physics, we combined all Black, Latinx, Multiracial, and Native applicants into a single category, which we will refer to as B/L/M/N following the recommendation of Williams [39]. We acknowledge that this may obscure important distinctions between groups, as Teranishi [40] and Williams suggest. We also acknowledge applicants identifying as a marginalized gender and race may face additional barriers and hence could stand out differently than an applicant identifying as either a marginalized gender or race. However, there are less than 50 applicants (~2% of the sample) identifying as a member of both a marginalized gender and marginalized race, limiting statistical power for analysis. Full demographics are shown in Table II. For information about how race and ethnicity categories were constructed and standardized, see Posselt *et al.* [41] who previously used the 2017–2018 academic year application data from this study in their study.

C. Mediation and moderation procedure

Given that to some degree, both the physics GRE score and undergraduate GPA measure physics knowledge, we expect that these two measures will be correlated with each other. Therefore, we first tested whether the physics GRE has any mediating effects when predicting admission and whether GPA moderates the relationship between the physics GRE and admission; that is, is one only related to admission because it influences the other and that one influences admission or is the strength of the relationship between one and admission affected by the other. Because admissions status is a binary outcome variable, we need to use logistic regression for Eqs. (1), (3), and (10).

When taking an applicant's GPA and physics GRE score into account, we first centered and scaled both variables so they both have means of zero and variances of 1. As we are treating GPA and physics GRE scores as continuous, we can use linear regression for Eq. (2).

To estimate the coefficients in Eqs. (1) to (3) and (10), we generated 5000 bootstrap samples with replacement as was done in Hayes and Scharkow [25]. For each trial, we computed the indirect effect ab. To get the estimate of each parameter, we took the average of the 5000 bootstraps. To get the lower end of the 95% confidence interval, we used the value that corresponded to the 2.5th percentile of the values generated by the bootstrap. Likewise, to get the upper end of the 95% confidence interval, we used the value that corresponded to the 97.5th percentile.

For the institutional features, we treat institutional selectivity and institution size as binary input variables (most selective or less selective and larger institution or smaller institution) and for demographic features, we treat gender and race as binary variables. Again, we use B/L/M/N as one category for race and white and Asian as the other. The applicant's physics GRE score and GPA are again treated as continuous mediating and moderating variables.

Because the physics GRE score and GPA can both act as mediators and GPA may also influence the physics GRE score, we used a serial mediation model instead of the simple mediation model [Eqs. (4) to (7)]. While moderation by the independent variable X can occur for any of the relations between the other variables, only moderating relationships between GPA and admission and the physics GRE score and admission are within the scope of this work. Therefore, we only include those interaction terms in our models. For all of these analyses, we used the same bootstrapping process used for the simple mediation and moderation cases.

IV. RESULTS

A. Probability of admission results

When comparing the GPAs and physics GRE scores of all applicants, we notice that most applicants who are admitted have both high GPAs and high physics GRE scores (Fig. 3). Furthermore, while a near perfect GPA or physics GRE score resulted in the highest chance of admission, having either a high GPA or high physics GRE and a modest score on the other seemed to still offer an admission fraction around the overall average. However, having a low GPA or low physics GRE and a modest score on the other is usually grounds for rejection. Overall admissions fractions for a given physics GRE score or GPA are shown in the top and right margins of Fig. 3 respectively.

In regard to having a high physics GRE score despite a low GPA, we first note that only a small fraction of all applicants fall in this regime. Second, there appears to be no pattern in terms of higher than average fraction admitted for these applicants. Some combinations of low GPA and high physics GRE score result in a few applicants being admitted, and hence, an above average fraction of applicants being admitted, while other score combinations have no applicants being admitted, and hence, a below average change of admission. For example, having a GPA in the 3.3 bin and a physics GRE score in the 1000 bin resulted in an above average fraction admitted while having a GPA in the 3.4 bin and a physics GRE score in the 1000 bin did not result in an above average fraction admitted, despite the applicants having a higher GPA.

To further understand whether a high physics GRE score can highlight those with low GPA, we divided all students into either a high or low GPA and high or low physics GRE score bins, Fig. 4. Based on Fig. 3 in terms of admissions probabilities, a low GPA seems to be below a 3.5, while a high physics GRE score seems to be above 700. However, 700 is a common cutoff score which could explain why admissions probabilities increase after that score. Because hitting the minimum score might not catch the admission committee's eyes, we instead selected a higher score of 880 which represents the 80th percentile.

From Fig. 4, we notice two things. First, among applicants in the low GPA bin, less than half (44%) even make it above the typical cutoff score of 700 and less than 10% of those applicants with low GPAs score 880 or higher. These represent approximately 11% and 2% of all applicants, respectively. Comparing the fraction of admitted applicants in each bin, applicants with high physics

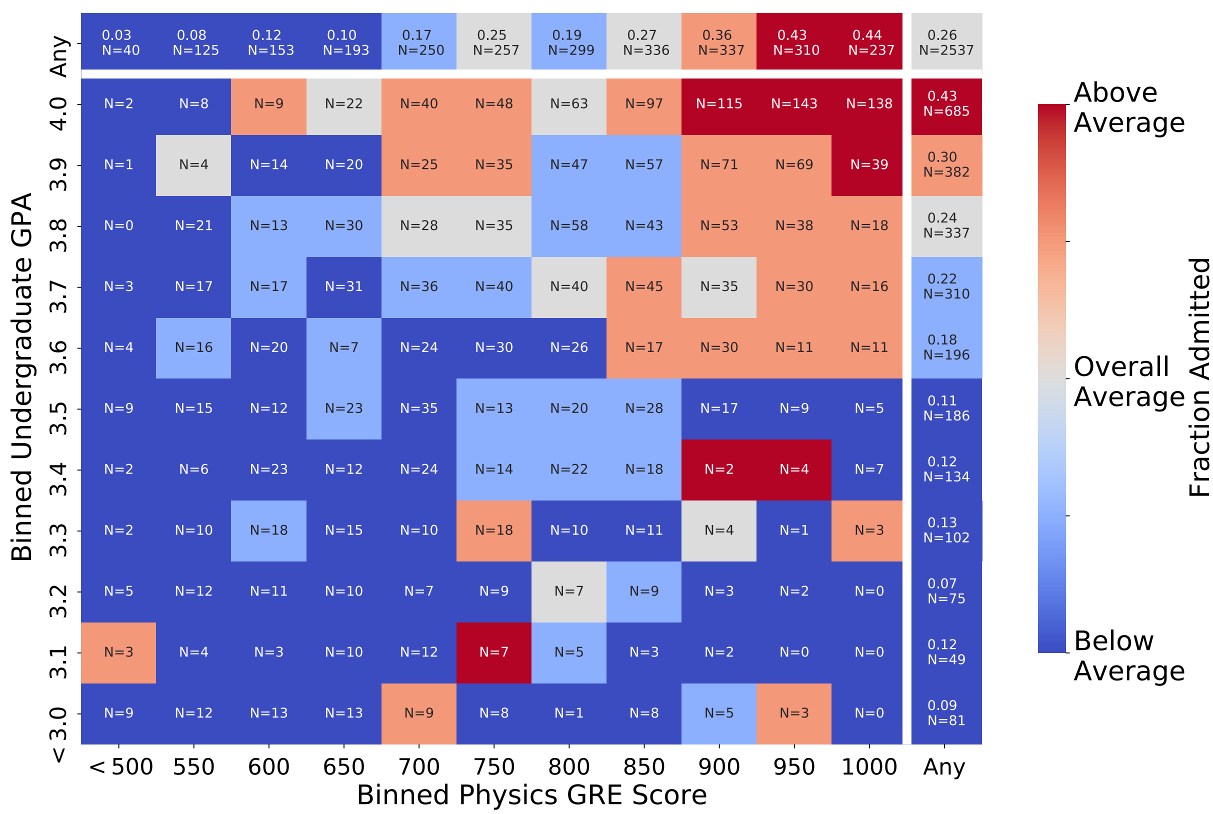

FIG. 3. Fraction of applicants admitted by undergraduate GPA and physics GRE score. The number of students in each bin is also shown. "Any" corresponds to the corresponding row or column totals. The bin label corresponds to the upper bound of values in the bin exclusive with the exception of the 4.0 GPA bin which includes 4.0. Values are colored based on whether they are above, below, or equal to the overall admissions rate. Admissions rates within 10% of the overall rate are colored the same as the overall rate. The above and below average colors are based on being above or below the midpoint between the max or min admission fraction and the overall average. These are based on raw numbers and not a statistical test.

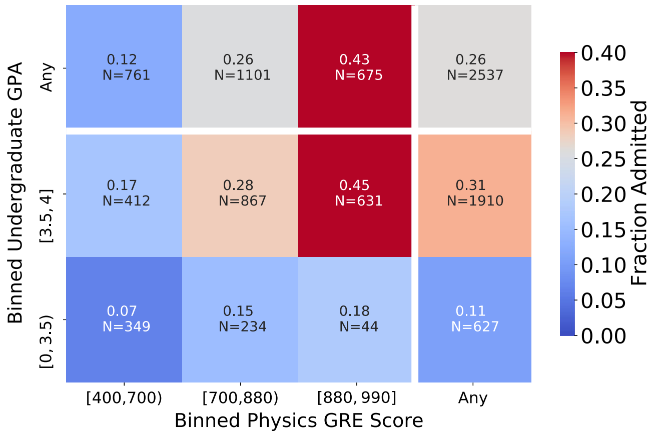

FIG. 4. A condensed version of Fig. 3 showing the fraction of applicants admitted by undergraduate GPA and physics GRE score.

GRE scores and low GPAs are admitted at nearly the same rate as applicants with high GPA and low physics GRE scores.

Second, we notice that 16% of all applicants score in the high GPA but low physics GRE score bin. That is, more applicants could be penalized for having a low physics GRE score despite a high GPA than could benefit from a high physics GRE score despite a low GPA.

When taking the size of the applicant's undergraduate program into account, (large or small), using either the highest or lowest quartile of bachelor's graduates over the two year period did not substantially change the results. Therefore, we only present results from the highest quartile reached, which are shown in Fig. 5. Because of the much smaller number of applicants per bin, we reduce the number of GPA and physics GRE bins. We use bins of 3.0 or less, which corresponds to a B or lower, 3.0 to up 3.3, a B+, 3.3 up to 3.7, an A−, and 3.7 up to 4, an A under the standard 4.0 scale.

Overall, by looking at the bin in the "Any" row and Any column of Figs. 5 and 6, we see that applicants from the largest undergraduate programs are nearly 40% more likely to be admitted (0.28 to 0.20) while applicants from selective institutions are nearly 70% more likely to be admitted (0.31 to 0.18). Looking at the individual admission fractions, there does not appear to be any advantage to applicants graduating from smaller institutions or less selective institutions. The physics GRE scores and GPAs where applicants are admitted at higher than average rates are the nearly same for large and small programs and selective and nonselective programs. Unsurprisingly, these tend to be higher physics GRE scores and higher GPAs. Outside of a few bins with a small number of applicants, no combination of low GPA (B+ or less) and high physics GRE score resulted in an above average admission fraction.

For the highest physics GRE scores, 900 and above, applicants from the largest or most selective universities seem to be admitted at a higher rate and a higher fraction of applicants from large or selective universities achieve these high scores compared to applicants from smaller universities. The fraction of applicants from both large universities, small universities, selective universities, and nonselective universities, as well as nationally, achieving each score is shown in Table III. Thus, it appears that even if higher scores did help applicants stand out, applicants from smaller and less selective schools most in need of standing out are less likely to achieve those scores in the first place.

Finally, the results from grouping by gender and race are shown in Figs. 7 and 8. Interestingly, we find that for most physics GRE scores, women are admitted at higher rates than men of equal score are. Likewise, we find that Black,

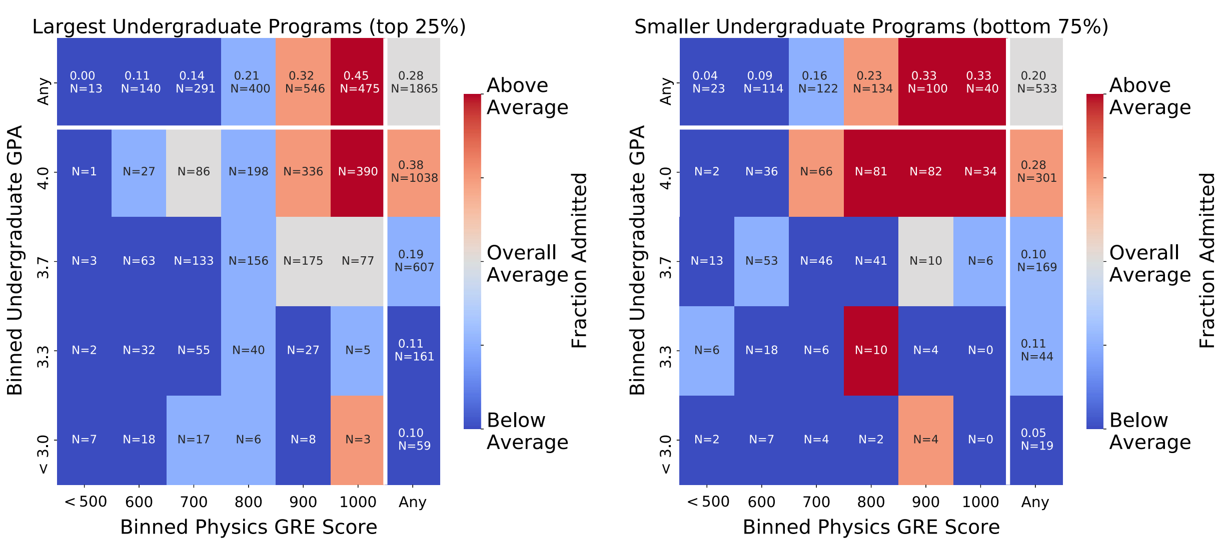

FIG. 5. Fraction of applicants admitted by undergraduate GPA and physics GRE score and split by large or small undergraduate university.

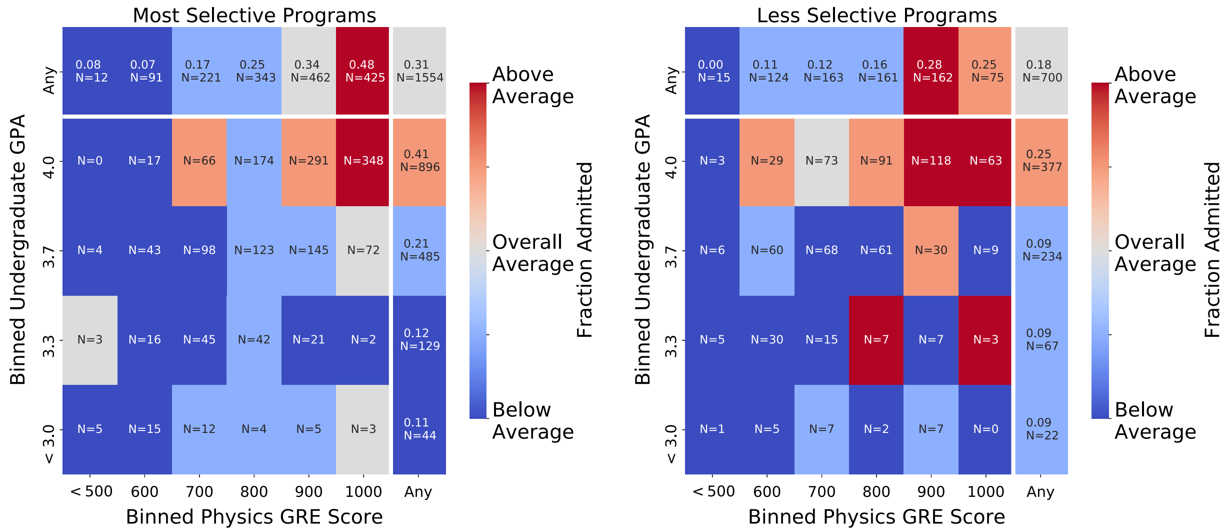

FIG. 6. Fraction of applicants admitted by undergraduate GPA and physics GRE score and split by selective or nonselective undergraduate university.

Latinx, Multiracial, and Native applicants are admitted at higher or similar rates as white or Asian applicants are for similar physics GRE scores. In addition, the same trend seems to hold for GPA as well. However, a high physics GRE score does not seem to help women with a low GPA. For B/L/M/N applicants, there appears to be a few places where applicants may stand out (such as the 800 physics GRE bin and 3.3 GPA bin). If these applicants were standing out due to their physics GRE score though, we would expect that pattern to continue for higher physics GRE scores but the same GPA. This does not appear to be the case, suggesting these applicants stood out for a reason other than their physics GRE scores. We address this in our discussion.

Because these applicants may have stood out for reasons other than their physics GRE score, we do not discuss any interactions between gender and race and selectivity and institution size. For completeness, plots showing these interactions are included in the Supplemental Material [42].

TABLE III. Distribution of applicants scoring in each physics GRE range by size of institution. ETS only publishes overall score distributions and hence, we cannot report national scores from only domestic students.

Score	Large schools	Small schools	Selective schools	Nonselective schools	National
[400,500)	0.7%	4.3%	0.8%	2.1%	9%
[500,600)	7.5%	21.4%	5.9%	17.7%	19%
[600,700)	15.6%	22.9%	14.2%	23.3%	20%
[700,800)	21.5%	25.1%	22.1%	23.0%	19%
[800,900)	29.3%	18.8%	29.7%	23.1%	16%
[900,990]	25.5%	7.5%	27.3%	10.7%	17%

B. Mediation and moderation results

1. Physics GRE and GPA

A visual representation of our mediation results with the physics GRE score and GPA is shown in Fig. 9. We find that all coefficients are statistically different from zero.

From Fig. 9, we see that an applicant's physics GRE score and GPA have about the same effect on whether the applicant is admitted. Given that applicants who had either a high physics GRE score or a high GPA had about the same chance of being admitted, this is not a surprising result.

Second, we find that the indirect effect is not zero, meaning that there is partial mediation. That is, whether an applicant is admitted depends on their physics GRE score and their GPA. In terms of the amount of mediation, we find that the indirect effect accounts for nearly 30% of the total effect.

Finally, doing moderation analysis, we find that $b_1' = 0.024(-0.114, 0.154)$. As zero is included in the confidence interval, we do not find evidence that GPA moderates the relationship between an applicant's physics GRE score and whether they are admitted. That is, the relationship between an applicant's physics GRE score and whether they are admitted is not influenced by their GPA.

2. Institutional features

A visual depiction of our results is shown in Figs. 10 and 11. We find that the applicant's physics GRE score partially mediates the relationship between the selectivity of their undergraduate institution and whether they were admitted and fully mediates the relationship between their institution's size and whether they were admitted. The fractions of mediation due to the indirect effects from the physics GRE score were

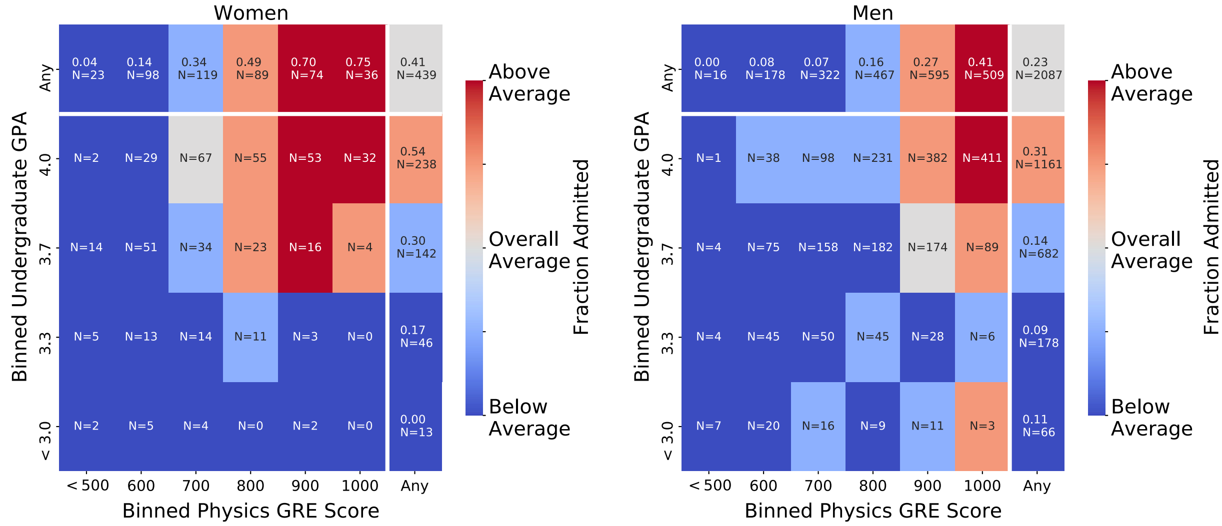

FIG. 7. Fraction of applicants admitted by undergraduate GPA and physics GRE score and split by the applicant's gender.

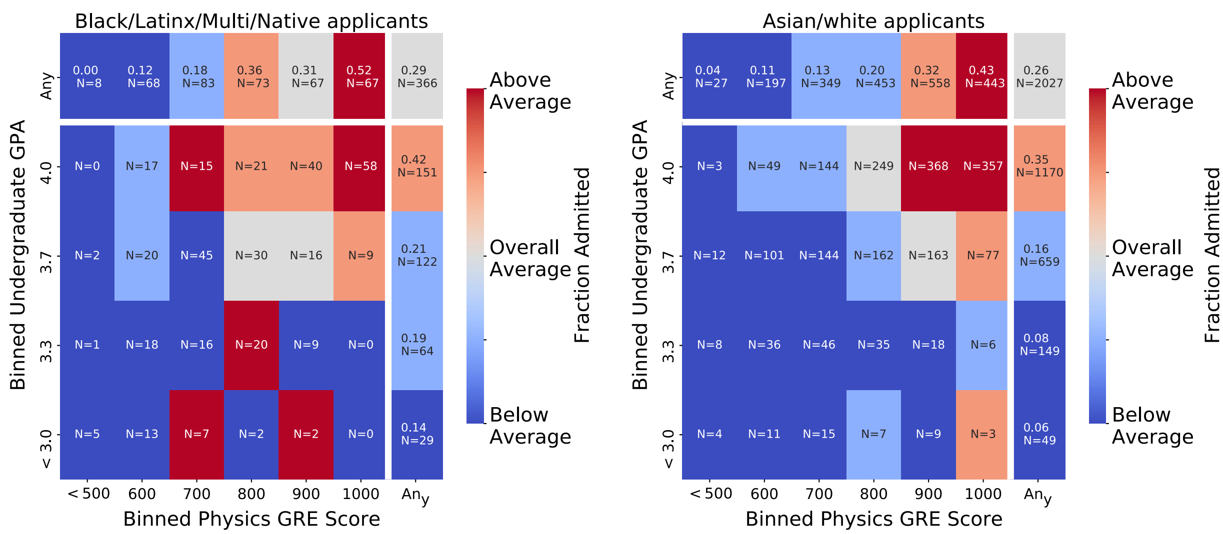

FIG. 8. Fraction of applicants admitted by undergraduate GPA and physics GRE score and split by the applicant's race.

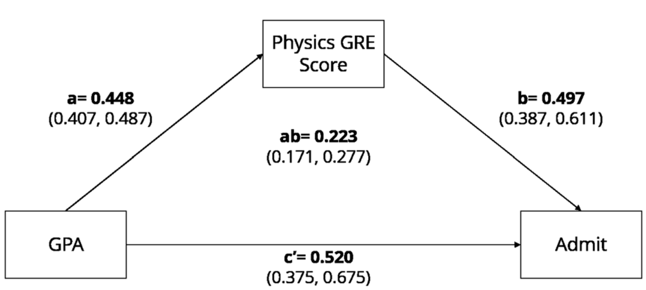

FIG. 9. Visual representation of the bootstrapped coefficients in Eqs. (1) to (3). We do find evidence of the physics GRE score mediating the relationship between GPA and admission status.

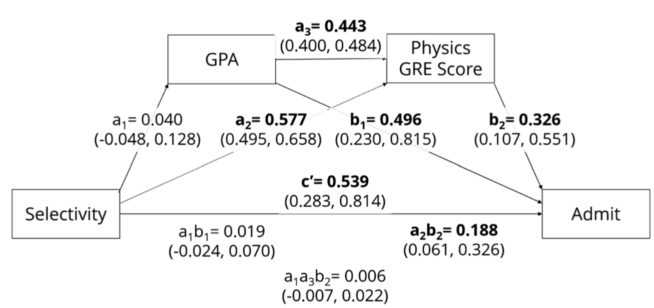

FIG. 10. Visual representation of the bootstrapped coefficients in Eqs. (5) to (7). We do find evidence of the physics GRE score mediating selectivity and admissions status but do not find evidence of GPA mediating selectivity and admissions status. We do not find evidence of a serial mediating relationship. Statistically significant coefficients are in bold.

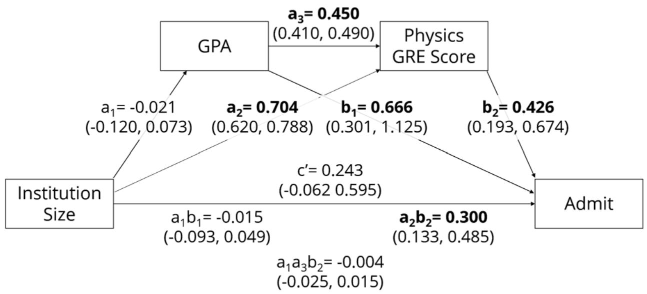

FIG. 11. Visual representation of the bootstrapped coefficients in Eqs. (5) to (7). We do find evidence of the physics GRE score mediating institution size and admission status but do not find evidence of GPA mediating institution size and admissions status. We do not find evidence of a serial mediating relationship. Statistically significant coefficients are in bold.

$$\frac{a_2 b_2}{|a_1 b_1| + |a_2 b_2| + |a_1 a_3 b_2| + |c'|} = 0.25 \text{ and } 0.533,$$

respectively.

In contrast, the applicant's GPA was not found to be a significant mediator in either case (zero was contained in the indirect effects' 95% confidence intervals), meaning that GPA is not a reason that there are differences in admission based on the applicant's undergraduate institution. Additionally, no serial mediation was observed for either case.

When looking at the results of the moderation analysis when the physics GRE is the mediating variable, we find that neither b'_2 value is statistically different from zero $[b'_{2,\text{selectivity}} = 0.136(-0.141, 0.411)$ and $b'_{2,\text{size}} = 0.080(-0.204, 0.361)]$, meaning that the relationship between the physics GRE score and admission is the same regardless of the type of institution the applicant attended.

Likewise, we do not find evidence of moderation when GPA is the mediation variable. In those cases, $b'_{1,\text{selectivity}} = 0.052(-0.314, 0.394)$ and $b'_{1,\text{size}} = -0.143(-0.630, 0.267)$.

3. *Demographic features*

Our results are shown visually in Figs. 12 and 13. Because we chose woman to be 1 and B/L/M/N to be 1 in our logistic regression equation, some of the coefficients are negative. For example, the negative a coefficient for gender and physics GRE score means that women score lower on the physics GRE than men do. Because the sign depends on our choice of which category should be 1 and are in that sense arbitrary, we use the absolute values of c' and $a_i b_i$ to calculate the fraction of mediation.

We find that the applicant's physics GRE score partially mediates the relationship between gender and admission but not race and admission meaning that gender affects admission in part because it affect physics GRE scores, which affect admission. The fraction of mediation for gender and admission due to the physics GRE score is

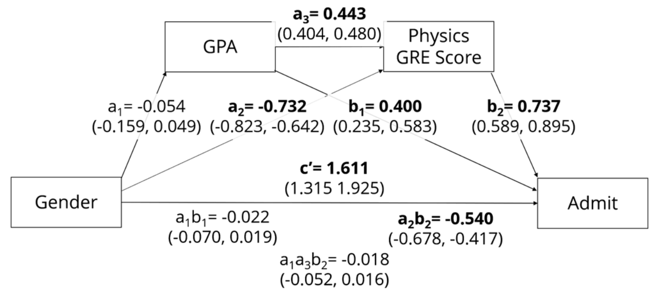

FIG. 12. Visual representation of the bootstrapped coefficients in Eqs. (5) to (7). We do find evidence of the physics GRE score mediating gender and admission status but do not find evidence of GPA mediating gender and admission status. We do not find evidence of a serial mediating relationship. Statistically significant coefficients are in bold.

$$\frac{a_2 b_2}{|a_1 b_1| + |a_2 b_2| + |a_1 a_3 b_2| + |c'|} = 0.246.$$

For GPA, we find the opposite. GPA partially mediates the relationship between race and admission but not gender and admission. The fraction of mediation for race and admission due to GPA is

$$\frac{a_1 b_1}{|a_1 b_1| + |a_2 b_2| + |a_1 a_3 b_2| + |c'|} = 0.299.$$

Likewise, we find a serial mediation effect for race and admission but not gender and admission. That is, admission is affected by race both because admission is related to GPA which is related to race and because admission is related to the physics GRE score which is related to GPA which is related to race.

When investigating whether any moderation effects exist, we do not find that to be the case. That is, we find that none of the interaction coefficients are statistically different from zero and hence, physics GRE scores and GPAs do not have a differential effect on admission based on the applicant's gender or race. Specifically,

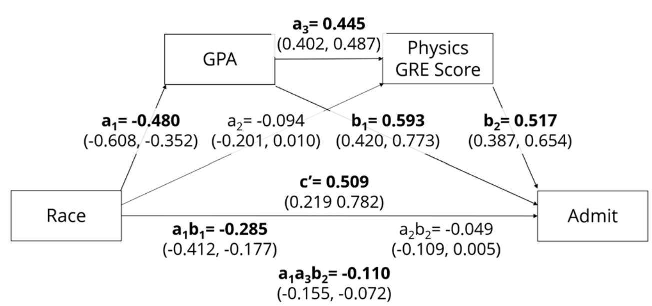

FIG. 13. Visual representation of the bootstrapped coefficients in Eqs. (5) to (7). We do find evidence of GPA mediating race and admission status and a serial mediation effect but do not find evidence of the physics GRE mediating race and admission status. Statistically significant coefficients are in bold.

TABLE IV. Summary of the mediating and moderation results. * signifies partial mediation is present, ** signifies full mediation is present, † signifies moderation is present. However no moderation effects were found.

Independent	Mediating	Indirect effect	Moderating effect
GPA	Physics GRE	0.223*	0.024
Selectivity	Physics GRE	0.188*	0.136
Selectivity	GPA	0.019	0.052
Selectivity	Serial	0.006	NA
Institution size	Physics GRE	0.300**	0.080
Institution size	GPA	−0.015	−0.143
Institution size	Serial	−0.004	NA
Gender	Physics GRE	−0.540*	0.154
Gender	GPA	−0.022	0.209
Gender	Serial	−0.018	NA
Race	Physics GRE	−0.049	−0.007
Race	GPA	−0.285*	−0.236
Race	Serial	−0.110*	NA

- $b'_{2,p\mathrm{GRE,gender}} = 0.154(-0.154, 0.462)$,
- $b'_{1,\mathrm{GPA,gender}} = 0.209(-0.103, 0.538)$,
- $b'_{2,p\mathrm{GRE,race}} = -0.007(-0.309, 0.314)$,
- $b'_{1,\mathrm{GPA,race}} = -0.236(-0.586, 0.143)$.

All results and interpretations from the mediation and moderation analyses are summarized in Table IV.

V. DISCUSSION

Here, we address each of our research questions and possible limitations or confounding factors.

A. Research questions

How does an applicant's physics GRE score and undergraduate GPA affect their probability of admission?—We find that scoring highly on the physics GRE and having a high GPA results in the highest chance of admission (Fig. 4). Likewise, having a low physics GRE score and low GPA results in the lowest chance of admission. If either the applicant's physics GRE score or GPA is high while the other is not, the chance of admission is approximately equal, regardless of which one is high.

However, the number of applicants with high GPAs but low physics GRE scores is 9 times as large as applicants with low GPAs and high physics GRE scores (i.e., scoring above the 80th percentile; Fig. 4). Even if we consider meeting the minimum cutoff score as a high physics GRE score, the number of applicants who have high GPAs but low physics GRE scores is 1.5 times greater than the number of applicants with low GPAs but high physics GRE scores. Thus, many more high GPA applicants could be penalized by the physics GRE than low GPA applicants could stand out or benefit from a high physics GRE score.

Finally, we note that for low-GPA applicants with high physics GRE scores, they are all essentially admitted at the same rate, regardless of whether they scored in the 700–870 range or the 880–990 range. If these applicants were standing out, we would expect low GPA applicants scoring above 880 to be admitted at a much higher rate than low GPA applicants scoring between 700 and 870. Thus, it is hard to determine if these applicants actually stood out to the committee or if they simply met the minimum physics GRE score needed for the committee to review the rest of the application.

How are these probabilities of admission affected by an applicant's undergraduate institution, gender, and race?—First, we find that for most physics GRE scores, applicants from larger and smaller institutions are admitted at similar rates (Fig. 5). However, for the highest scores (above 900), applicants from larger universities are admitted at higher rates. Interestingly, for applicants from smaller programs, scoring above 900 does not appear to provide any additional benefit in terms of the fraction of applicants admitted compared to scoring between 800 and 900.

In contrast, applicants from less selective institutions are less likely to be admitted than applicants from more selective institutions for all physics GRE scores above the common cutoff score (Fig. 6). That is, the physics GRE does not seem to counteract any potential biases from admissions committees toward applicants from less selective institutions.

Overall, attending a large or selective institution and scoring highly on the physics GRE does result in a higher chance of admission than scoring highly on the physics GRE and attending a smaller or less selective institution.

It is important to note that there might be selection bias in our data because test takers with high scores from smaller universities might not choose to apply to these schools. However, this seems unlikely because (i) these programs are highly regarded and hence, these would not be "safety schools" to high scoring applicants (as indicated by many high scoring applicants from large programs applying here) and (ii) while there is research suggesting students with low physics GRE scores might view their scores as barriers to applying [7], to our knowledge, there is no evidence that students with high scores do not apply to physics graduate programs. Given that students with low test scores might not apply, it is expected that our data are not representative of test takers on the lower end of scores (as shown in Table III).

When looking at the demographic variables, we find that women are admitted at higher rates than men with similar scores (Fig. 7) and B/L/M/N applicants are also admitted at higher rates than white or Asian applicants (Fig. 8). As prior work has shown [5], women and B/L/M/N test takers tend to score lower than white men on the physics GRE and hence, scoring highly could cause these applicants to stand out to admissions committees.

How might the above relationships be accounted for through mediating and moderating relationships?—Our mediation and moderation analysis further supports the results found through the probability of admissions procedure.

We find that the physics GRE score and GPA have similar regression coefficients when modeling admission, suggesting they have similar effects (Fig. 9) and that there is a mediation effect. In addition, we did not find any evidence of moderation. That means the relationship between GPA and admission is not different due to the applicant's physics GRE score. If a high physics GRE score did help a low-GPA applicant stand out, we would expect to see a moderation effect.

Combining the results of probability of admission analysis and the mediating and moderation analysis, we find that there is mediation but no moderation between an applicant's physics GRE score and their GPA when it comes to admission probability. In practice then, an applicant with a low GPA cannot simply overcome that low GPA by scoring highly on the physics GRE.

When we performed mediation analysis on the institutional factors, we found that the relation between institutional selectivity and admission was partially mediated by the applicant's physics GRE score and the relation between institutional size and admission was fully mediated by the applicant's physics GRE score (Figs. 10 and 11). Neither of these relationships was mediated by the applicant's GPA or serially, however.

The results of the mediation analysis show that physics GRE scores seem to explain some of the differences in admission probability based on the applicant's undergraduate institution. Therefore an applicant from a smaller or less selective institution may be able to stand out by scoring highly on the physics GRE. However, looking at the fraction of applicants admitted by physics GRE scores, especially the highest scores, suggests that is not what happens in practice.

In terms of gender and race, we do find some mediating relationships, but no moderation relationships (Figs. 12 and 13). We find that the physics GRE partially mediates the relationship between gender and admission. We also find GPA and GPA plus the physics GRE score partially mediates the relationship between race and admission. That is, some of the differences in admission rates between men and women can be explained by the differences in their physics GRE scores and some of the differences in admission rates between B/L/M/N applicants and non-B/L/M/N applicants can be explained by differences in their GPAs or physics GRE scores and GPAs.

These results then suggest that a female or B/L/M/N applicant may be able to stand out by doing well on the physics GRE. In practice, the probability of admission results do suggest that women and B/L/M/N applicants are admitted at higher rates than their male, white, or Asian peers are. However, as the five programs studied here were interested in increasing their diversity, our data do not allow us to disentangle standing out from highlighting. Therefore, our results should be interpreted with caution regarding any claims that the physics GRE may help applicants from groups underrepresented in physics stand out.

It should also be noted that women and B/L/M/N applicants are less likely to reach these higher scores than their male, white, and Asian peers. In our data, 75% of men and 72% of white or Asian applicants scored above 700 compared to 45% of women and 57% of B/L/M/N applicants. Thus, even if the physics GRE does allow these applicants to stand out, any potential benefit must be weighed against known scoring discrepancies.

Finally, it could be argued that even though we did not show that the physics GRE helps these applicants stand out, doing well on the test could still provide some benefit for them in the admissions process. We would agree with that argument not because of any properties of the test but because of the structure of graduate admissions in physics. In theory, any part of the application could be weighted highly and therefore, doing well on that part would provide some benefit. Given that prior work has established that the physics GRE is weighted highly [1–4], we would expect that good performance on the test would provide some benefit to applicants. Our goal, however, was not to determine whether a high physics GRE score benefits applicants in any capacity. Instead, our goal was to determine whether a high physics GRE score offers a disproportional benefit that would justify using it in graduate admissions given the disproportional harms the physics GRE can cause, which we were unable to show in practice.

B. Limitations and researcher decisions

Data biases.—As previously noted, applicants with lower physics GRE test scores may be less likely to apply, resulting in an overrepresentation of high scoring applicants. In addition, the programs in this study are well-regarded programs and there is likely a secondary bias toward applicants with high GPAs and high physics GRE scores applying overall. As a result, the results may not generalize to graduate programs whose applicants tend to have lower GPAs or low physics GRE scores.

In addition, it is possible that an applicant could be represented multiple times in the dataset, as an applicant could have applied to more than one of the five universities in this study. However, each applicant applies to each program independently and thus, we can treat them as separate events for the admissions probabilities. On the other hand, results based on distributions such as Table III and Figs. 17 and 18 would be affected by the duplicates.

To see if possible duplicates affected our results, we compared the distributions with and without possible duplicates. We assumed an application represented the same applicant and hence was a duplicate if two records had the same physics, verbal, written, and quantitative GRE scores, GPA, undergraduate university, and demographic features as the chance of all of these matching for a nonduplicate seems exceedingly low.

When we compare the distributions both with and without possible duplicates, Kolmogorov-Smirnov tests

[43] suggest the distributions are not significantly different. Therefore, because we cannot actually determine which applicants are duplicates and excluding possible duplicates does not change our results, we did not remove possible duplicates.

Our choice of low GPA and high physics GRE.—While percentiles are available for the physics GRE, a "high score" is left to interpretation. Even among admissions committees, individual members may have different ideas of what a high score is. In our work, we have taken the common cutoff score of 700 as the minimum possible high score [6]. Even around this minimum score, the number of applicants with low GPAs who could benefit from scoring highly on the physics GRE is less than the number of high GPA applicants who could be penalized by having a score below the cutoff.

We find that the number of low GPA applicants who could benefit from a high physics GRE score is greater than the number of high GPA applicants who could be penalized by a low score when the high score cutoff is less than or equal to 670, which is lower than the typical cutoff score and is around the 43rd percentile (Fig. 14). Assuming a high score should be at least above the 50th percentile, our specific choice of a high score does not affect our result that more applicants could be penalized than could benefit.

The previous argument is also affected by what we consider a high GPA. We have chosen any GPAs less than 3.5 to be low based on the results shown in Fig. 3 where applicants with GPAs at or above 3.5 are nearly twice as likely to be admitted to as applicants with GPAs below 3.5. If we were to pick a lower threshold, there would be even fewer applicants in the low GPA-high physics GRE score group and more applicants in the high GPA-low physics GRE score group, meaning even more applicants would

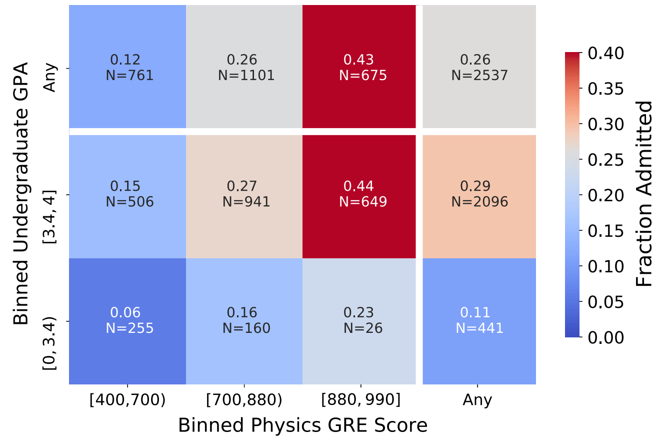

FIG. 15. A revised version of Fig. 4 showing the fraction of applicants admitted by undergraduate GPA and physics GRE score when the cutoff score for a high undergraduate GPA is 3.4.

possibly be penalized rather than standout. Using a GPA cutoff of 3.4 instead of 3.5, the ratio of applicants who could be penalized compared to stand out changes from the original 9∶1 to nearly 19∶1 (Fig. 15).

If we instead picked a higher GPA such as 3.6, there would be more applicants who could potentially benefit, but even then, the number of applicants who could benefit is only greater than the number of applicants who could be penalized around a physics GRE score of 730, which is not a high physics GRE score (approximately 54th percentile) and does not significantly change our results (Fig. 16). If we were to pick an even higher GPA cutoff, we could be hard pressed to justify why anything other than an "A" GPA is considered a low GPA, especially because admissions committees seem to group applicants with GPAs between

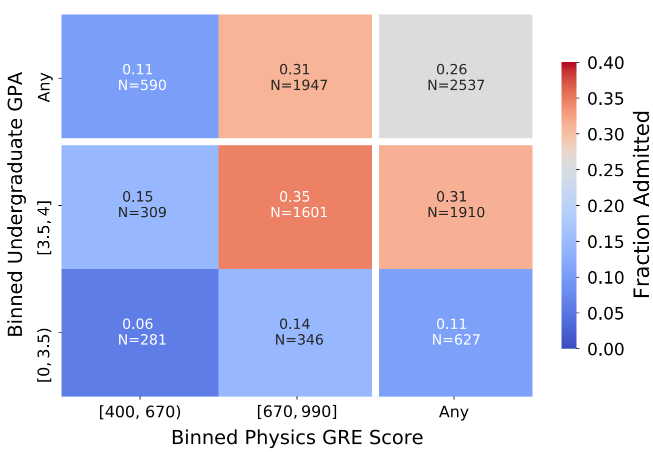

FIG. 14. A revised version of Fig. 4 showing the fraction of applicants admitted by undergraduate GPA and physics GRE score when the cutoff score for a high physics GRE score is 670. Here, the number of applicants who could benefit from a high physics GRE score is approximately equal to the number of applicants who could be penalized by a low physics GRE score.

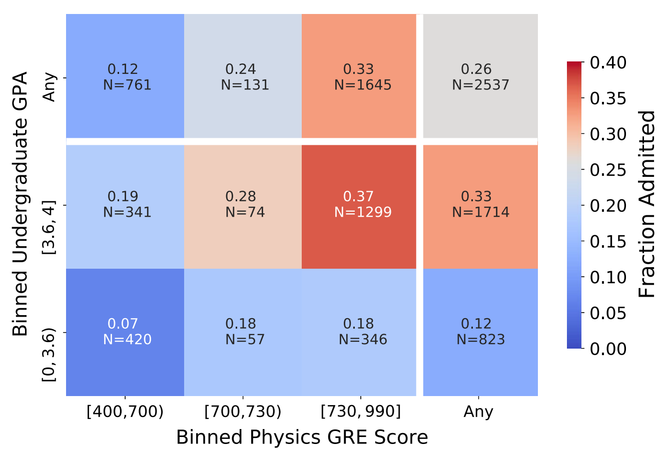

FIG. 16. A revised version of Fig. 4 showing the fraction of applicants admitted by undergraduate GPA and physics GRE score when the cutoff score for a high undergraduate GPA is 3.6. Here, the number of applicants who could benefit from a high physics GRE score is approximately equal to the number of applicants who could be penalized by a low physics GRE score.

3.5 and 3.6 more closely with applicants with GPAs between 3.7 and 3.8 than applicants with GPAs between 3.4 and 3.5 (based on the fraction of applicants admitted).

Based on our data and the fact that some universities use 3.5 as the only separation between 3.0 and 4.0, using 3.5 seems to represent the best option for separating high and low GPA students. Using any other choice either strengthens our claims or seems unrealistic to use as a cutoff.

Our choice of nonselective school.—We choose to follow a modified version of Chetty *et al.*'s groupings of programs [37]. However, many large, state universities have a Barron's selective index of 3 and fall in Chetty *et al.*'s fourth group. For our analysis, we would have included these large, state institutions as part of the less selective programs. As we are concerned with whether the physics GRE helps applicants stand out, saying applicants from large, state universities (for example, the University of Colorado-Boulder, the University of Washington, and Michigan State University) may fall in the traditionally missed category may not be correct.

We reran the analysis with these large, state institutions as part of what we called the most selective programs. We find that the conclusions are then more aligned with the large vs small program results. Using this grouping, applicants from less selective programs are admitted at similar rates to applicants from more selective programs for most physics GRE scores. However, applicants from more selective institutions with physics GRE scores above 900 are still more likely to be admitted than applicants from less selective institutions with similar physics GRE scores.

In terms of the mediating and moderation analysis, our results would be strengthened under this choice. While the physics GRE score would no longer mediate the relationship between selectivity and admission status, it would moderate the relationship [$b_2' = 0.311(0.015, 0.612)$]. This positive moderation means that the physics GRE score has a greater effect on admission status for applicants from more selective programs. In terms of standing out arguments, the positive moderation result means that doing well on the physics GRE would provide more of a benefit to applicants from more selective universities and not to applicants from smaller programs who are the intended beneficiaries of the standing out argument.

Thus, even though the details change, the overall conclusions are not weakened by changing our groupings. In fact, changing the groupings may strengthen our conclusions instead.

Our choice of a "small" school.—We chose small schools to be any university not in the top quartile of yearly bachelor's degrees awarded. We acknowledge that using quartiles is an arbitrary decision. However, when we used halves instead of quartiles to divide large and small schools, our results were unchanged, both in terms of the probability of admission analysis and the mediation and moderation analysis. Using the bottom quartile as small schools and all other programs as large schools would not have yielded insightful results as less than 2% of applicants would have attended a small school under this choice.

Of the possible physics specific measures, the number of bachelor's degrees seems most appropriate because programs with more graduates are more likely to be known by admissions committees simply because there are more students to apply from those programs. For example, the programs in the top quartile by number of bachelor's graduates produce nearly two-thirds of all physics bachelor's graduates [33,34]. In addition, we assume that programs with strong physics reputations attract more students and hence, produce more graduates. While this is likely to be more true at the graduate level, not all physics programs offer graduate degrees and hence, using the number of Ph.D.s awarded would not be useful. Thus, we believe the number of bachelor's graduates serves as a rough proxy for physics reputation.

VI. FUTURE WORK

While the five universities included in this study were interested in increasing their diversity and reducing inequities in their programs, their admissions processes still resembled the traditional metrics-based admissions model. Recently, many programs, including the ones studied here, have begun to employ holistic admissions, which looks at the overall application, taking into account noncognitive competencies and contextualizes the accomplishments of the applicant in terms of the opportunities that were available to them [44,45]. Often these holistic admissions use rubrics to weight the various components of each applicant (e.g., see Refs. [46,47]). Evidence from biomedical science graduate programs suggests that the GRE can even be included in holistic admissions without reproducing its known gender and racial biases [48]. Furthermore, their two-tiered approach to holistic admissions did not significantly increase the workload of admission committee members. These findings could persuade faculty reluctant to remove GRE due to its ease and supposed ability to measure some innate quality to try holistic admissions. Whether these results would hold for decentralized admissions as is typical in physics and for the physics GRE though are still open questions.

Our future work will then examine how our results may be affected when a department uses holistic admissions. In theory, we should no longer see the discrepancies between admitted applicants from large and small programs and more selective and less selective universities. In addition, the sample rubric developed by the Inclusive Graduate Education Network (as shown in Ref. [46]) suggests ranking applicants by high, medium, or low on each part of their applicant. Therefore, we would expect to see a flatter distribution of admission fractions based on physics GRE scores because, for example, all scores within the "high" range should be treated equally in the admissions process. Our future work will determine if this is indeed the case.

VII. CONCLUSION AND IMPLICATIONS

Our work suggests that, in practice, scoring highly on the physics GRE does not help applicants from small or less selective schools or applicants with a low GPA stand out. Indeed, having a high physics GRE and low GPA is no better than having a low physics GRE score and high GPA in terms of the fraction of applicants admitted. Similarly, for average physics GRE scores, the selectivity or size of the applicant's institution does not offer any advantage. For the highest scores though, attending a smaller or less selective institution does appear to result in an admissions penalty.

We also find that women and B/L/M/N applicants do have higher rates of admission based on physics GRE scores. However, given that the departments included in this study were actively trying to improve the diversity of their graduate student population [41], we are unable to attribute that standing out to the physics GRE.

While ETS's claim that the physics GRE can help applicants stand out from other applicants may be true in theory, we do not find evidence to support that claim in practice. In fact, our results suggest the opposite: the physics GRE may penalize applicants due to a low score rather than help applicants due to a high score.

As Small points out, facts and data do not unambiguously prescribe a course of action [22] and as others have noted, making such courses of action require a framework of assumptions and commitments [49]. Thus, we do not make a specific recommendation regarding whether the physics GRE should be kept or removed as a result of our work because the answer to that question depends on the priorities of the department. However, if departments are using the physics GRE to identify applicants who might be missed by other metrics to achieve their admissions priorities, we suggest against this practice as it does not appear to be backed by evidence.

ACKNOWLEDGMENTS

We would like to thank Julie Posselt, Casey Miller, and the Inclusive Graduate Network for providing us with the data for this project. We would also like to thank Rachel Henderson for her feedback on this project and guidance on the mediation and moderation analysis. This project was supported by the Michigan State University College of Natural Sciences and the Lappan-Phillips Foundation.

APPENDIX: ANALYSIS OF THE VARIABLES

In this Appendix, we describe the data used to answer research questions 2 and 3 to give the reader a better idea of the distributions of physics GRE scores and GPA in the dataset. Because the data are skewed left and exhibit ceiling effects (many applicants have 4.0 GPAs or 990 physics GRE scores), quartiles are used to describe the various features. To maximize the amount of information shown about the data, we use raincloud plots [50,51], which show the distribution, the density plot, and traditional box plot. Kolmogorov-Smirnov tests suggest the distributions are not significantly different whether we include applicants who may have applied to multiple schools in our dataset, so we include possible duplicates in our analysis.

Figure 17 shows the physics GRE scores and undergraduate GPAs of each applicant based on whether they attended a large undergraduate physics program (top 25% nationally in yearly physics bachelor's degrees) or attended a selective university (categorized as most competitive or highly competitive based on Barron's selectivity index). We notice that the physics GRE score distributions are shifted to the right for applicants from large physics departments or selective institutions, signifying higher scores. Indeed, the median physics GRE scores of applicants from large programs or selective institutions are nearly 100 points higher than those of applicants from smaller or less selective institutions. However, in terms of GPA, the median GPA is approximately the same, regardless of whether the applicant graduated from a larger or smaller physics department or attended a more or less selective institution.

Figure 18 shows the physics GRE and undergraduate GPAs by gender and race. As expected, men score higher on the physics GRE than women do, and Asian and white applicants score higher than Black, Latinx, Multiracial, or Native applicants, though the gaps appear larger than those reported in Ref. [6].

When comparing GPAs, we find that men and women have similar GPAs, as recently reported in Ref. [52] when comparing men and women's STEM GPAs. Likewise, our data also show a racial GPA gap with non-B/L/M/N applicants having a median GPA higher than that of B/L/M/N applicants by 0.15.

When looking across both figures, we notice that the physics GRE score distributions of smaller and less selective programs resemble the physics GRE distributions of women and B/L/M/N applicants while the physics GRE score distributions of the largest and most selective programs resemble the physics GRE score distributions of men and non-B/L/M/N applicants. To see if gender and race are confounding variables in our analysis, we examined the fraction of women and B/L/M/N applicants in each group. If this were the case, the smaller and less selective programs should have a greater fraction of women and B/L/M/N applicants than the larger and more selective programs.

However, we did not find this to be the case. Applicants from more selective institutions were 16% women while applicants from less selective institutions were 18% women (15% and 14%, respectively, for B/L/M/N applicants). For institution size, applicants from larger institutions were

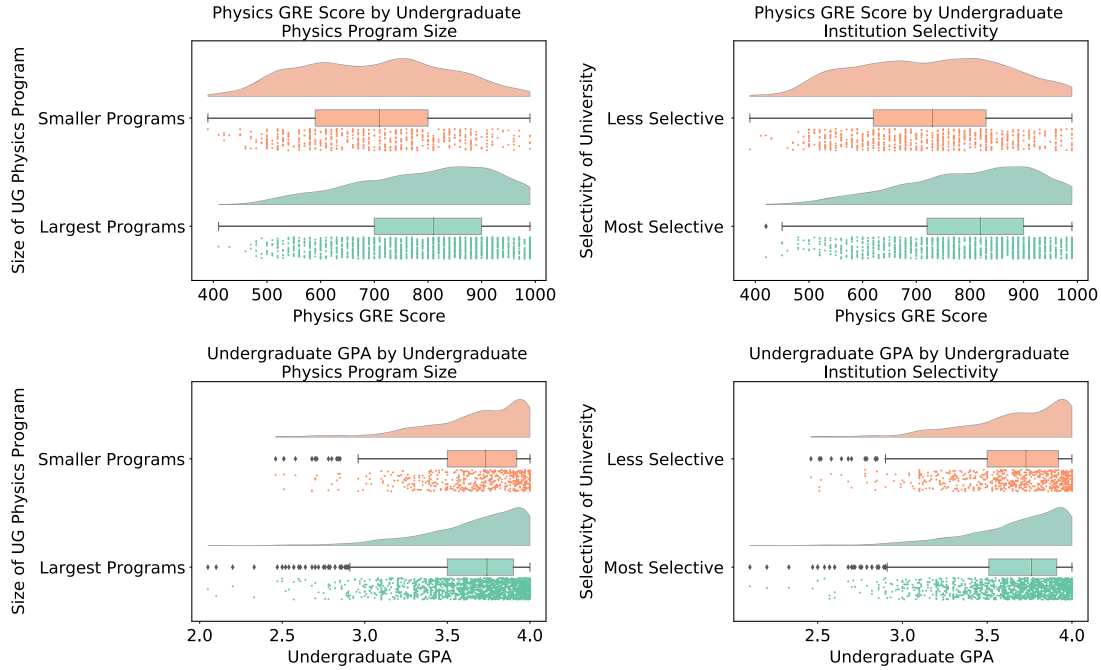

FIG. 17. Distribution of physics GRE scores and undergraduate GPAs by the size of the undergraduate physics program and institutional selectivity for each applicant.

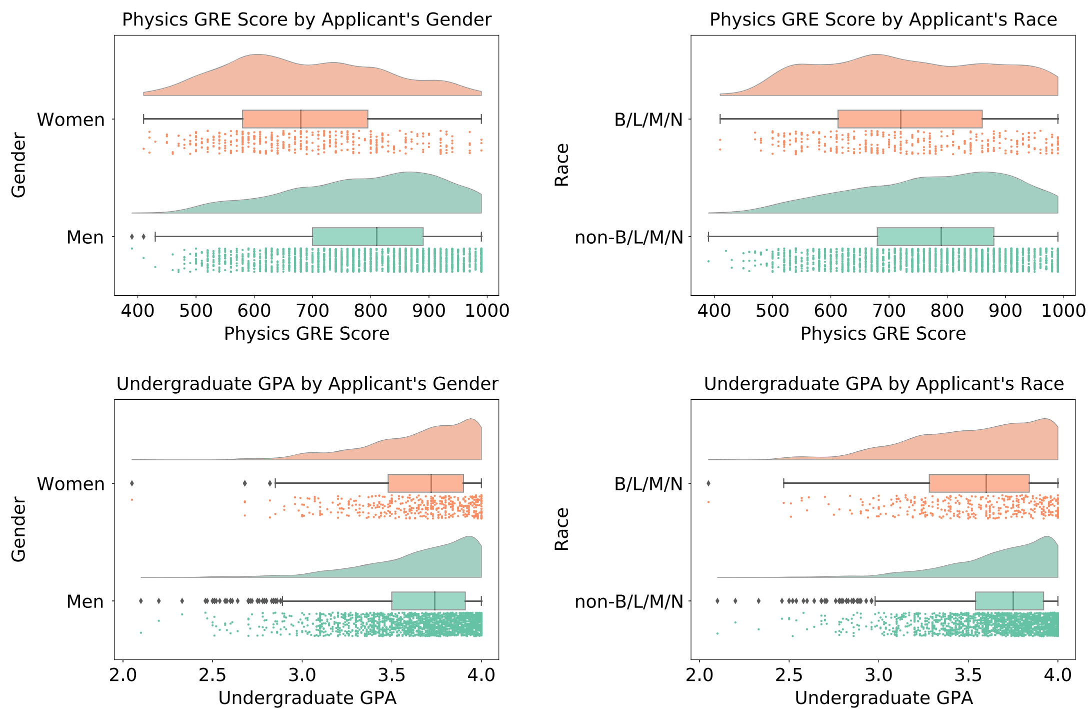

FIG. 18. Distribution of physics GRE scores and undergraduate GPAs by gender and whether the applicant identified as a member of a racial or ethnic group currently underrepresented in physics.

16% women compared to 21% women from smaller institutions (14% and 17% for B/L/M/N applicants, respectively). Thus, it does not appear that differences in who attends (in terms of gender and race) larger or more selective institutions are responsible for the observed differences in scores.

[1] Deepa Chari and Geoff Potvin, Understanding the importance of graduate admissions criteria according to prospective graduate students, Phys. Rev. Phys. Educ. Res. **15**, 023101 (2019).

[2] Geoff Potvin, Deepa Chari, and Theodore Hodapp, Investigating approaches to diversity in a national survey of physics doctoral degree programs: The graduate admissions landscape, Phys. Rev. Phys. Educ. Res. **13**, 020142 (2017).

[3] Julie R. Posselt, *Inside Graduate Admissions* (Harvard University Press, Cambridge, MA, 2016).

[4] Nicholas T. Young and Marcos D. Caballero, Using machine learning to understand physics graduate school admissions, in *Proceedings of the 2019 Physics Education Research Conference, Provo, UT* (AIP, New York, 2020).

[5] Casey Miller and Keivan Stassun, A test that fails, Nature (London) **510**, 303 (2014).

[6] Casey W. Miller, Benjamin M. Zwickl, Julie R. Posselt, Rachel T. Silvestrini, and Theodore Hodapp, Typical physics Ph.D. admissions criteria limit access to underrepresented groups but fail to predict doctoral completion, Sci. Adv. **5**, eaat7550 (2019).

[7] Geraldine L. Cochran, Theodore Hodapp, and Erika E. Alexander Brown, Identifying barriers to ethnic/racial minority students‘ participation in graduate physics, in *Proceedings of the 2018 Physics Education Research Conference, Cincinnati, OH* (AIP, New York, 2018), pp. 92–95.

[8] Raeshanda Wilson, Predicting Graduate School success: A critical race analysis of the Graduate Record Examination, Doctor of Education in Secondary Education Dissertations, 2020, https://digitalcommons.kennesaw.edu/seceddoc_etd/22.

[9] Lindsay M. Owens, Benjamin M. Zwickl, Scott V. Franklin, and Casey W. Miller, Physics GRE requirements create uneven playing field for graduate applicants, in *Proceedings of the 2020 Physics Education Research Conference, Virtual Conference* (AIP, New York, 2020), pp. 382–387.

[10] Emily M. Levesque, Rachel Bezanson, and Grant R. Tremblay, Physics GRE scores of prize postdoctoral fellows in astronomy, arXiv:1512.03709.

[11] American Association of Physics Teacher, *Statement on the Use of the GRE in Admissions to Graduate Physics Programs* (American Association of Physics Teacher, College Park, MD, 2019).

[12] American Astronomical Society, *AAS Statement on Limiting the Use of GRE Scores in Graduate Admissions in the Astronomical Sciences* (American Astronomical Society, Washington, DC, 2018).

[13] Laura A. Lopez, *Demographic Effects of Removing the Physics GRE Requirement in Graduate Admissions* (2019).

[14] Rachel E. Scherr, Monica Plisch, Kara E. Gray, Geoff Potvin, and Theodore Hodapp, Fixed and growth mindsets in physics graduate admissions, Phys. Rev. Phys. Educ. Res. **13**, 020133 (2017).

[15] Sarah-Jane Leslie, Andrei Cimpian, Meredith Meyer, and Edward Freeland, Expectations of brilliance underlie gender distributions across academic disciplines, Science **347**, 262 (2015).

[16] Katie Langin, A wave of graduate programs drops the GRE application requirement, Science https://doi.org/10.1126/science.caredit.aay2093 (2019).

[17] About the GRE Subject Tests (For Test Takers), https://www.ets.org/gre/subject/about.

[18] GRE Guide to the Use of Scores (2019), https://www.ets.org/s/gre/pdf/gre_guide.pdf.

[19] Nancy D. Morrison, William V. Dixon, Casey W. Miller, and The Women in Astronomy IV Graduate School Admissions White Paper Group, Women in astronomy IV white paper: Graduate admissions in a post-GRE world, Bull. AAS **51** (2020).

[20] Emily M. Levesque, Rachel Bezanson, and Grant R. Tremblay, Why astronomy programs are moving on from the physics GRE, Phys. Today https://doi.org/10.1063/PT.5.9090 (2017).

[21] Julie R. Posselt, Trust networks: A new perspective on pedigree and the ambiguities of admissions, Rev. High. Educ. **41**, 497 (2018).

[22] Alexander R. Small, Range restriction, admissions criteria, and correlation studies of standardized tests, arXiv:1709.02895.

[23] Reuben M. Baron and David A. Kenny, The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations, J. Personality Soc. Psychol. **51**, 1173 (1986).

[24] Judith J. M. Rijnhart, Jos W. R. Twisk, Iris Eekhout, and Martijn W. Heymans, Comparison of logistic-regression based methods for simple mediation analysis with a dichotomous outcome variable, BMC Med. Res. Methodol. **19**, 19 (2019).

[25] Andrew F. Hayes and Michael Scharkow, The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: Does method really matter?, Psychol. Sci. **24**, 1918 (2013).

[26] Nathaniel Amos and Andrew F. Heckler, Mediating relationship of differential products in understanding integration in introductory physics, Phys. Rev. Phys. Educ. Res. **14**, 010105 (2018).

[27] Susanne Ditlevsen, Ulla Christensen, John Lynch, Mogens Trab Damsgaard, and Niels Keiding, The mediation proportion: A structural equation approach for estimating the proportion of exposure effect on outcome explained by an intermediate variable, Epidemiology **16**, 114 (2005).

[28] Laurence S. Freedman, Confidence intervals and statistical power of the ‘Validation’ ratio for surrogate or intermediate endpoints, J. Stat. Plan. Inference **96**, 143 (2001).

[29] Andrew F. Hayes, PROCESS: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling, Tech. Rep., 2012.

[30] Kristopher J. Preacher, Derek D. Rucker, and Andrew F. Hayes, Addressing moderated mediation hypotheses: Theory, methods, and prescriptions, Multivariate Behav. Res. **42**, 185 (2007).

[31] M. B. Weissman, Do GRE scores help predict getting a physics Ph.D.? A comment on a paper by Miller *et al.*, Sci. Adv. **6**, eaax3787 (2020).

[32] Jayson Nissen, Robin Donatello, and Ben Van Dusen, Missing data and bias in physics education research: A case

for using multiple imputation, Phys. Rev. Phys. Educ. Res. **15**, 020106 (2019).

[33] Starr Nicholson and Patrick J. Mulvey, Roster of physics departments with enrollment and degree data, 2017, American Institute of Physics, New York, Tech. Rep., 2018, https://www.aip.org/statistics/reports/roster-physics-2017.

[34] Starr Nicholson and Patrick J. Mulvey, Roster of physics departments with enrollment and degree data, 2018, American Institute of Physics, New York, Tech. Rep. 2019, https://www.aip.org/statistics/reports/roster-physics-2018.

[35] National Center for Education Statistics, *NCES-Barron's Admissions Competitiveness Index Data Files: 1972, 1982, 1992, 2004, 2008, 2014* (National Center for Education Statistics, Washington, DC, 2017).

[36] Lindsay Owens, Benjamin M. Zwickl, Scott V. Franklin, and Casey W. Miller, Identifying qualities of physics graduate students valued by faculty, in *Proceedings of the 2019 Physics Education Research Conference, virtual conference* (AIP, New York, 2020).

[37] Raj Chetty, John Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan, Mobility Report Cards: The Role of Colleges in Intergenerational Mobility, National Bureau of Economic Research, Cambridge, MA, Tech. Rep. No. w23618, 2017.

[38] Adrienne L. Traxler, Ximena C. Cid, Jennifer Blue, and Ramón Barthelemy, Enriching gender in physics education research: A binary past and a complex future, Phys. Rev. Phys. Educ. Res. **12**, 020114 (2016).

[39] Tiffani L. Williams, Underrepresented minority considered harmful, racist language (2020), https://cacm.acm.org/blogs/blog-cacm/245710-underrepresented-minority-considered-harmful-racist-language/fulltext.

[40] Robert T. Teranishi, Race, ethnicity, and higher education policy: The use of critical quantitative research, New Dir. Inst. Res. **2007**, 37 (2007).

[41] Julie Posselt, Theresa Hernandez, Geraldine Cochran, and Casey Miller, Metrics first, diversity later? Making the shortlist and getting admitted to physics Ph.D. Programs, J. Women Minorities Sci. Eng. **25**, 283 (2019).

[42] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevPhysEducRes.17.010144 for plots that show institutional effects by gender and race.

[43] Frank J. Massey, Jr., The Kolmogorov-Smirnov test for goodness of fit, J. Am. Stat. Assoc. **46**, 68 (1951).

[44] Casey Miller and Julie Posselt, Broadening participation in graduate education through holistic review, in *Proceedings of the Bridge Program and National Mentoring Community Conference* (2018), https://meetings.aps.org/Meeting/BPNMC18/Session/W3A.1.

[45] Julia D. Kent and Maureen Terese McCarthy, *Holistic review in graduate admissions: A report from the Council of Graduate Schools* (Council of Graduate Students, Washington DC, 2016).

[46] Alexander Rudolph, Gibor Basri, Marcel Agüeros, Ed Bertschinger, Kim Coble, Meghan Donahue, Jackie Monkiewicz, Angela Speck, Keivan Stassun, Rachel Ivie, Christine Pfund, and Julie Posselt, Final Report of the 2018 AAS Task Force on Diversity and Inclusion in Astronomy Graduate Education, Bull. AAS **51**, (2020).

[47] Kris Dunlap, Journey to a more holistic admissions review process by implementing an evaluation rubric (2018), https://cpb-us-w2.wpmucdn.com/u.osu.edu/dist/d/39848/files/2018/07/Kris-AGPA-preso-for-publication-abbrev-w45oyd.pdf.

[48] Marenda A. Wilson, Max A. Odem, Taylor Walters, Anthony L. DePass, and Andrew J. Bean, A model for holistic review in graduate admissions that decouples the GRE from race, ethnicity, and gender, CBE Life Sci. Educ. **18**, ar7 (2019).

[49] John R. Searle, How to derive "Ought" from "Is", Philos. Rev. **73**, 43 (1964).

[50] Micah Allen, Davide Poggiali, Kirstie Whitaker, Tom Rhys Marshall, and Rogier A. Kievit, Raincloud plots: A multi-platform tool for robust data visualization, Wellcome Open Res. **4**, 63 (2019).

[51] Kirstie Whitaker, Tom Rhys Marshall, Tim Van Mourik, Paula Andrea Martinez, Davide Poggiali, Hao Ye, and Marius Klug, Raincloudplots/raincloudplots: Wellcomeopenresearch, https://doi.org/10.5281/ZENODO.3368186 (2019)

[52] K. M. Whitcomb and C. Singh, Not all disadvantages are equal: Racial/ethnic minority students have largest disadvantage of all demographic groups in both STEM and non-STEM GPA, arXiv:2003.04376.