Investigating approaches to diversity in a national survey of physics doctoral degree programs: The graduate admissions landscape

Geoff Potvin,[1,*] Deepa Chari,[2] and Theodore Hodapp[3]

[1]*Department of Physics, and STEM Transformation Institute, Florida International University, Miami, Florida 33199, USA*

[2]*STEM Transformation Institute, Florida International University, Miami, Florida 33199, USA*

[3]*American Physical Society, College Park, Maryland 20740, USA*

(Received 20 July 2017; published 29 December 2017)

Graduate admissions play a critical gatekeeping role in the physics community not only because they select students who are allowed to begin their graduate studies, but also because they influence how students perceive graduate school, and in some cases whether or not they will even choose to apply. In conjunction with the APS Bridge Program, we conducted a national survey of graduate directors (and related faculty) of physics Ph.D. programs in the United States to explore graduate admissions practices. Our focus was on criteria used in determining admissions, mechanisms through which graduate applicants are handled, and how student representation considerations are incorporated into admissions (if at all). We report here on existing graduate admission practices in physics departments and highlight some critical issues for understanding barriers for diversifying graduate physics, including the use of GRE scores (and the relative importance placed on them). We find that the use of a minimum GRE score for admission, a practice in opposition to recommendations made by the tests designers, is reported to be used in many departments (more than one in three). We also find letters of recommendation to be highly valued in admissions decisions. Our data describe various initiatives at the institutional or individual level to increase gender diversity in admissions. A sizable number of departments also express a latent demand for greater numbers of students from traditionally marginalized racial or ethnic groups, but simultaneously report a lack of such applicants.

DOI: 10.1103/PhysRevPhysEducRes.13.020142

I. INTRODUCTION

Graduate admissions play a critical role in shaping the future of physics, both in terms of choosing who is allowed to participate in physics and in affecting (often implicitly) the research (and other) agendas of the discipline. Only those individuals who are successful in their application to graduate school will have the opportunity to conduct doctoral-level research and participate in, and have significant influence over, shaping future research activities. In addition, students who perceive they have little chance of being accepted to graduate programs may not make the attempt [1]—a process that involves significant effort, time, and money. Similarly, students who are less knowledgeable or poorly mentored about admission processes may not even consider applying for graduate studies in the first place, and are thus left out. In this study, we aim to bring transparency about the process of graduate admissions in doctoral degree-granting physics departments in the United States and to reveal some practices that work against establishing an inclusive pool of scientific talent. Through this work, we seek to identify normative admission practices and to inform institutions as well as prospective graduate students about present practices and future considerations for diversifying graduate education in physics. Doctoral education plays a critical role in the sustainability of the physics enterprise. Individuals pursuing doctoral education are prepared for “the problem-solving, -identifying, and strategic-brokering activities” [2] necessary to drive physics forward in its research, education, and other endeavors. A doctorate not only rewards individuals’ accomplishments in graduate school, but also recognizes their future potential to lead the development of physics. In this way, the awarding of a doctorate empowers individuals, in both real and symbolic ways, to set the vision and scope of the physics community’s agenda.

*gpotvin@fiu.edu

II. RESEARCH QUESTIONS

The goal of the current study was to establish a nationally representative understanding of the current admissions practices in place at doctoral-granting physics departments in the United States. As such, the overarching research goal is to answer the following question:

What are the normative practices used by physics departments to determine admissions to doctoral programs?

To address the specific concerns summarized in the introduction, in the rest of this paper we address the following specific research questions:

(1) How are student factors (prior performances, experiences, etc.) ranked in importance when evaluating student applications (as reported by graduate program directors and related faculty)?
(2) How are GRE scores (quantitative, verbal, written, and/or physics subject) being used by departments in the admissions process, and are they being used in a consistent manner, especially in light of ETS' guidelines on the use of cutoff scores?
(3) How are considerations of diversity (race or ethnicity, gender) being accounted for in current admissions decisions, if at all?

III. BACKGROUND AND LITERATURE REVIEW

Graduate school experiences, although involving unique elements for each graduate student, position individuals as "key performers" in driving physics forward in research and education. However, given the existing lack of diversity in the physics community, it is worth investigating whether graduate admissions processes fairly allow all individuals the opportunity to have these experiences. To bring clarity to this issue, we begin with a review of research on graduate education with an eye towards gender and racial or ethnic diversity, although many of these lessons might also be easily applied to first generation college students or those from lower social-economic status. A greater understanding of graduate admission practices from this perspective motivates the current study and highlights the need to understand institutional practices for attracting and retaining underrepresented populations in graduate education. Thus, it is relevant to examine if (and how) considerations are made in current admission processes to fairly evaluate and/or attract a diverse pool of applicants who can thrive in graduate school.

A. Trends of traditionally marginalized students in physics

Physics continues to have an appallingly low fraction of women and students from traditionally marginalized racial or ethnic groups (in this study we include students who identify as Hispanic Americans, African Americans, and Native Americans) in comparison to the general population as well as to several other STEM fields. According to recent census estimates, black or African-American-identified individuals represent 16%, Hispanic-identified individuals represent 21%, and American Indian or Alaskan Native-identified individuals represent 1.5% of 20–24 year olds, respectively [3].

In terms of STEM participation, of all science and engineering bachelor's degrees awarded to U.S. citizens and permanent residents, the proportion awarded to black-identified students was 9%, to Hispanic students was near 10%, and to Native American or Alaskan Native students was 0.6% [4]. By contrast, students from these groups earned only 3%, 6%, and 0.6% of physics bachelor's degrees, respectively [4,5]. These proportions are disquieting; however, even more alarming is the fact that the long-term trend for African American students is slightly negative, meaning a smaller percentage of such students are completing bachelor's degrees in recent years than in the past, going against a significant overall increase in the number of physics bachelor's degrees in recent years (having rebounded from historic lows in the mid-1990s).

In the same period, black-identified and Hispanic-identified students each earned about 6% of all science and engineering doctoral degrees awarded to U.S. citizens and permanent residents, while American Indian or Alaskan Native students earned 0.5% [4]. Simultaneously, black-identified students earned only 2% of domestic physics doctorates, with Hispanic-identified students earning 4% and American Indian or Alaskan Native students earning 0.3% (the latter representing, in fact, just three American Indian or Alaskan Native students in the entire country) in 2011 [4,6].

Similarly, while women earn about 50% of all the science and engineering bachelor's degrees awarded to U.S. citizens and permanent residents and 47% of all science and engineering doctoral degrees, they represent slightly less than 1 in 5 bachelor's degrees awarded in physics, and a similar percentage of physics doctorates, the latter essentially catching up to the stagnant (actually, slightly declining) proportion of female bachelor's degree recipients [4,5,7].

Only those individuals who successfully receive doctorates form the main talent pool for faculty recruiting, so the lack of diversity in graduate education directly limits diversity in the faculty ranks. At the tenure or tenure-track level, more than 85% of all physics departments have no African American faculty members [8]. Similarly, about 80% of physics degree-granting departments have no Hispanic-identified faculty [9]. Despite women still being a minority within physics, there are positive signs. In 2010, the National Research Council reported that 24.1% of assistant professors of physics are women—a percentage larger than that of recent doctoral classes and an indication that departments are making an effort overall to recruit women into the academy. Levels of representation at the associate and full professor ranks are also consistent with hiring and retention from cohorts associated with the years in which these groups originated [10]. While these numbers are encouraging, there is also a small difference in gender representation of graduate students based on the size and prestige of the university [11]. This lower rate of women with Ph.D.s at more elite schools when compared to the larger average has implications on faculty hiring at research universities, as hires are more often made from such institutions. Improving the representation of marginalized groups beyond graduate education will be difficult to achieve without considering the limits to access and choices imposed in undergraduate and graduate education in the first place.

B. Systematic effects in admissions that may affect graduate diversity

There are many possible reasons that students may or may not apply, and be accepted, to any particular graduate program. However, it has been widely discussed that there are systemic factors that may lead to traditionally marginalized students being systematically selected out of the admission processes [12–14]. Notably, the Graduate Record Examinations (GRE) are widely used as criteria in graduate admissions across a broad range of disciplines, and questions have been raised about the appropriate interpretation and use of GRE scores, with implications for diversity. There are several different components of the GRE with independent scores including quantitative, verbal, and analytical reasoning sections as well as several subject-specific tests including physics. GRE tests are written and administered by the Educational Testing Service (ETS). ETS itself recommends against the use of the GRE as an exclusive determinant in the admission process [15]. ETS comments in its official guidebook on the use of the GRE:

> *The GRE Board believes that GRE scores should never be the sole basis for an admissions decision and that it is inadvisable to reject an applicant solely on the basis of GRE scores. A cutoff score below which every applicant is categorically rejected without consideration of any other information should not be used [16] (p. 13).*

It is a longstanding, well-known fact that the GRE test results have gender- and race or ethnicity-specific testing effects [12]. Specifically, women score lower, on average, than men in both quantitative and verbal GRE tests [17]. In addition, there are systematic differences in the average scores of students of different racial or ethnic groups [12]. Despite this, the GRE remains a frequent gatekeeping criteria (including being used as a "first cut") due to broad beliefs about its ability to account for issues such as grade inflation and differential grading systems at varying undergraduate institutions, thus (apparently) easing the process of comparing graduate applications from a wide pool of applicants [18]. On this issue, Miller has noted how the use of a simple cutoff score may have a significant, negative impact on graduate diversity [12] by eliminating a large proportion of such applicants who fall below a fixed cutoff, despite the previously mentioned warnings from the ETS. In this paper, we present direct evidence on the prevalence of such practices in admissions decisions for physics doctoral programs.

C. Understanding the validity of admissions criteria

The success of graduate students is framed most commonly in terms of outcome indicators such as graduate GPA, research publications, citation counts, and times to degree [19]. The predictive validity of admission criteria is often understood by examining how admission criteria are correlated with such outcomes. The most commonly discussed admission criteria are GRE scores, which have been found to be positively correlated to first-year graduate GPA and success in qualifying exams in a range of fields [20–22] and are often presented as being a measure of undergraduate content knowledge. Nonetheless, GRE scores are much more weakly related to outcomes like citations counts and research products [20,23], and time-to-degree [19,24]. This is an important, though often overlooked reality: it may be assumed that high GRE scores are strongly predictive of all relevant graduate student outcomes, but this is not apparently the case. It is research-oriented outcomes that are highly valued in the long run, and for good reason: many would agree that activities such as citation counts and research products are the best indicator of novel science and of productive scientists. To this end, in prior work, Potvin and Tai reported no significant correlation between prior grades (both high school and undergraduate) and time to Ph.D. [19]. The latter was, in fact, predicted by several factors mainly outside the control of the student (e.g., number of required courses in program, etc.). This earlier study raised questions about graduate admissions decisions possibly over-relying on prior performance indicators (e.g., GREs, which are highly correlated to grades, see Kuncel and Hezlett [20]).

A few studies have indicated that other criteria such as students' writing samples or interviews may be potentially useful to graduate admissions [25]. Further, variables relating to students' perceptions and motivations, independently or in combination with quantitative GRE, are also reported to be better predictors of success of graduate students in physics and other disciplines [26–29]. It was reported that students' prior interests in science and their motivation to pursue graduate studies predict career productivity in science [27,28]. Specifically, career scientists who reported they pursued graduate studies because they "loved thinking about science" were found to be more productive (in terms of grant funding and primary or first-author publications) than average, while those who pursued graduate studies because of good grades or fellowships were no more productive than average. Furthermore, such correlations with students' success may be greater for students coming from traditionally marginalized groups [30], a point to note in light of the paucity of such students in graduate physics. Additionally, other studies have worked in the context of adult learners. Such students may gain rich work-related experiences and, when evaluated solely on admissions criteria like undergraduate GPA or GRE score, they may be afforded fewer opportunities for graduate studies as these factors may not reflect their skills [31].

IV. CONTEXT OF CURRENT STUDY: THE APS BRIDGE PROGRAM

In an effort to stop the marked leak of students from traditionally marginalized racial and ethnic groups from the academic pipeline, the American Physical Society (APS) began, in 2012, the APS Bridge Program—a project with national scope intended to increase the fraction of students from traditionally marginalized racial or ethnic groups who complete doctorates in physics [32]. Specifically, the Bridge Program is focused on increasing the number of students who identify as Hispanic American, African American, and Native American who complete physics Ph.D.s. This project has developed an alternate application process for prospective graduate students who have not otherwise been successful in gaining admissions to graduate school, has established programs at departments that have demonstrated their interest and capability in supporting such students ("Bridge Sites" and "Partnership Institutions"), and developed a national dialogue and support network of students, faculty, and administrators focused on these issues through annual conferences, publications, a presence at national meetings, and other activities.

In concert with the other Bridge Program activities, the survey analyzed in this manuscript was initiated in 2012 to systematically document, in some ways for the first time, admissions processes in physics departments that award doctorates throughout the United States. With a dearth of research about graduate education and admissions practices in particular, the research team was motivated to conduct an exploration of criteria in graduate admissions as a step to understand if (and how) these processes may influence the participation of underrepresented groups. A purpose of this research study is to disseminate these findings to relevant stakeholders (prospective students, current students, physics faculty, departmental and institutional administrators, etc.) in order to make the admissions process more transparent so that everyone can make better-informed decisions in physics graduate admissions.

V. DATA COLLECTION AND ANALYSIS

A. Survey design and validity

A survey consisting of multipart questions was developed in late 2012 to probe the admissions practices of active doctoral programs in the United States. An active Ph.D. program was defined as a program having graduated at least one Ph.D. in the previous five years. Initial survey development was based on prior instruments of the lead author (e.g., see Ref. [19]), existing literature in this area [20,33], focus group discussions with graduate students, and discussions between the authors and other APS Bridge Program management team members.

Following initial development, in January 2013, a draft instrument was circulated to participants of the 2nd Graduate Education Conference in College Park, MD. At the conference, focus groups were organized with graduate student attendees separately from faculty and other senior attendees. The purpose of these focus groups was to help establish some aspects of content and construct validity—to determine whether the instrument was comprehensible, coherent, and probed the "right" issues in graduate admissions. Based on the discussion and feedback of these focus groups, a second draft was developed which was then given to another graduate student focus group. The final instrument contains 30 multipart questions (see Supplemental Material [34] for the final version of the survey) including items probing

- The importance of a wide variety of student criteria in determining admission,
- How student representation considerations are incorporated into admissions decisions (if at all),
- Recruiting and processing mechanisms through which graduate applicants flow (e.g., the proportion of students who receive teaching assistantships, how admissions committees are constituted, etc.), and
- Respondents' academic status and demographic information.

B. Data collection

The survey responses were solicited via email beginning in August 2013. The APS maintains a contact list of program directors and/or department chairs of all active physics Ph.D. programs in the United States (approximately 185 departments have active Ph.D. programs in any given year). These individuals were contacted directly and solicited to complete the final, anonymous survey instrument online. They were also encouraged to distribute the survey link to faculty in their departments who were active in their graduate admissions processes. Nonrespondents were resolicited periodically during September and October of 2013.

C. Response rate

The APS list of Ph.D. program directors contained 200 different institutions at the time of survey administration. In total, 170 individuals identifying themselves from 149 different institutions responded to the survey. Thus, the response rate is estimated to be 75% of those departments that were solicited (e.g., 149 of 200 institutions), and is estimated at approximately 83% of the active Ph.D. programs in 2013 (there are approximately 180–185 active physics Ph.D. programs in the U.S. in any given year). As another estimate of this sample, the institutions appearing in the sample represent upwards of 80% of Ph.D. student output in the U.S.: this was estimated by comparing the four-year average of the number of Ph.D.s awarded in these institutions to the total number of Ph.D.s awarded in the U.S. at the same time [6]. Note that these response rates, at the institutional level at least, are quite high by typical survey standards. The high response rate can be attributed at least in part to the position that the APS holds in the physics community as well as the importance placed on these topics by graduate directors and other physics faculty. Some departments are represented by more than one of their faculty in the data (up to three, in a few cases). Since the focus of the paper is to estimate departmental-level practices, in the quantitative results reported below, responses are weighted so that no one department weighs more than any other (e.g., if there are two responses from one department, each of these respondents is weighted as 0.5).

D. Quantitative data analysis

All of the data cleaning and processing was conducted in R [35]. We used the specific packages car [36], ggplot2 [37], and plotrix [38] for computation and graphing. Throughout the quantitative analyses, the maximum allowed probability of Type I error, or "false positives," was set to be 5%. Again, as mentioned, the data set contains representation from a large majority of the entire population of Ph.D.-granting physics departments; however, for prudence, we continue to report inferential statistics when making claims about real differences in means, etc., throughout the quantitative analysis.

E. Thematic analysis of open-ended responses

Several survey questions, including three that will be discussed in this paper, involved open-ended responses. Using thematic analysis, we categorized the emergent responses into themes following their relations with the specific research questions [39]. This approach allows us to build an emergent, but robust, understanding of this non-quantitative data, which was necessary to understand several issues related to graduate admissions; in particular, for research questions 2 and 3 in this paper, where there was potentially significant inter-institutional variation in practices and beliefs.

VI. RESULTS

In this section, we present the findings for each research question in order, with a detailed discussion afterward.

Research question 1: How are student factors (prior performances, experiences, etc.) ranked in importance when evaluating student applications (as reported by graduate program directors and related faculty)?

To address this question, we analyzed the responses of a survey question that probed the respondents on the importance of 21 possible student criteria to their admissions decisions. Each of these items was scored on independent, seven-point anchored scales running from "Least important" to "Most important;" separately, respondents could indicate that they "Don't Use" each of the criteria. Specifically, the 21 criteria probed are listed below:

(a) GPA or grades–general
(b) GPA or grades–physics or math
(c) Undergraduate courses taken
(d) Undergraduate institution type or reputation
(e) GRE quantitative scores
(f) GRE verbal scores
(g) GRE written scores
(h) GRE physics subject scores
(i) TOFEL scores
(j) Quality of letters of recommendation
(k) Reputation of recommenders
(l) Recommenders' ranking of students
(m) Quality of interviews (conducted by your department)
(n) Proximity or familiarity to department
(o) Personal statement
(p) Prior research experiences
(q) Prior publications
(r) Prior conference presentations
(s) Student research interest and/or stated faculty advisor
(t) Departmental research opportunities (Specific availability of research groups)
(u) Other

The mean responses along with associated standard error for each factor is summarized in Fig. 1, with the overall mean indicated with a solid line. The category "u. Other" had a low response rate (with different fill-in responses, of course) so it is not included in the analysis below. Again, responses were weighted so that each institution had equal weight.

The solid black line indicates the midpoint of the response scale; the dotted line represents the mean response. For nearly every factor (with the exceptions of "g. GRE written scores" with mean 3.75 ± 0.11, "n. Proximity or familiarity to department" with mean 3.17 ± 0.17, and "r. Prior conference presentations" with mean 3.92 ± 0.12), the mean response is above 4, which is the midpoint of the scale (1 to 7). This clearly indicates the tendency of respondents to simultaneously consider the factors they use in admissions as "relatively important." However, within this context, several differences between factors are statistically significant. We performed non-parametric, pairwise Wilcoxon rank-sum tests between each pair of factors to look for differences in the relative importance placed on each criterion. Those factors which appear vertically separated (e.g., whose mean estimates do

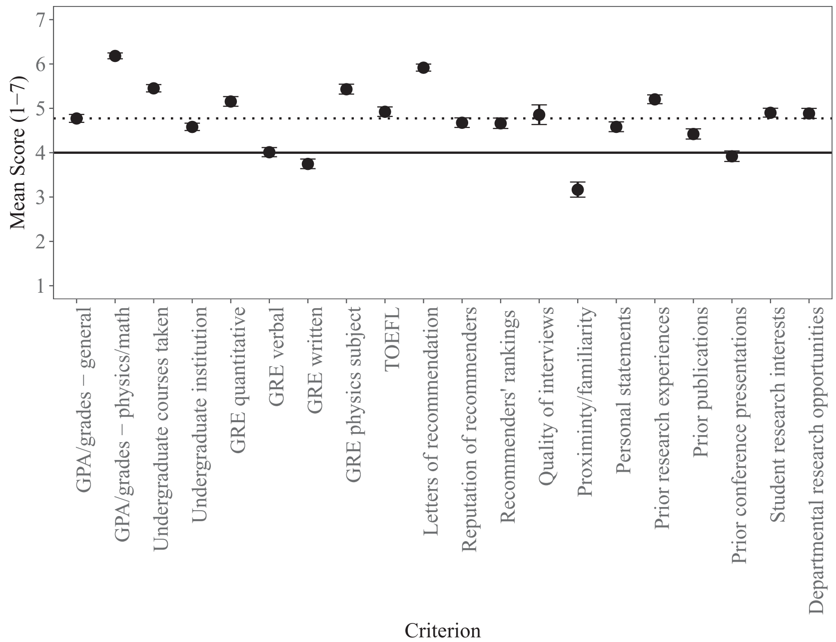

FIG. 1. Mean importance assigned to each admission criteria.

not have overlapping ranges, according to their standard errors) in Fig. 1 are generally statistically significantly different (at the $p < 0.01$ level), as one would expect. The complete table of pairwise tests (p values) is reported in the Appendix. Interestingly, the factors rated as most important are (in decreasing order):

(b) GPA or grades—physics or math (6.18 ± 0.07),
(j) Quality of letters of recommendation (5.92 ± 0.08), and
(c) Undergraduate courses taken (5.45 ± 0.08), statistically tied with
(h) GRE physics subject scores (5.43 ± 0.11).

At the other end of the scale, the factors rated least important are (in increasing order):

(n) Proximity or familiarity to department (3.17 ± 0.17),
(g) GRE written scores (3.75 ± 0.11),
(r) Prior conference presentations (3.92 ± 0.12), and
(f) GRE verbal scores (4.01 ± 0.1).

One might also wonder if these responses are systematically different for larger programs compared to smaller programs (which might not receive as many applications every year). Thus, we considered whether the size of each program (as measured by the average number of Ph.D.s awarded annually and reported by the NRC [40]) is predictive of these responses. We broke institutions into quartiles according to their reported program size, and conducted regression analyses to compare responses to the factors appearing in Fig. 1. In most cases, there is no association between program size and response; however, some significant differences are seen for interesting factors: at the $p < 0.01$ level, compared to the smallest programs, the largest give significantly lower importance to GRE quantitative scores (by 1.1 points, on average) and significantly higher importance to letters of recommendation (0.73 points), personal statements (1.0 points), prior research experiences (1.4 points), and prior publications (1.1 points).

Separate from the ratings of importance of each factor, another way to consider these factors is the rate at which each criterion was indicated as not being used at all (indicated as *Don't Use* by respondents, which supplanted responses on the anchored scales above). The fraction of institutions reportedly not using each of these factors is summarized in Fig. 2. Three somewhat unsurprising factors are most often reported as unused; namely, "m. Quality of interviews (conducted by your department)" (71% not using), "n. Proximity or familiarity to department" (46% not using), and "l. Recommenders' rankings of students" (24% not using). Interestingly, the fourth most-frequently unused criterion is "h. GRE physics subject scores" (14% not using). Similar to the analysis of Fig. 1, we also considered whether program size is predictive of responses appearing in Fig. 2. In this case, we see no significant differences in whether a program reports not using any criteria at the $p < 0.01$ level in Kruskal-Wallis tests.

Further, a detailed investigation indicates that those institutions who report not using the GRE physics subject test do not, on the surface, appear to fit a simple profile—e.g., only public state schools respond thusly, or only top ranked schools, etc.

Research question 2: How are GRE scores (quantitative, verbal, written, and/or physics subject) being used by departments in the admissions process, and are they being used in a consistent manner, especially in light of ETS' guidelines on the use of cutoff scores?

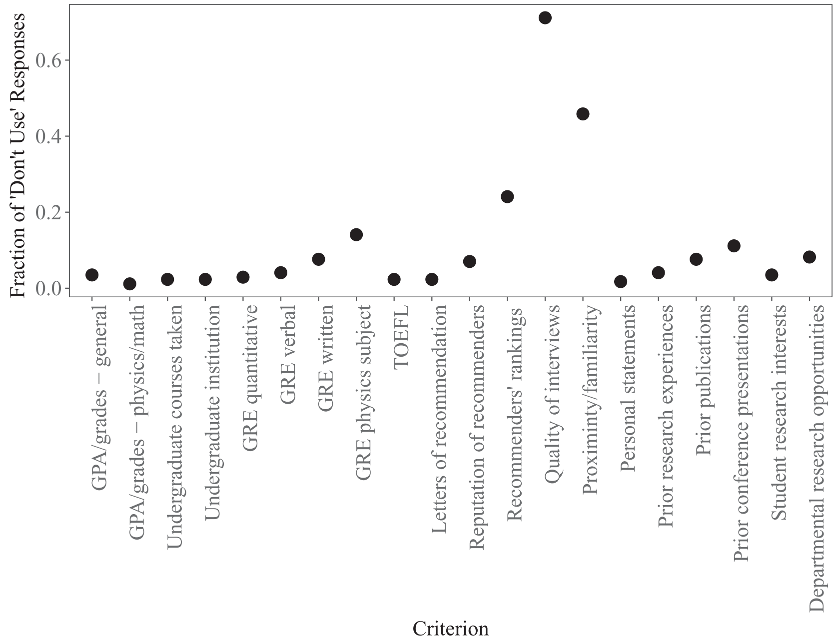

FIG. 2. Proportion of respondents indicating they "Don't Use" specific admissions criteria.

The previous results that highlighted the use of GRE scores further motivated us to examine their importance and use in greater detail, so we now turn to our second research question. Specifically, we analyzed responses to two questions; first, to the closed-ended yes or no question:

> *Q13. Are GRE scores (quantitative, verbal, written, or physics subject) used as a minimum cutoff in admissions decisions?*

and the immediate open-ended follow up of

> *Q13a. Please briefly describe how GRE scores are used.*

To analyze these data, we first provide a summary of the counts of responses to these questions, and then present the results of the thematic analysis of the open-ended Q13a. To the first question (170 responses), 32% indicated "Yes," meaning their department uses a GRE cutoff, and 68% indicated "No." For the follow-up question, 48 respondents in the yes category and 29 in the no category provided further details, as summarized in Table I.

Amongst the no responses to Q13, a thematic analysis of the responses to Q13a underlines the long shadow cast by the Graduate Record Examination in several cases. Again, note that these respondents initially indicated that their department did not use a GRE cutoff. Nonetheless, we noted several phrases such as "a rough cutoff" "preferable score" "as a first cutoff" "strongly recommended but no [sic] required" all indicating the influence of the GRE. We present a few responses as exemplars:

> *No fixed cutoff, but GRE quantitative should be about 90 [sic] percentile or higher. (Respondent 150)*

> *No hard cutoff, but used as a first cut in going through applications and GRE scores trump GPA scores in assessing students. (Respondent 49)*

Among other responses, a common sentiment is that low GRE scores trigger a more comprehensive look at an application—a use of the GRE that is more in line with ETS guidelines. The following excerpt from a respondent typifies well this sentiment:

> *There is no strict cutoff, but a very low score means other parts of the application must compensate and explain why the student may nonetheless be successful in our program. (Respondent 73)*

We also extended this analysis by breaking out the responses to Q13 according to the research ranking of the departments (which is not synonymous with program size, discussed previously). Using the average NRC research ranking [40] of physics departments (averaging

TABLE I. Response counts for questions 13 and 13a.

	Response given for Q13a?	
Response to Q13	No further details given	Further details given
No (115 total)	86	29
Yes (55 total)	7	48

the 5th and 95th percentile estimates), we sorted departments into quartiles and performed a Kruskal-Wallis rank sum test to test whether or not this research categorization is associated to the prevalence of using a GRE cutoff. Indeed, we found this to be the case (at the $p < 0.01$ level): the highest ranked quartile reported yes to Q13 in only 4 of 35 (12%) of cases, while the lowest ranked quartile responded yes in 18 of 35 (51%) cases. The small number of responses to Q13a in each subcategory of research ranking that also indicated no to Q13 prevents us from drawing further conclusions about which of these ranking groups (if any) are more likely to be using GRE cutoffs in the contradictory manner discussed above, but there appears to be an association between being ranked lower in the NRC listings and being more likely to use GRE cutoffs. We also compared the prevalence of yes responses to Q13 according to the program size ranking reported earlier. Again, performing a Kruskal-Wallis rank sum test between program size quartile and Q13 responses, we see no significant differences. Thus programs which are larger (but not ranked higher in research) do not appear to be more likely to report using a GRE cutoff.

Based on the data available, we cannot make an exact estimate for the prevalence of a GRE cutoff (there is sufficient ambiguity in some responses as well as several individuals who did not provide sufficient detail to make such an assessment), however, it is clear that the fraction of departments using cutoffs is well above 32%, as the responses to Q13 provide a lower bound. Despite the ETS direct caution against using GRE scores as an exclusive determinant in admissions, they are clearly being used in this way in a significant number of departments.

Research question 3: How are considerations of diversity (race or ethnicity, gender) being accounted for in current admissions decisions, if at all?

In the survey, we asked two further open-ended questions probing the ways in which gender and race or ethnicity are factored into admissions decisions (if at all). We address these responses by discussing considerations for women separately and somewhat independently of race or ethnicity.

A. Gender

We asked participants how gender-specific considerations are accounted for in admissions decisions through the question,

Please explain how applicants'gender is factored into application reviews and decisions, if at all:

Of the 170 respondents, 51 indicated gender is not factored at all (plus 17 nonresponses), while 102 responses indicated some consideration. A thematic analysis identified two themes from these 102 responses:

1. Recognizing the availability of awards or fellowships

Several respondents recognized and articulated that their institutions provide diversity fellowships for women. Such fellowships may be awarded before or after admissions decisions. An exemplar response in this theme was

Some fellowships are available only for female students. (Respondent 118)

Being aware of such institutional resources to promote women's representation at the graduate level itself may be catalytic in advancing gender equity, although relying on these resources may not be sufficient to support gender diversity in physics in particular. That is, diversity fellowships may support individuals to gain access to graduate school but may not drive the improvement of general practices in admissions (or postadmissions). It is noteworthy that the availability of these types of fellowships may or may not have any association with admissions decisions, but the fact that several respondents talked about such fellowships in response to this question indicates there is some level of awareness that gender diversity is something to be considered.

2. Efforts to tackle the challenge of gender diversity

Some participants discussed their efforts (including resulting successes or failures) to address gender diversity in their graduate programs. These efforts included devoting extra time to reading women's applications, or helping female graduates in searching and applying for scholarships to pursue their graduate education. The following two excerpts provide a sense of the responses in this category:

We spend more time per student reviewing applications from women in order to make offers to a high proportion of female applicants than male. (Respondent 3)

Approximately 15%–20% of our applicants are female. We make an effort to increase the representation of women in our program and, other things being equal, we always give priority to female applicants in our admissions. Even then, we end up with only 10%–15% female students in our incoming classes. (Respondent 134)

This latter response is instructive: this respondent appears to indicate a latent demand for improving the representation of women in their program, however, they nonetheless appear to be unable to recruit improved numbers of women from their applicant pool. Recall that these sorts of efforts, while not necessarily effective even according to our respondents, were reported relatively rarely in our data.

To summarize, several physics departments indicate some initiative at the institutional and personal level to increase gender diversity, usually in the form of identifying the availability of fellowships available to women (whether or not the female applicants are made aware of these or if this availability has any impact on admissions decisions is unclear from many responses), while a sizable fraction (about one-third) report not taking gender into account in admissions decisions at all.

B. Race or ethnicity

Similar to the previous survey item, we then asked respondents to explain how they took the race or ethnicity of applicants into consideration in their admissions decisions:

> *Please explain how applicants' race or ethnicity (e.g., students who identify as underrepresented minorities) is factored into application reviews and decisions, if at all:*

Of the total 170 responses, 48 indicated race or ethnicity is not taken into account at all (plus 19 further nonresponses) while 103 responses indicated some consideration. (Note that the 48 respondents who indicated they did not account for race or ethnicity were not completely overlapping with the 51 respondents from the previous question who indicated gender was not taken into account.) From the thematic analysis of these responses, two themes again emerged:

1. Reliance on the existence of diversity fellowships

Many respondents to this question focused exclusively on the availability of diversity fellowships at their institutions. One respondent also indicated efforts towards nominating graduate students for such fellowships. By comparison to the previous question about gender considerations, there was a relatively less involved treatment given to these considerations. A typical response in this theme is

> *We have fellowships available for students that are eligible for [specific institutional diversity fellowship]. We do check if the applicants are eligible for this fellowship. (Respondent 80)*

2. Recognizing a dearth of underrepresented applicants

Some responses outlined an approach of giving priority to domestic and/or underrepresented students, "other factors being equal." It is worth reflecting on what "other factors being equal" really means, given the prior work highlighting systematic disparities in other measures of merit [12,14]. Some respondents emphasized that they would readily accept underrepresented students but lamented the lack of such students (often zero) who appear in their application pool.

> *Unlike the male or female situation, we are not very successful in recruiting underrepresented minorities. If we find a candidate, we find a fellowship. The numbers are just not there in our pool. (Respondent 16)*

> *We get very few (to none) applicants that identify themselves as underrepresented minorities, the ones we get we look at carefully to see if we can accept them. (Respondent 132)*

Overall, our analysis related to research question 3 identified a latent demand at some institutions for enrolling greater numbers of students from traditionally marginalized racial or ethnic groups, as indicated by the above and similar other responses. However, these responses may be worrisome for those interested in improving the diversity of graduate physics students because they indicate a lack of effective departmental recruitment efforts. After all, the fraction of underrepresented students who complete bachelor's degrees in physics is higher than the fraction of such students who complete graduate school. So there is at least a *prima facia* opportunity to recruit students from undergraduate programs to consider applying to graduate school.

VII. DISCUSSION AND CONCLUSIONS

A. Research question 1

We found the factors rated as most important in admissions decisions to be physics and math GPA or grades, letters of recommendation, undergraduate course taking, and GRE physics subject scores (though this latter factor was also somewhat commonly reported as not being used at all). The factors rated least important were students' proximity or familiarity to department, GRE written and verbal scores, and prior conference presentations. One might expect students' prior course experiences (e.g., having taken certain critical subjects in physics or math) and their performance in physics or math to be central to their consideration for graduate school. This does, however, raise a concern: many students who attend undergraduate institutions with small physics programs may not have the opportunity to take higher-level "canonical" courses. This suggests that these students may have limited opportunities to pursue a graduate research career, unless there are other ways to show evidence of promise, at least as judged by admissions committees. Related to this, we found that the largest programs rate GRE quantitative scores as a bit less important than do the smallest programs, but they simultaneously rate several other application factors as more important, including personal statements, letters, prior research experiences, and prior publications, which emphasizes the importance of students showing their potential in one of these ways.

The appearance of letters of recommendation as a highly valued criterion (with a mean importance of 5.92 out of 7) indicates an avenue through which students can succeed in transitioning to a Ph.D., possibly even if they have not taken all the canonical undergraduate courses or received high GRE scores. Given that letters of recommendation are a typical part of a graduate application in which students' "other" qualities may be highlighted including research experiences, motivations towards school and/or research, their interest, creativity, grit, etc., as well as opportunity to explain the details of an academic record, it is likely that this is why letters are rated as so important [41]. The importance placed on letters of recommendation may also reflect an understanding on the part of faculty that the quantitative indices of grades, GPA, and GRE scores may not tell the "whole" story of a students' potential, and that letters are a place to gain this knowledge. For students who lack critically required coursework, a strong recommendation letter is likely to improve the chances of getting into graduate school. Applicants should understand the critical importance given to recommendation letters, and consider ways to ensure letter writers highlight their strengths and potential for research, and address weaknesses that may appear in other parts of an application. This also raises the issue of implicit bias, which can arise in these and other contexts and may have significant diversity implications [42]. Writers and readers of letters should work to be aware of potential implicit bias and interpret letters accordingly.

By contrast, it is noteworthy that GRE physics subject scores were simultaneously rated as one of the most important factors amongst departments considering them but also one of the factors most commonly identified as not being used in admissions decisions at all. Students in the APS Bridge Program have often reported to the project team of not applying to schools because of their GRE requirements. As a result, departments preselect their applicant pool (perhaps unconsciously) by choosing to require the physics subject test. We are pursuing a deeper investigation into students' views of admissions policies such as this in a separate study [43].

It is notable that two so-called "nonquantitative" components of the GRE appear to be considered amongst the least important factors, both of which may be relevant to producing successful future researchers. While our data do not speak to the importance of these factors to the success of individuals, we note that as science becomes increasingly collaborative, entrepreneurial, and team-based, effective interpersonal and communication skills may become ever more critical to the professional preparation of physicists. It is ironic, then, that both written and verbal GRE scores were indicated as amongst the least important criteria in the admissions process.

B. Research question 2

We identified many institutions to be using GRE cutoffs (at least a third of responding departments), and a significant number of other respondents who did not believe themselves to be using a GRE cutoff, but nonetheless articulate the *de facto* use of a cutoff through stated practices. Further, lower NRC-ranked departments were more likely to report using cutoff scores. This latter fact may be a reflection of the lack of variance in GRE scores seen by the highest-ranked institutions—applicants to the top programs self-select so that they reflect a relatively narrow band of possible GRE scores. Hence, this score may not be perceived by these programs as containing much distinguishing information amongst this relatively homogeneous pool of applicants.

Unfortunately, using an initial cutoff is a simple way of filtering applications for faculty who have little time to carefully scrutinize a large number of applications. However, it appears that the prevalence of such practices are reducing the chances of underrepresented students from gaining admission. There are two mechanisms through which this can happen: (a) student may choose not to apply for graduate school due to low GRE scores and their perception that this will eliminate their chances of success, or, (b) student may apply but still get eliminated solely due to not reaching a minimum GRE threshold. The first problem can be addressed to some extent by increasing undergraduate students' awareness about different institutions' varying use of graduate admission criteria. Students should at least be made aware of departments that do not consider GRE scores as the most important (or sole) criteria and/or do not use it. The second mechanism is more challenging to address in light of the apparent prevalence of GRE cutoffs reported here. Based on these numbers, the risks of already underrepresented students getting further, disproportionately eliminated in the application process after the initial consideration are sizable. As Miller has emphasized, the heavy reliance on the GRE score as a cutoff should raise concerns about the number of otherwise-qualified underrepresented students [12] who are losing the opportunity to get Ph.D.s.

C. Research question 3

In addressing the third question, we identified many institutions (again, about one-third) that are not taking into consideration gender or race or ethnicity (or both) in their admissions decisions. Of those that do, the most common theme was to identify the availability of diversity-related institutional or departmental fellowships, and to discuss departmental efforts to help students acquire such resources. While this approach counts for something, and at least tangentially recognizes the need for diversification in physics, we further note that some respondents recognize that their efforts are largely unsuccessful. On the other hand, it is notable, and cause for some hope, that several respondents recognized (and lamented) the scarcity of applications from underrepresented students.

VIII. CONCLUSIONS AND FUTURE WORK

The underrepresentation of women and traditionally marginalized racial or ethnic groups in graduate education has been consistently reported for years [6,9,44]. Not coincidentally, the risk of overreliance on a narrow band of admission criteria (as discussed here, GRE scores in particular) may substantially reduce students' opportunities for graduate school [12]. The current work does confirm that such practices are widespread, though not universal, in graduate physics departments. Given the increasing amount of evidence that traditional admissions measures fail to recognize and recruit capable individuals [32], especially from underrepresented groups, departments, and institutions broadly, should develop coherent strategies to more successfully evaluate applications of traditionally marginalized students in order to remove existing biases that work against creating an inclusive graduate student cohort.

Future work should seek to understand in more detail the student-level factors—both experiential and institutional—that can encourage undergraduate students, particularly women and students from traditionally marginalized racial or ethnic groups, to apply for graduate school. Relatedly, it would be of value to better understand which recruitment practices are most effective. The current findings should motivate a reconsideration of the various approaches to admissions and their implications, and to implement more inclusive admissions practices [45] that hold the possibility to admit students of great potential rather than just those who have experienced great advantage. It should be noted that all of these efforts, including the current paper, only support the first step in improving diversity in graduate physics, i.e., they may help to get students "in the door." At that time, we must also begin the process of fully supporting students through to the completion of their Ph.D., a process that can have challenges and potential threats for marginalized students. Much remains to be done.

A. Limitations of this study

In addressing the third research question, we discussed gender separately, and somewhat independently, of race or ethnicity. This is a limitation of the current work: after all, it is clear that the experiences of individuals do not neatly disaggregate according to one's race, ethnicity, gender, ability status, or any other personal characteristic. That is to say, an intersectional approach to understanding student experiences and marginalization that takes into consideration many aspects of the students' experiences would deepen our understanding of diversity in physics. Indeed, we advocate for such a holistic approach in the graduate admission process itself: understanding applicants as whole individuals, rather than as collections of indicators. This is an approach that is likely to be more successful in recruiting and retaining highly motivated and creative students from all backgrounds and origins.

Another limitation of this work is that, in most cases in this data set, a single individual responded from each institution. There is undoubtedly variance within departments about admissions practices. Hence, our findings should be interpreted cautiously and this is why we treated our data as a sample, not necessarily as a whole population, and provided statistical inference where possible, despite having data from well over three-fourths of Ph.D.-granting departments in the U.S.

Finally, this work uses a retrospective cross-sectional design, and, consequently, we cannot establish causality in relationships we have identified. Nor can we state definitively, for example, exactly how prevalent the use of a GRE cutoff is, given that some individuals provided no or ambiguous responses in this regard. It is clear that it is a widespread practice but we cannot state for certain exactly how prevalent it is.

ACKNOWLEDGMENTS

The authors would like to acknowledge the current and former APS Bridge Program staff and research team including Asmaa Khatib, Bushraa Khatib, Arlene Modeste Knowles, Erika Alexander Brown, Brian Beckford, Brián Clash, Monica Plisch, Geraldine Cochran, Kathryne Woodle, and Rachel Scherr. Further, the authors would like to thank Idaykis Rodriguez, Zahra Hazari, Casey Miller, and Gautam Bhattacharyya for discussions at various stages of this work. This material is based upon work supported by the National Science Foundation under Grant No. 1143070.

APPENDIX: RESULTS OF PAIRWISE WILCOXON TESTS ON ADMISSIONS CRITERIA

Results of pairwise, Wilcoxon rank-sum tests between each factor appearing in the analysis of research question 1. Here, n/s indicates a nonsignificant difference, * indicates a difference at the level of $p < 0.05$, ** indicates a difference at the level of $p < 0.01$, and *** indicates a difference at the level of $p < 0.001$.

	Q12a	Q12b	Q12c	Q12d	Q12e	Q12f	Q12g	Q12h	Q12i	Q12j	Q12k	Q12l	Q12m	Q12n	Q12o	Q12p	Q12q	Q12r	Q12s
Q12b	***	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
Q12c	***	***	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
Q12d	n/s	***	***	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
Q12e	**	***	*	***	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
Q12f	***	***	***	***	***	…	…	…	…	…	…	…	…	…	…	…	…	…	…
Q12g	***	***	***	***	***	***	…	…	…	…	…	…	…	…	…	…	…	…	…
Q12h	***	***	n/s	***	**	***	***	…	…	…	…	…	…	…	…	…	…	…	…
Q12i	n/s	***	***	*	*	***	***	***	…	…	…	…	…	…	…	…	…	…	…
Q12j	***	*	***	***	***	***	***	***	***	…	…	…	…	…	…	…	…	…	…
Q12k	n/s	***	***	n/s	**	***	***	***	n/s	***	…	…	…	…	…	…	…	…	…
Q12l	n/s	***	***	n/s	***	**	***	***	n/s	***	n/s	…	…	…	…	…	…	…	…
Q12m	n/s	***	n/s	n/s	n/s	*	**	n/s	n/s	***	n/s	n/s	…	…	…	…	…	…	…
Q12n	***	***	***	***	***	**	*	***	***	***	***	***	***	…	…	…	…	…	…
Q12o	n/s	***	***	n/s	***	***	***	***	n/s	***	n/s	n/s	n/s	***	…	…	…	…	…
Q12p	***	***	n/s	***	n/s	***	***	n/s	*	***	***	***	n/s	***	***	…	…	…	…
Q12q	n/s	***	***	n/s	***	**	***	***	*	***	n/s	n/s	*	***	n/s	***	…	…	…
Q12r	***	***	***	***	***	n/s	n/s	***	***	***	***	***	**	***	***	***	***	…	…
Q12s	n/s	***	***	**	*	***	***	***	n/s	***	n/s	n/s	n/s	***	*	**	**	***	…
Q12t	n/s	***	***	*	*	***	***	***	n/s	***	n/s	*	n/s	***	*	*	**	***	n/s

[1] A. Bandura, The explanatory and predictive scope of self-efficacy theory, J. Soc. Clinical Psych. **4**, 359 (1986).

[2] R. B. Reich, *The Work of Nations: Preparing Ourselves for 21st-Century Capitalism* (Random House, New York, 1991).

[3] U.S. Census Bureau, Current population survey: Annual social and economic supplement 2013, (Released March 2016).

[4] National Science Board, Science and engineering indicators 2014 (National Science Board: Arlington, VA, 2014).

[5] P. J. Mulvey and S. Nicholson, Physics bachelor's degrees: Results from the 2010 survey of enrollments and degrees. Focus on (Statistical Research Center of the American Institute of Physics, College Park, MD, 2012).

[6] P Mulvey and S. Nicholson, *Trends in Physics Ph.D.s* (Statistical Research Center of the American Institute of Physics, College Park, MD, 2014).

[7] National Center for Science National Science Foundation and Engineering Statistics, *Women, minorities, and persons with disabilities in science and engineering: 2017. special report nsf*, (National Science Foundation, Washington, DC, 2017), pp. 17–310.

[8] R. Czujko and S. Nicholson, African Americans among degree recipients in physics and geoscience. Focus on (Statistical Research Center of the American Institute of Physics, College Park, MD, 2010).

[9] R. Czujko and S. Nicholson, Hispanic americans among degree recipients in physics and geoscience, Focus on (Statistical Research Center of the American Institute of Physics, College Park, MD, 2010).

[10] T. Hodapp and Z. Hazari, Women in physics: Why so few?, APS News **24**, 8 (2015).

[11] T. D. Snyder, C. de Brey, and S. A. Dillow, *Digest of Education Statistics 2015* (National Center for Educational Statistics, Jessup, MD, 2016).

[12] C. W. Miller, Admissions criteria and diversity in graduate school, APS News **22**, 8 (2013).

[13] C. Miller and K. Stassun, A test that fails, Nature (London) **510**, 303 (2014).

[14] J. R. Posselt, Toward inclusive excellence in graduate education: Constructing merit and diversity in Ph.D. admissions, Am. J. Educ. **120**, 481 (2014).

[15] M. Walpole, N. W. Burton, K. Kanyi, and A. Jackenthal, Selecting Successful Graduate Students: In-Depth Interviews with GRE® Users, ETS Research Report Series **2002**, i (2002).

[16] *Guide to the Use of Scores* (Educational Testing Service, Princeton, NJ, 2017).

[17] A. Bleske-Rechek and K. Browne, Trends in GRE scores and graduate enrollments by gender and ethnicity, Intelligence **46**, 25 (2014).

[18] K. Perez, GRE as a predictor of graduate student success at a Hispanic serving institution of higher education, 361 (2011).

[19] G. Potvin and R. H. Tai, Examining the relationships among doctoral completion time, gender, and future salary prospects for physical scientists, J. Chem. Educ. **89**, 21 (2012).

[20] N. R. Kuncel and S. A. Hezlett, Standardized tests predict graduate students' success, Science **315**, 1080 (2007).

[21] N. R. Kuncel, S. Wee, L. Serafin, and S. A. Hezlett, The validity of the Graduate Record Examination for master's and doctoral programs: A meta-analytic investigation, Educ. Psychol. Meas. **70**, 340 (2010).

[22] N. R. Kuncel, S. A. Hezlett, and D. S. Ones, A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: implications for graduate student selection and performance., Psychol. Bull. **127**, 162 (2001).

[23] A. R. Marston, It is time to reconsider the Graduate Record Examination, Am. Psychol. **26**, 653 (1971).

[24] R. J. Sternberg and W. M. Williams, Does the Graduate Record Examination predict meaningful success in the graduate training of psychology? A case study, Am. Psychol. **52**, 630 (1997).

[25] J. D. Kent and T. M. McCarthy, *Holistic review in graduate admissions: A report from the council of graduate schools* (Council of Graduate Schools, Washington, DC, 2016).

[26] W. I. Pacheco, R. J. Noel, J. T. Porter, and C. B. Appleyard, Beyond the GRE: Using a composite score to predict the success of Puerto Rican students in a biomedical Ph.D. program, CBE-Life Sci. Educ. **14**, 1 (2015).

[27] Z. Hazari, G. Potvin, R. H. Tai, and J. Almarode, For the love of learning science: Connecting learning orientation and career productivity in physics and chemistry, Phys. Rev. ST Phys. Educ. Res. **6**, 010107 (2010).

[28] Z. Hazari, G. Potvin, R. H. Tai, and J. T. Almarode, Motivation toward a graduate career in the physical sciences: Gender differences and the impact on science career productivity, J. Coll. Sci. Teach. **41**, 90 (2012).

[29] W. E. Sedlacek, *Beyond the Big Test: Noncognitive Assessment in Higher Education* (Jossey-Bass, Indianapolis, IN, 2004).

[30] M. Oliveri and C. Ezzo, The role of noncognitive measures in higher education admissions, J. World University Forum **6**, 55 (2014).

[31] Laurie Dodge and Ellen Baker Derwin, Overcoming barriers of tradition through an effective new graduate admission policy, J. Continuing Higher Educ. **56**, 2 (2008).

[32] T. Hodapp and K. S. Woodle, A bridge between undergraduate and doctoral degrees, Phys. Today **70**, No. 2, 50 (2017).

[33] B. E. Lovitts, *Leaving the Ivory Tower: The Causes and Consequences of departure from Doctoral Study* (Rowman & Littlefield Pub. Inc., Lanham, MD, 2001).

[34] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevPhysEducRes.13.020142 for the doctoral admissions survey.

[35] R Development Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2017).

[36] J. Fox and S. Weisberg, *An R Companion to Applied Regression*, 2nd ed. (Sage, Thousand Oaks, CA, 2011).

[37] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, New York, 2009).

[38] J. Lemon, Plotrix: A package in the red light district of R, R-News **6**, 8 (2006).

[39] V. Braun and V. Clarke, Using thematic analysis in psychology, Qual. Res. Psych. **3**, 77 (2006).

[40] Committee to Assess Research-Doctorate Programs; National Research Council, *A Data-Based Assessment of Research-Doctorate Programs in the United States*, edited by J. P. Ostriker, C. V. Kuh, and J. A. Voytuk (National Academies Press, Washington, DC, 2011).

[41] As an aside, from another question in the survey, we identified only 29% of departments as taking the opportunity to interact with prospective graduate students before decisions are made. Thus, this adds to the concern that the totality of students' experiences and potential may not get considered by admission committees. This practice (of interviewing) is a key component used by several bridge sites in making final decisions on the top candidates.

[42] K. Dutt, D. L. Pfaff, A. F. Bernstein, J. S. Dillard, and C. J. Block, Gender differences in recommendation letters for postdoctoral fellowships in geoscience, Nat. Geosci. **9**, 805 (2016).

[43] D. Chari and G. Potvin (to be published).

[44] National Research Council *et al.*, *From Scarcity to Visibility: Gender Differences in the Careers of Doctoral Scientists and Engineers* (National Academies Press, Washington, DC, 2001).

[45] R. E. Scherr, M. Plisch, K. E. Gray, G. Potvin, and T. Hodapp, Fixed and growth mindsets in physics graduate admissions, Phys. Rev. Phys. Educ. Res. **13**, 020133 (2017).