Time to PhD completion is no different between men and women despite score gap on physics GRE

Mike Verostek
Department of Physics and Astronomy, University of Rochester and
School of Physics and Astronomy, Rochester Institute of Technology[*]

Casey W. Miller and Benjamin Zwickl
School of Physics and Astronomy, Rochester Institute of Technology

Using analysis of variance on a sample consisting of 1,499 US students across 21 US PhD programs, we show that there is no significant difference in the time it takes US male and female physics PhD students to complete their degree programs. This result comes in spite of a statistically significant 18 percentile point gap in median GRE-P scores between genders. Additional analyses reveal that there is no statistical difference between US students reported as White, Black/Hispanic/Multiracial/Native American, and Asian. Expanding our sample to also include 1,143 Non-US students, we find a small but significant effect of citizenship status on time to PhD completion where the average time for Non-US students to complete a physics PhD is about two months less than their US student counterparts. These results show that in spite of known gaps in standardized admissions exams between genders, these differences are not reflected in subsequent graduate school performance. Our findings reinforce the need for graduate admissions committees to go beyond quantitative metrics and conduct a holistic assessment of an applicant's potential to perform research effectively and to earn a PhD.

I. INTRODUCTION

Despite progress in recent years, physics remains one of the least diverse of all the STEM fields. Only 5% of physics PhDs are granted annually to students identifying as an underrepresented racial/ethnic category (i.e., Black, Hispanic, and Native American), and women earn only 20% of physics PhDs [1]. Taken together with the fact that there are significant differences in typical GRE performance between students of different demographic backgrounds [2], the prospect that GRE tests may limit the ability of certain students to enter graduate school has led researchers to begin questioning the utility of GRE exam scores in identifying successful physics graduate students [2–6]. Some recent work has shown that although there is a major gap in GRE Physics scores between male and female test takers (approximately a 20 point difference in median GRE Physics percentile) [3], this gap does not manifest itself in several subsequent metrics of graduate performance. For instance, there is no statistical difference between average graduate grades earned by male and female physics graduate students [4]. However, previous work has not included analyses of the time it takes students to complete their PhD programs. This study aims to fill this gap by extending the analyses of [3] and [4] to investigate whether there are differences in time to PhD completion between demographic groups based on gender, race/ethnicity, and citizenship status.

Along with graduate grades and overall PhD completion, faculty often consider graduation in a timely manner to be an important measure of graduate student success [7]. For departments tasked with supporting students financially for the duration of their doctoral study, identifying the best ways to support students of diverse backgrounds throughout their graduate careers is imperative [8]. However, studies of differences in overall PhD completion and time to completion by demographics such as gender have traditionally been hindered by a lack of data [9]. Some of this research has suggested that although women may be just as likely as men to complete a PhD, their time to degree completion is longer, particularly in the natural sciences [10–13]. Yet others have found no statistical difference in time to completion between male and female PhD students regardless of discipline, although students in the natural sciences tended to complete faster overall than those in the humanities [14–16]. Notably, one study of engineering PhDs with a sample size of over 9,000 students found no overall differences by gender in the completion percentage of PhDs. With regard to time to PhD completion, the authors found that female engineering students progressed faster than their male counterparts [17]. Thus, further research on this topic is clearly necessary, particularly within the context of specific disciplines such as physics.

———————
* mveroste@ur.rochester.edu

This analysis seeks to answer two primary research questions:

1. For US physics students, is there a gap in time to PhD completion depending on gender and/or race/ethnicity?
2. Among both US and Non-US students, is there a gap in time to completion by citizenship status and/or gender?

We begin in Sec. II by describing the data used in this study to answer these research questions along with the statistical methods used in results Section III. Data visualization and two-way analysis of variance (ANOVA) tests are used to answer the primary research questions. Lastly, interpretation of these results and their context within recent research into graduate admissions is discussed in Sec. IV.

II. METHODS

*a. **Data:*** To investigate whether male and female students experience different time to degree completion, this study utilizes the data collected and analyzed in two prior studies [3, 4]. This consists of student level data from a subset of U.S. physics and astronomy departments that awarded more than 10 PhD's per year for students who matriculated between 2000 and 2010, including information on the final disposition of students (PhD earned or not) and degree program start and finish years. Time to degree completion is calculated by subtracting the year a student started their PhD program from the finish year. Since the data collected only includes each student's start and finish years, the time to completion is only available at the scale of integer valued years.

The data was collected from 27 programs (approximately a 42% response rate) spanning broad range of National Research Council (NRC) rankings. However, time to completion was only available for the 21 programs for which start and finish years were reported for students who completed their degrees. These data covered 3565 students (see Table I). Of this subset, a total of 953 students did not complete their PhD programs. Since we are interested in studying time to degree completion, the sample for this study excludes these students, thereby reducing the sample size to 2642 students across 21 programs. This corresponds to approximately 9% of matriculants to all U.S. physics PhD programs during the years studied. Within this sample, $N = 1499$ are US students and $N = 1143$ are Non-US students.

Demographic data collected includes gender, race/ethnicity, and citizenship status. Among the sample of US students who completed a PhD, 17% are women ($N = 248$). Although the authors generally advocate for a nuanced treatment of gender in physics education research and recognize the deficits associated with treating gender as a fixed binary variable [18], the present data set spans the years 2000 to 2010 during which the data reported by programs only allowed for the binary option of male/female. Hence, we treat gender as a dichotomous variable in this analysis.

The racial/ethnic composition of the data set is 65.8% White, 0.9% Black, 2.1% Hispanic, 0.1% Native American,

		US Students Only					All students		
		White	B/H/M/N	Asian	Unknown	**Total**	US	Non-US	**Total**
PhD	Male	832	43	43	333	*1251*	1251	926	*2177*
	Female	154	9	23	62	*248*	248	217	*465*
	Total	986	52	66	395	**1499**	1499	1143	**2642**
No PhD	Male	248	38	12	143	*441*	441	285	*726*
	Female	62	9	6	39	*116*	116	81	*197*
	Total	*310*	*47*	*18*	*182*	**557**	*557*	*366*	**923**

Table I: A descriptive summary of the number of students in the data, broken down by demographics and PhD earned.

4.4% Asian, 0.5% multiple or other races, and 26.4% undisclosed. Excluding the cases for which race/ethnicity was unavailable, the sample is roughly representative of annual PhD production in U.S. physics for gender, race/ethnicity, and citizenship [19]. However, due to the limited number of students identifying as part of an underrepresented group in physics, we choose to combine all students identifying as Black, Hispanic, Multiracial, or Native American into a single category referred to as B/H/M/N as recommended by Williams [20]. Although we acknowledge that grouping students in this manner may mask distinct experiences of specific racial and ethnic groups of students [21], failing to do so would limit the power of subsequent analyses.

b. Statistical Methods: First, we analyze the sample of US students who completed their PhD programs using a two-way ANOVA with gender and race/ethnicity as independent variables and time to completion as the dependent variable. This allows us to determine whether there are significant differences in mean time to completion between students grouped by gender and race/ethnicity, and whether an interaction exists between these variables. Then in a second two-way ANOVA, we expand the data set to include all US and Non-US students, using gender and citizenship status as independent variables and time to completion as the dependent variable. This subsequent analysis allows for the investigation of whether the mean time to degree completion is significantly different between US and international students, and whether those differences vary based on gender.

Analysis of variance tests are similar to t-tests in that both are applied to determine whether significant differences exist between the means of different groups of data. Our study tests two independent variables at a time, and therefore the tests are referred to as two-way analysis of variance tests.

ANOVA tests produce an F-statistic, which is interpreted as a ratio of the amount of variation in the data explained by a model to the unexplained variation [22–24]. So for example, if there exists a large difference in average time to degree completion between men and women, then we would expect gender to explain a lot of the variance in how long it takes students to complete their degree programs. The more variance that is explained by gender, the larger the F value becomes. On the other hand if the average time to completion varies little between men and women then gender will not explain

much variation in the data, and the F statistic will be close to 1. Thus larger values of F mean that the effect under investigation is more likely to be significant.

We also report several measures of effect size, including ω^2 and Cohen's d, associated with each independent variable. Similar to R^2 in regression analysis, ω^2 describes the fraction of variance in the dependent variable explained by an independent variable and is an indication of strength of association between the two. Cohen's d provides a standardized difference between two specific group means, calculated as the difference in means between the two groups divided by their pooled standard deviation.

III. RESULTS

a. Gender and Race/Ethnicity: We begin our investigation into the effects of gender and race/ethnicity on time to completion with a visual comparison of the data. Using the student data on time to completion, we first calculate and plot the proportion of students who complete their PhD by each particular year in the program. These are shown by solid lines in Fig. 1 and are referred to as empirical cumulative distribution functions (ECDF). The ECDF is discrete because programs only reported start and finish years, meaning the time to completion is only available at the scale of integer valued years. To aid in visualization, a smoothed version of the ECDF was generated using kernel density estimation, and is shown by the dashed lines overlaying the data's empirical distribution.

Group	Sum of Squares	*df*	Mean Square	*F*	*p*
Gender	3.44	1	3.44	2.33	0.13
Race/Ethnicity	1.17	3	0.39	0.26	0.85
Interaction	6.13	3	2.04	1.38	0.25
Residual	2202.34	1491	1.48	–	–

Table II: A two-way ANOVA using gender and race/ethnicity as independent variables and time to completion as the dependent variable. No significant differences were found in time to PhD completion by gender or race/ethnicity.
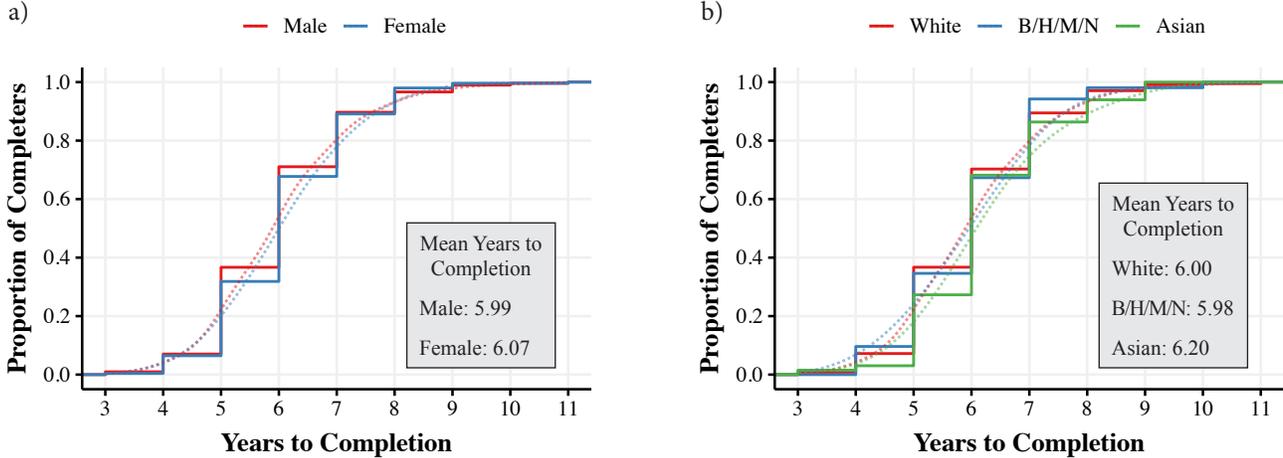
Figure 1: The empirical distribution functions (ECDF) for time to completion for the sample of US students. The solid line in a) shows the ECDF of years to completion for male and female PhD completers, while the solid line in b) shows the empirical distribution of years to completion for White, B/H/M/N, and Asian PhD completers. We exclude students for whom race/ethnicity is unknown from this plot to make visualization of the other groups easier. A smoothed version of the ECDF was generated using kernel density estimation, and is shown by the dashed lines overlaying the data's empirical distribution.

Fig. 1a suggests little difference in the time taken by males and females to complete their PhD programs, indicated by the dashed lines showing the ECDF for each group closely tracking one another. After the 5th year, our data indicates that approximately 36% of male students and 32% of female students had completed their PhD; by the 7th year, 89% of both groups had completed their programs. Fig. 1b similarly shows little difference in time to degree completion between White, B/H/M/N, and Asian students. However, we emphasize that these plots are purely descriptive.

A two-way analysis of variance test of the effects of gender and race/ethnicity on time to completion reveals no significant main effects for gender or race/ethnicity. The main effect of gender on time to degree completion was not statistically significant, $F(1, 1491) = 2.33$, $p = 0.127$, $\omega^2 = .001$. Thus time to completion is not statistically different between male ($M = 5.99$, $SD = 1.23$) and female ($M = 6.07$, $SD = 1.16$) PhD completers within the US student sample. Indeed, there is less than a tenth of a standard deviation difference between the means, $d = -0.06$, a negligible effect size.

Similarly, the main effect of race/ethnicity on time to degree completion was not statistically significant, $F(3, 1491) = 0.264$, $p = 0.851$, $\omega^2 = -.002$. Although the main effect is not significant, we still conduct a Tukey post-hoc test to explore individual differences between groups, as recommended in [25]. However, none of the pairwise comparisons are statistically different. Hence, the time it takes White ($M = 6.00$, $SD = 1.223$), B/H/M/N ($M = 5.98$, $SD = 1.18$), and Asian ($M = 6.197$, $SD = 1.21$) students to complete PhDs are not statistically different from one another. There was also no statistically significant interaction term, $F(3, 1491) = 1.38$, $p = 0.246$, $\omega^2 = 0.001$.

b. Gender and Citizenship Status: Turning our attention to the sample of all students (both US and Non-US), we again begin with a visual comparison of the data. Fig. 2 shows the ECDF of years to completion for US and Non-US PhD completers, and suggests little difference in the time to completion between the two groups. However, we note that the slightly higher proportion of Non-US students to complete their programs in years 4 and 5 may lower the mean time to completion for Non-US students enough for subsequent ANOVA tests to yield a statistically significant result.

A two-way analysis of variance test of the effects of gender and citizenship status on time to completion reveals that the main effect of citizenship status is small but statistically significant, $F(1, 2638) = 4.83$, $p = 0.03$, $\omega^2 = 0.002$. Specifically, Non-US students ($M = 5.87$, $SD = 1.24$) take slightly less time to complete their degree programs than US students ($M = 6.01$, $SD = 1.22$). The difference in means is small,

Group	Sum of Squares	df	Mean Square	F	p
Gender	1.29	1	1.29	0.86	0.35
Citz	7.25	1	7.25	4.83	0.03
Interaction	1.41	1	1.41	0.94	0.33
Residual	3960.64	2638	1.50	–	–

Table III: A two-way ANOVA using gender and citizenship status as independent variables and time to completion as the dependent variable. No significant difference in time to PhD completion is found by gender. A small but significant effect of citizenship status on time to PhD completion indicates that Non-US students complete physics PhD programs slightly faster than US students by about 2 months on average.
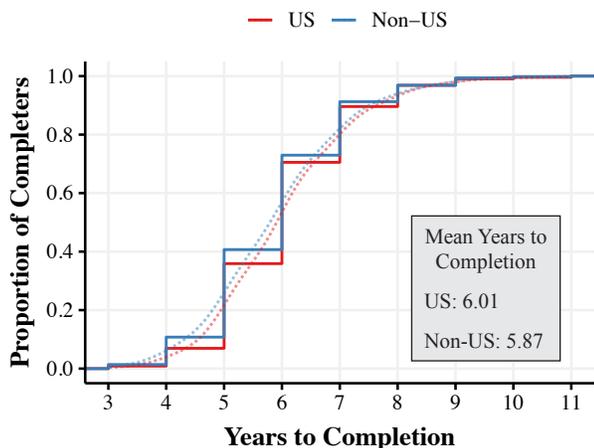
Figure 2: The empirical distribution functions (ECDF) for time to completion for US and Non-US students (solid lines). A smoothed version of the ECDF was generated using kernel density estimation (dashed lines).

$d = 0.11$, approximately a tenth of a standard deviation.

Reinforcing the previous result found using just the sample of US students, the main effect of gender on time to completion is not statistically significant $F(1, 2638) = 0.86$, $p = 0.354$, $\omega^2 < .001$. Time to completion is therefore not significantly different between male ($M = 5.94$, $SD = 1.24$) and female ($M = 5.96$, $SD = 1.18$) PhD completers among the entire student sample, regardless of citizenship status. For this sample we find $d = -0.01$, a negligible difference between the means.

The difference in time to completion between US and Non-US students was the same across each level of gender, indicated by the non-significant interaction term, $F(1, 2638) = 0.94$, $p = .333$, $\omega^2 < 0.001$.

IV. DISCUSSION AND CONCLUSIONS

The results presented in Section III show that in spite of known gaps in standardized admissions exams between different demographic groups [2, 3], there are no significant differences in time to degree completion by gender or race/ethnicity.

In both the sample consisting of only US students and the sample consisting of both US and Non-US students, male and female students complete their PhD programs at similar rates. Thus despite the existence of a large gender gap in GRE-P scores (the median GRE-P percentile is 35 for females and 57 for males), the disparity in GRE-P scores between male and female test takers is anomalous. The disparity does not appear to be related to differences in ability or level of preparation and is not reflected in subsequent graduate performance. Male and female graduate students earn nearly indistinguishable graduate grades, there is no practical relationship be-

tween gender and PhD completion, and there is no significant difference in the time it takes for male and female physics graduate students to complete doctoral degrees [4]. The differences in time to completion between White, B/H/M/N, and Asian students are not significant either. Although large gaps exist in GRE-P scores by racial/ethnic groups (Median GRE-P score for White students is 59, B/H/M/N students is 42, and Asian students is 70), these differences are not reflected in subsequent time to degree completion either.

Although there is a statistically significant difference in time to completion between US and Non-US students, the difference is small. The mean time to completion for US students is 6.01 years while the mean time to completion for Non-US students is 5.87 years. The difference of 0.14 years translates to approximately a difference of just under 2 months. Such a small difference is unlikely to be observable in most programs.

There are several limitations to this work. The analyses presented here only discuss differences in time to completion between demographic groups. They do not answer questions related to whether quantitative metrics such as undergraduate GPA, GRE scores, or graduate GPA predict time to degree completion. Future work will incorporate these metrics in order to address these questions. We also note that exclusion of students who did not complete their PhD from the study has the potential to introduce bias in the results [26] since we cannot know how long those students would have taken to complete their programs had they indeed finished. However, given previous results [3] indicating that gender and race are not predictive of overall PhD completion for U.S. students, we suspect this effect, if any, to be small. Lastly, due to the data collection process we were only able to measure time to completion at the scale of integer years. Although here we considered time to completion to be a continuous variable, we recognize it may be more appropriate to model it as a count variable. We plan to address both of these concerns in future work by investigating the appropriateness of linear regression versus other types of regression to model time to completion.

The findings presented in this paper reinforce the need for graduate admissions committees to conduct a holistic assessment of an applicant's potential to perform research effectively and to earn a PhD. Gaps in standardized test scores between various demographic groups are unexplained and anomalous. Yet no current standardized test measures the research and project management skills it takes to successfully complete a multi-year research project, despite the fact that these skills that are highly valued in PhD graduates. Hence, identifying a broad set of applicant characteristics that predict graduate student outcomes is essential. We encourage the continued investigation of both the physics graduate admissions process and the experiences of students in PhD programs, as well as how those students are taught, mentored, and supported through their growth as individuals within a larger scientific community.

[1] NSF - Science and Engineering Doctorates: (SED Interactive), Available at https://www.nsf.gov/statistics/2016/nsf16300/digest/ (2021).

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

[3] C. W. Miller, B. M. Zwickl, J. R. Posselt, R. T. Silvestrini, and T. Hodapp, Typical physics Ph.D. admissions criteria limit access to underrepresented groups but fail to predict doctoral completion, Science Advances **5**, eaat7550 (2019).

[4] M. Verostek, C. Miller, and B. Zwickl, Analyzing admissions metrics as predictors of graduate gpa and whether graduate gpa mediates phd completion (2021), arXiv:2104.08591 [physics.ed-ph].

[5] N. T. Young and M. D. Caballero, The Physics GRE does not help applicants "stand out" (2020), arXiv:2008.10712 [physics.ed-ph].

[6] E. M. Levesque, R. Bezanson, and G. R. Tremblay, Physics GRE Scores of Prize Postdoctoral Fellows in Astronomy (2015), arXiv:1512.03709 [physics.ed-ph].

[7] 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**, 10.1002/j.2333-8504.2002.tb01875.x (2002).

[8] J. Rodwell and R. Neumann, Predictors of timely doctoral student completions by type of attendance: the utility of a pragmatic approach, Journal of Higher Education Policy and Management **30**, 65 (2008).

[9] S. J. Ceci, D. K. Ginther, S. Kahn, and W. M. Williams, Women in academic science: A changing landscape, Psychological science in the public interest **15**, 75 (2014).

[10] J. Abedi and E. Benkin, The effects of students' academic, financial, and demographic variables on time to the doctorate, Research in Higher Education **27**, 3 (1987).

[11] H. Tuckman *et al.*, *On Time to the Doctorate. A Study of the Increased Time To Complete Doctorates in Science and Engineering.* (ERIC, 1990).

[12] I. Moses, Planning for quality in graduate studies, in *Quality in postgraduate education* (Routledge, 1994) pp. 4–13.

[13] J. G. Baker, Gender, race and Ph. D. completion in natural science and engineering, Economics of Education Review **17**, 179 (1998).

[14] R. G. Ehrenberg and P. G. Mavros, Do doctoral students' financial support patterns affect their times-to-degree and completion probabilities?, Journal of human resources , 581 (1995).

[15] B. C. Seagram, J. Gould, and S. W. Pyke, An investigation of gender and other variables on time to completion of doctoral degrees, Research in higher education **39**, 319 (1998).

[16] T. Wrightand R. Cochrane, Factors influencing successful submission of phd theses, Studies in higher education **25**, 181 (2000).

[17] M. T. Nettles, C. M. Millett, and C. M. Millett, *Three magic letters: Getting to Ph. D.* (JHU Press, 2006).

[18] A. L. Traxler, X. C. Cid, J. Blue, and R. Barthelemy, Enriching gender in physics education research: A binary past and a complex future, Physical Review Physics Education Research **12**, 020114 (2016).

[19] U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Integrated Postsecondary System (IPEDS), Available at https://nces.ed.gov/ipeds/ (2021).

[20] T. L. Williams, 'Underrepresented Minority' Considered Harmful, Racist Language, Available at https://cacm.acm.org/ (2021).

[21] D. Shafer, M. S. Mahmood, and T. Stelzer, Impact of broad categorization on statistical results: How underrepresented minority designation can mask the struggles of both asian american and african american students, Physical Review Physics Education Research **17**, 010113 (2021).

[22] A. Field, J. Miles, and Z. Field, *Discovering Statistics Using R*, 1st ed. (SAGE Publications Ltd, London; Thousand Oaks, Calif, 2012).

[23] D. C. Howell, *Statistical methods for psychology* (Cengage Learning, 2012).

[24] G. R. Iversen, H. Norpoth, and H. P. Norpoth, *Analysis of variance*, 1 (Sage, 1987).

[25] J. Hsu, *Multiple comparisons: theory and methods* (CRC Press, 1996).

[26] A. R. Small, Range restriction, admissions criteria, and correlation studies of standardized tests (2017), arXiv:1709.02895 [physics.ed-ph].