Analysis of Praxis physics subject assessment examinees and performance: Who are our prospective physics teachers?

Lisa Shah,¹ Jie Hao,² Christian A. Rodriguez,¹ Rebekah Fallin,² Kimberly Linenberger-Cortes,³
Herman E. Ray,^{2,4} and Gregory T. Rushton^{1,5}

¹*Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, USA*

²*Department of Statistics and Analytical Science, Kennesaw State University,
Kennesaw, Georgia 30144, USA*

³*Department of Chemistry and Biochemistry, Kennesaw State University, Kennesaw, Georgia 30144, USA*

⁴*Analytics and Data Science Institute, Kennesaw State University, Kennesaw, Georgia 30144, USA*

⁵*Institute for STEM Education, Stony Brook University, Stony Brook, New York 11794, USA*

(Received 9 November 2017; published 10 May 2018)

A generally agreed upon tenant of the physics teaching community is the centrality of subject-specific expertise in effective teaching. However, studies which assess the content knowledge of incoming K–12 physics teachers in the U.S. have not yet been reported. Similarly lacking are studies on if or how the demographic makeup of aspiring physics educators is different from previously reported analyses of the actual high school physics teaching workforce. Here we present findings about the demographics and subject knowledge of prospective high school physics teachers using data from Praxis physics subject assessments administered between 2006 and 2016. Our analysis reveals significant variations in exam participation and performance between men and women, as well as those with different undergraduate majors and academic performance over the past decade. Findings from this work inform understandings and decisions about the quality, recruitment, and preparation of the high school physics teaching workforce.

DOI: [10.1103/PhysRevPhysEducRes.14.010126](https://doi.org/10.1103/PhysRevPhysEducRes.14.010126)

I. INTRODUCTION

Teachers have a significant impact on the success of their students in the classroom and beyond into the workforce, as the effectiveness of STEM teachers is particularly critical for preparation of highly qualified STEM professionals. Studies have reported strong correlations between teacher qualifications, which are often used as proxies for assessing effective teaching, and student achievement in STEM [1–3]. In-field degrees (i.e., those consistent with the subject being taught) have been among the most impactful teacher qualifications, as they are typically an indication of having had extensive academic training in a subject. What and how much a teacher understands about a subject is reasonably expected to influence their ability to teach it effectively [4–6].

In recognition of the importance of subject matter knowledge in effective teaching, one of the final requirements on the pathway to becoming a beginning teacher is often the demonstration of content knowledge proficiency on a subject certification exam. In the United States, the

Praxis physics subject assessments is one such exam administered in 36 states and Washington D.C. over the past decade [7]. The 100-question multiple choice assessment includes topics commonly encountered in an introductory college physics course, including mechanics; electricity and magnetism; optics and waves; and heat, energy, and thermodynamics [8]. To be recommended for certification, prospective teachers must demonstrate some minimum level of proficiency as designated by content experts and state officials [9].

II. BACKGROUND AND PRIOR RESEARCH

Previous studies of physics teacher content knowledge have established the centrality of subject expertise in effective teaching. These studies have examined physics teachers' topic-specific content knowledge [10–12] and its effect on their self-confidence in teaching physics [13]. Knowledge of physics has been identified as an important foundation for being able to appropriately relay this knowledge to students [4–6]. While these studies have highlighted the importance of strong teacher content knowledge, to what extent incoming physics teachers in the U.S. demonstrate understanding as they prepare to lead students in physics classrooms has not yet been investigated. Such an assessment would serve to evaluate the current state of physics education and identify ways to improve it.

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

Additionally, a number of recent reports from the physics education community have generally concluded that the teaching workforce suffers from a lack of diversity with respect to gender [14,15], though far fewer studies have acknowledged the lack of racial diversity [16]. Several reports by the American Institute for Physics (AIP) have found (i) lagging participation of African American, Black, and Hispanic high school students in physics relative to other races or ethnicities [17], (ii) no detectable changes in the participation of African American and Black students in undergraduate physical sciences and engineering programs [18], and (iii) and a recent decline in the already low percentage of women and African American and Black students pursuing physics degrees [19]. Furthermore, our work and that of others in this area has also built on these findings, establishing that the diversity of the high school physics teaching workforce has not changed appreciably in the past two decades with respect to race and ethnicity (i.e., 90% white in 1987 to 87% white in 2011) and has made just minor improvements with respect to gender equity (i.e., 20% female in 1987 to 32% female in 2011) [20–22]. However, evaluations of the diversity of *prospective* physics teachers (i.e., aspiring professionals) relative to the diversity of the resulting workforce have not yet been conducted. Insights from these analyses should help inform future policy and initiatives aimed at improving the diversity of the physics teaching workforce.

III. STUDY CONTEXT AND RESEARCH QUESTIONS

The goal of the presented work was to understand the demographics of prospective high school physics teachers in the United States in the past decade and the extent to which these test takers with different personal and professional characteristics demonstrated knowledge of the subject prior to beginning their first physics teaching assignment. We report findings from our analysis of Praxis physics subject assessments data spanning past ten years to (i) characterize examinees by their personal and professional demographics, and (ii) evaluate their exam performance in relation to these demographic factors. The presented work is guided by the following research questions:

1. What have been the personal and professional characteristics of those who have taken the Praxis physics subject assessments in the past decade?
2. How have these personal and professional characteristics correlated with Praxis physics subject assessment performance in the past decade?

IV. DATA SOURCES AND METHODOLOGY

A. Praxis physics subject assessments

The Praxis subject assessments are a series of exams intended to assess the subject-specific content knowledge and related skills of beginning K–12 teaching candidates.

Those pursuing careers as teachers take the exam as part of the certification process in the majority of U.S. states (currently 36 states and Washington D.C., see Appendix). Exam questions are prepared by panels of expert educators, teacher preparation faculty, and subject specialists and subsequently reviewed by ETS for validity, reliability, and issues of bias before administration [9]. Once examinees have taken the exam, passing scores are set at the state level (described in more detail in Sec. C).

B. Study sample

All Praxis subject assessments examinees between the ages of 18 and 75 who took the exam sometime between June 2006 and May 2016 were included in our analysis ($N = 9667$). Restricting the age range removed outlier examinees who may have misreported their date of birth. To avoid multiple scores for examinees, only the highest score was analyzed for repeat test takers. Information regarding test takers' personal and professional characteristics (e.g., race or ethnicity, gender, undergraduate major) was obtained from their self-reported responses to the demographic questionnaire accompanying their exam. A detailed analysis of examinee demographics is reported in the Appendix (Table III). To ensure an adequate level of privacy for examinees, the state postal codes of each test takers' reported home address were used to group examinees by U.S. census region. As our analysis is performed at the level of the entire Praxis physics subject assessment population, it does not warrant the use of statistical parameters such as p values that are typically applied to analyses of population samples. We, therefore, do not report such error estimates in our findings.

C. Common passing standard

Praxis physics subject assessment performance is reported as a scaled score between 100 and 200, which is a function of both the raw percentage of correct responses and the difficulty of the exam. To assist in the determination of state passing scores, ETS conducts a multistate standard-setting study, where content experts and current or former physics teachers evaluate the probability that a beginning physics teacher would correctly respond an exam item. After several rounds of discussion, these judgments are summed and averaged to yield a final recommended passing score. Individual states are provided with this information, which is considered when setting their own passing standard (i.e., “cut” score), which may change from year to year [9]. Examinees that obtain scaled scores at or above a state's cut score are considered to have passed the exam. Following the precedence of Gitomer and colleagues [23], we used publicly available data on individual state passing scores and assumed a common, national passing standard. Using the median cut score across all Praxis physics subject assessment states in the analyzed time frame (see Appendix, Table IV), we assumed a passing

standard of 140 for all analyses. To approximate the corresponding percent correct (as an evaluation of the threshold level of content knowledge needed for entry into the profession), we took the median percent correct scores of all individuals who earned an exact scaled score of 140 in the past decade, which resulted in an estimated corresponding percent correct score of 54%.

D. Model selection

Since the scaled scores were approximately normally distributed, a linear regression model was used to identify the demographic and/or personal characteristics of test takers that were most significantly related to their exam performance. A stepwise linear regression approach was used to identify the most appropriate linear model from the set of candidate independent variables (i.e., gender, race or ethnicity, undergraduate major). This stepwise procedure enters or removes one candidate per step based on specified information criteria, such as Akaike's information criterion (AIC) and Schwarz Bayesian criterion (SBC) [24,25]. We utilized SBC as the selection criteria, which tends to suggest simpler models with lower dimensionality than AIC. We chose the tenfold cross validation (CV) as the stopping criterion, which reduces bias caused by variable selection. The SBC is defined to be

$$\text{SBC} = n \times \ln(\text{SSE}/n) + p \times \ln(n), \quad (1)$$

where SSE is the sum of squared errors, n is the sample size, and p is the number of parameters included in the model. The stepwise approach yields a single optimal model; however, there is typically more than one equivalent optimal model with slightly different combinations of variables. It was, therefore, necessary for us to manually include several key variables based on our previous experience and knowledge of these analytic methodologies. This process was employed to generate an aggregate model that identified which personal and professional characteristics were most strongly associated with Praxis physics subject assessment performance. Regression analysis was performed for each of the ten years independently to identify the most significant characteristics or combinations of characteristics for each year. Undergraduate major, undergraduate GPA, and gender were present in the year-level models for eight, eight, and seven of the ten years analyzed, while all other variables were present in the models for no more than two years. Only these most prevalent demographic characteristics were included as variables in the final model and are reported on in Sec. V. The final model explained 24% of the variation in scores (total η^2 , Table I). The total η^2 for each model can be used to determine the partial contribution of each variable or interaction to the overall total variance at each step. Variables that are not included in the final model produced partial η^2 values that did not significantly

TABLE I. Stepwise linear regression models including top examinee characteristics most strongly associated with performance on the Praxis physics subject assessment from 2006 to 2016.

Stepwise models	Total η^2
UMajor	0.15
UMajor*Gender	0.20
UMajor*Gender*UGPA	0.24

improve the predictive capability of the model. A detailed analysis of only the most predictive variables or interactions are presented in Sec. V.

E. Differential item functioning (DIF)

To examine whether individual test questions functioned differently for similarly performing examinees with respect to gender and race or ethnicity, we performed differential item functioning (DIF) analysis following closely the procedures previously published by ETS [26,27] and briefly described here. For each test form used between June 2006 and May 2016,¹ a new exam score was calculated for examinees after removing items that displayed differences in the likelihood of responding correctly with respect to race or ethnicity or gender. Test takers were then matched into quartiles of similar performers for each test form using this new total exam score. To estimate the relative probabilities of the focal group (i.e., female, Black, Hispanic) responding correctly to a question relative to the reference group (i.e., male, White), a logistic regression model was used to calculate the MH D-DIF statistic [27,28] for each item across forms. MH D-DIF was then used to sort items into one of three categories: A, B, and C, where (i) "A" items had MH D-DIF statistics that were less than 1.0, (ii) "C" items had MH D-DIF values statistically greater than 1.5, and (iii) all remaining items with MH D-DIF statistics between 1.0 and 1.5 fell into category "B." Findings from DIF analyses are reported in Tables V and VI (see Appendix).

V. RESULTS

A. Median cut scores across states using the Praxis physics subject assessment

Figure 1 shows the state-level, median scaled score required for examinees to be recommended for a physics teaching certification in a particular state between 2006 and 2016. For example, candidates seeking licensure in

¹ETS convened a National Advisory Committee and conducted a job survey and revised test specifications went into effect in 2014; however, the changes were small enough that the new version of the test was able to be equated to the old version, avoiding rescaling and allowing equated scaled scores from the old and new versions to be comparable.

FIG. 1. Median Praxis physics subject assessment cut scores from 2006 to 2016. Individual data reported by each state was used to determine the minimum scaled score needed to be awarded certification in the state. The median of these values for each state between 2006 and 2016 is depicted. States were assigned a score if they accepted Praxis physics subject assessment testing for certification in any year within this timeframe. Those that did not are shaded in gray. A scaled score of 140 corresponds to a raw percentage score of approximately 54%. Source: Derived from data provided by Educational Testing Service.

Pennsylvania during this timeframe would have needed a score of at least 140 (i.e., the state's median cut score in this 10-year period) to be recommended for certification in that state. These cut scores ranged from 126 to 153 (which correspond to approximate percent correct values of 46% and 64%) across all of the states that accepted the Praxis physics subject assessment in the past decade. The national median value of 140 (approximately 54%) was set as our estimated standard across all states for passing the exam when performing the analyses described in the subsequent sections.

B. Demographics

Several relevant personal and professional characteristics of Praxis physics subject assessment test takers in the past decade are summarized in Table II. A total of 9667 individuals took the Praxis physics subject assessment between June 2006 and May 2016, 73.1% of whom were likely to have passed by our previously defined standard (i.e., scaled score greater than or equal to 140). 36.9% of examinees were female, who were 20% less likely to pass relative to their male counterparts (i.e., 60.4% compared to 80.4%). An overwhelming majority of the test-taking

population reported their race as White (85.6%), while Black and Hispanic individuals represented just 3.6% and 1.7% of examinees, respectively. These underrepresented groups have historically passed the exam at much lower rates, with Black test takers being approximately 1.7 times less successful at passing (43.6%) relative to White test takers (73.7%). Black or White and Hispanic or White DIF analysis revealed an average of 13.2% and 6.3% category C items (i.e., those exhibiting relatively large performance differentials), respectively, per test form (see Appendix, Table V), though this percentage has dropped to 0% for both sets of focal and reference groups in the two most recent administrations of the exam. More than 78.7% of examinees reported undergraduate GPAs above 3.0, and those with higher GPAs have had a higher pass rate. Additionally, physics and engineering majors, who have had relatively high pass rates, have made up just under one-third (32.5%) of the test-taking population in the past decade.

C. Undergraduate and graduate major

A prospective physics teaching candidate's undergraduate major explained 15% of the overall variance in Praxis

TABLE II. Personal and professional characteristics of Praxis physics subject assessment examinees from 2006 to 2016.

	N_{Total}	% Pass
Gender		
Male	6073	80.4
Female	3555	60.4
Race or ethnicity		
White	7803	73.7
Black	326	43.6
Hispanic	156	62.8
Other	826	74.3
Undergraduate GPA		
3.5–4.0	3916	76.6
3.0–3.49	3693	71.1
2.5–2.99	1402	67.1
Below 2.5	184	61.4
Undergraduate major		
Physics or engineering	3138	86.4
Chemistry	946	70.8
Biology	945	55.1
(STEM) Education	1005	60.8
STEM Other	957	65.9
Non-STEM Other	568	54.6
Overall	9667	73.1

physics subject assessment performance (Table I). Physics and engineering majors outperformed others by about 10–20 scaled points, with examinees with chemistry, biology, (STEM) education, and both types of other majors performing within 10 points of one another [Fig. 2(a)]. However, both biology and non-STEM other majors have performed just at the national median cut score, while all other examinees have had average scores above this threshold. Though graduate major was not identified as a significant predictor of performance [possibly due to the lower reporting rate (59%) to this questionnaire item and, therefore, limited data available for analysis], it is included

for comparison purposes [Fig. 2(b)]. Physics and engineering majors were still the highest performers, but the subpopulation of graduate majors in the field did not perform much better than the population of undergraduate majors. Test takers with reported graduate majors in (STEM) education now outperformed examinees with majors in chemistry or other STEM disciplines, and performed roughly 8 scaled points higher (151) than the population of (STEM) education undergraduate majors (143). Examinees reporting graduate majors in chemistry and both "other STEM" and non-STEM disciplines performed several scaled points higher than those with only undergraduate majors in these areas. Finally, graduate majors in biology performed as well as undergraduate majors in this field.

D. Gender

An examinee's gender was also identified by our model as a significant predictor of Praxis physics subject assessment performance [Fig. 3(a)]. As shown in Fig. 3, males consistently scored approximately 10 points higher than females, an achievement gap that has persisted in each of the past ten years with little year-to-year variation in the mean difference. Examinees of both genders performed at or above the median passing score across states (dashed line) in each of the past ten years, though average scores for females have been within 5 points of the median in each year. A comparison of performance between male and female physics and engineering majors [Fig. 3(b)] reveals that while scores for both groups are higher than those for average males and females [Fig. 3(a)], the difference in performance between genders remains even after controlling for undergraduate major. Gender DIF analysis revealed an average of 17.6% category C items per test form, or that 17.6% of exam items (on average) exhibited a difference in performance such that the probability of answering correctly was at least 89% higher for males than females.

FIG. 2. Praxis physics subject assessment performance by examinees' reported (a) undergraduate major and (b) graduate major. Scaled score is plotted on the y axis. Dashed line represents the median passing score of 140 across states. Source: Derived from data provided by the Educational Testing Service.

FIG. 3. Praxis physics subject assessment performance by (a) all examinees' reported gender and (b) the indicated gender of examinees reporting physics and engineering undergraduate majors. Scaled score is plotted against academic year (i.e., academic year 2015 = June 2015–May 2016). Dashed line represents the median passing score of 140 across states. Source: Derived from data provided by the Educational Testing Service.

Notably, however, this percentage of category C items has decreased to approximately 7% and 5% on two of the most recent (i.e., 2016) administrations of the exam (see Appendix, Table V).

E. Undergraduate GPA

An examinee's undergraduate GPA was also shown to be significantly correlated with performance on the Praxis physics subject assessment between 2006 and 2016 (Fig. 4). Those with higher GPAs tended to outperform those with lower academic standing, and examinees with GPAs in the highest bracket outperformed others by 5–10 points overall. However, on average, even examinees with GPAs below 2.5 have scored above the median (dashed line) by several points.

F. Race or ethnicity

While an examinee's race or ethnicity was not a significant predictor of Praxis physics subject assessment performance as determined by our model (possibly due to the small number of non-White examinees available for analysis), the documented issue of diversity in the physics teaching community warrants a presentation of these results [Fig. 5(a)]. White test takers and those of other races or ethnicities have performed similarly during the analyzed timeframe, with a decade-wide average scaled scores of 151 and 152, respectively. The variability in scores for Hispanic examinees is likely a function of the low numbers of reported test takers in each year, but they have performed at the level of White test takers and those of other race or ethnicities in six of the ten years studied. While analyses of the performance of Black examinees are similarly limited by low exam participation, these test takers have consistently scored lower than these groups by as much as 24 scaled points in 2006 to 11 scaled points in 2011. Further, Black test-takers have scored below the national median cut score (dashed line) in all but two of the ten years analyzed,

whereas average scaled scores for all other reported races or ethnicities have been above our established standard. Even after controlling for undergraduate majors in physics and engineering [Fig. 5(b)], Black examinees averaged 20–25 scaled points lower during this timeframe than all other test takers. Race or ethnicity DIF analysis revealed an average of 16.5% and 10.5% category C items when comparing performance between equally performing (i) White and Black test takers and (ii) White and Hispanic test takers, respectively (see Appendix, Table VI). Specifically, this data indicates that 16.5% of exam items (on average) exhibited a difference in performance such that the probability of answering correctly was at least 89% higher for

FIG. 4. Praxis physics subject assessment performance by examinees' reported undergraduate GPA. Scaled score is plotted on the y axis. Dashed line represents the national median passing score of 140. Source: Derived from data provided by the Educational Testing Service.

FIG. 5. Praxis physics subject assessment performance by (a) examinees' reported race or ethnicity and (b) the reported race or ethnicity of physics and engineering majors. Scaled score is plotted on the y axis. Dashed line represents the national median passing score of 140. Source: Derived from data provided by the Educational Testing Service.

White examinees than Black examinees, and that the same was true of 10.5% of items with respect to White and Hispanic examinees.

VI. DISCUSSION

Our analysis of Praxis physics subject assessment examinees in the past ten years reveals a number of findings about the demographic makeup of prospective high school physics teachers and their exam performance. First, state-to-state passing scores required for candidates to be recommended for certification have ranged from 126 to 153 across the nation (Fig. 1), with a median of 140 across states corresponding to an estimated 54% correct. While state-to-state variations in cut scores may be partially influenced by external factors, such as teacher shortages [9], it is important to consider the educational impact of certifying candidates who have not demonstrated a firm grasp of the subject. Encouragingly, candidates reporting GPAs between 3.5 and 4.0 comprised almost 80% of examinees over the past decade (Table II) and have outperformed candidates with lower GPAs (Fig. 4). While physics and engineering majors have earned some of the highest exam scores relative to those of other majors (Fig. 2), they have comprised just under one-third of the overall examinee population. Additionally, the participation and performance of women in physics has been a growing concern of the physics education community and is apparent in our sample of prospective high school physics teachers taking the Praxis physics assessment [15,16]. The percentage of females who sat for the Praxis physics subject assessment in the past decade has increased only slightly over this time frame (33.7% in 2006 to 40.2% in 2015), and a gender achievement gap has persisted across all exam years analyzed (Fig. 3). While an average of 30% of exam items exhibited differential performance between males and females in the past decade,

it is encouraging to note that gender DIF analysis has demonstrated substantially lower percentage of category C items on the past two of the most recent exam forms. The lack of racial diversity that has been evident in the Physics education community is also reflected in our analysis [17,19,29]. Black and Hispanic Praxis physics subject assessment test takers represented just 3.6% and 1.7% of test takers (Table II), while they comprise 13% and 18% of the U.S. population, respectively [29]. Additionally, using our estimated passing standard, both Black and Hispanic examinees have passed at lower rates than White and Asian candidates, with less than half of all Black test takers passing the Praxis physics subject assessment over the past ten years, and have underperformed relative to all other races or ethnicities even after controlling for undergraduate major.

Our findings should inform understandings and decisions about the quality, recruitment, and preparation of the high school physics teaching workforce. While physics and engineering majors have averaged a scaled score of 158, far exceeding the scores of examinees reporting other majors, this corresponds to an estimated percentage score of 66%. This performance should alert physics faculty to issues with long-term retention of knowledge from introductory physics courses (which these majors are likely to have taken). It is possible that more traditional curricula and approaches to teaching physics may not appropriately facilitate conceptual retention beyond end-of-course exams [30]. Instructional shifts toward reform-based practices and curricula may be especially conducive to improving student learning in the discipline [31,32]. STEM education majors (who are most likely to pursue teaching careers) were among the lowest performers, averaging a scaled score of 143 (~56%). Attempts to more closely coordinate the efforts of STEM education departments with disciplinary departments may provide both students and faculty with the resources necessary for mutual success. Learning assistant

models, for example, have proven mutually beneficial to both future K–12 educators and participating university faculty in shifting teaching practices [33].

Additionally, even after controlling for physics and engineering majors, a gender achievement gap was apparent on the Praxis physics subject assessment. Furthermore, this achievement gap has remained consistently large during the 2006–2016 timeframe. These data speak to the need to more critically evaluate underlying factors of the U.S. physics educational system that may contribute to these differences, including efforts aimed at changing the culture of physics departments, physics courses, and the discipline of physics in general [34–36]. Furthermore, a number of studies have suggested that reform-based pedagogies have served to not only improve the learning of physics for students broadly but reduce or eliminate gender achievement gaps in university physics classrooms [37–40]. As many of these suggested strategies for change rely on the localized efforts of individual faculty members, this may be an important population to target as part of continued efforts to eliminate performance disparities between males and females.

Lastly, the lack of diversity among physics teaching candidates, which has been an ongoing problem for the community [20,22,41], is evident in the low participation rates and performance among Black and Hispanic examinees. Black examinees in particular have underperformed relative to others even among physics and engineering majors. As with efforts related to gender inequities, the culture of physics likely plays a large role in the success of underrepresented groups. For example, studies have noted that candidates' race or ethnicity are often not considered as part of the criteria for acceptance into physics doctoral degree programs [42], and this population of students is most likely to transition into future physics faculty members. It is possible that until the face of physics more equitably reflects the diversity of the population at large (which largely depends on current university educators), underrepresentation will continue to afflict the discipline at all levels.

VII. LIMITATIONS

We acknowledge that while the Praxis physics subject assessment series is the most nationally representative

source of information about the content knowledge of teaching candidates, data from individual teacher certification exams across several of the most populous states is not encompassed in our analysis. Our established national cut score of 140, while necessary for the interpretation of our data and based on a precedent set by previous work [23], is somewhat arbitrary and we recognize that other approaches to determining passing status may have yielded slightly different results. For example, using the average of the average state cut scores during this timeframe (138) would have produced higher percent passing rates in Table I than current reported. Additionally, the estimated passing standard does not account for whether candidates *actually* passed the exam in the state in which they sought certification. It is important to note the nuances in examinee motivation for taking the Praxis physics subject assessment as it pertains to our reported results. While some test takers do sit for the exam to become certified in physics as their primary specialty, a number of examinees are likely to have taken the exam to obtain a secondary or tertiary subject certification to improve their competitiveness on the job market. It is, therefore, likely that some percentage of the test takers captured in our data set will not ever go on to teach physics. With respect to the interpretation of race or ethnicity data, it is important to note that the number of nonrespondents (566 examinees) is larger than the number of Black or Hispanic respondents, which may impact the observed results. DIF analysis does not incontrovertibly define test items as biased. Items exhibiting differential performance should be more carefully examined to uncover possible explanations to observed differences. Finally, although certification exams are used as a proxy for teacher effectiveness by measuring content knowledge, we also recognize that several other key variables not accounted for here (e.g., pedagogical content knowledge, school context) may have a significant impact on the success of beginning teachers.

ACKNOWLEDGMENTS

The authors would like to thank Jonathan Steinberg of Educational Testing Service (ETS). This work was supported by National Science Foundation Grants No. DUE-1557285 and No. DUE-1557292. The opinions set forth in this publication are those of the authors and not ETS.

APPENDIX: ANALYSIS OF PRAXIS® PHYSICS SUBJECT ASSESSMENT EXAMINEES AND PERFORMANCE: WHO ARE OUR PROSPECTIVE PHYSICS TEACHERS?

This section containing a complete list of examinee characteristics, individual state cut scores from 2006 to 2016, and summary of DIF analysis.

TABLE III. Detailed list of personal and professional characteristics of Praxis physics subject assessment Examinees from 2006 to 2016. Source: Derived from examinee questionnaire data filled out during test registration, provided by the Educational Testing Service.

Characteristic	Response options	<i>N</i>
Total number of examinees		9667
<i>Gender</i>	Male	6073
	Female	3555
<i>Race or ethnicity</i>	White	7803
	Black	326
	Hispanic	156
	Other	
	Examples: Asian; Southeast Asian; Native American; Pacific Islander	826
<i>Age</i>	Sequential	9667
<i>Years since attended college</i>	Sequential	9199
<i>Undergraduate GPA</i>	3.5–4.0	3916
	3.0–3.49	3693
	2.5–2.99	1402
	Below 2.5 ^a	184
<i>Education level at first time tested</i>	Freshman	8
	Sophomore	30
	Junior	239
	Senior	1166
	Earned Bachelor's degree	1487
	Bachelor's degree + additional credits	2517
	Earned Master's degree	1243
	Master's degree + additional credits	1504
	Earned Doctoral degree	466
<i>Census region^b</i>	Northeast	3503
	Midwest	2461
	South	2776
	West	890
<i>Undergraduate major</i>	Physics and engineering	3138
	Chemistry	946
	Biology	945
	(STEM) Education	1005
	STEM Other	
	Examples: Math and computer information; general science; geology, or earth and space science	957
	non-STEM Other	
	Examples: Art; philosophy; Spanish	568
<i>Graduate major</i>	(STEM) Education	1892
	Physics and engineering	1196
	Chemistry	343
	Biology	248
	STEM Other	425

(Table continued)

TABLE III. (Continued)

Characteristic	Response options	<i>N</i>
<i>Enrolled in teacher prep. program</i>	non-STEM Other	1611
	Currently	3630
	Formerly	3668
	Never	1836
<i>Type of teacher prep. program</i>	Undergraduate	2301
	Fifth year	395
	Master's degree	2112
	Alternative route	2135
	Other	289
<i>Teaching status</i>	Planning to enroll or enrolled in teacher prep. program	3463
	Recently graduated and will begin teaching in the near future	1104
	1–3 years teaching experience	1824
	3+ years teaching experience	2724
	Not planning to teach at this time	49
<i>Geographic area to Teach In^c</i>	Urban	2057
	Rural	1798
	Suburban	2891
	I do not plan on teaching next year	342
<i>Attended Noyce Institution</i>	Yes	3243
	No	6424

^aUndergraduate GPA categories of 2.5–2.99, 2.0–2.49, 1.5–1.99, and below 1.5 were grouped into this collective category.

^bExaminees were grouped into census regions using a state postal code in the original data set.

^cThis question only became part of the questionnaire beginning in 2008.

TABLE IV. Minimum passing scaled scores for 37 U.S. states and D.C. that have accepted the Praxis physics subject assessment from 2006 to 2016. Data was retrieved from annually published Praxis II series passing score by test and state documents from ETS [43–53].

State	Minimum passing (cut) score (scaled score)										
	2016	2015	2014	2013	2012	2011	2010	2009	2008	2007	2006
AK	127	127	129	129	129	129	129	129	129	129	129
AL	138	138	138	138	138	138	138	138	138	138	141
AR	135	135	135	^b	^b	^b	^b	^b	^b	^b	^b
CT	141	141	141	141	141	141	141	141	141	141	141
DC	143	143	143	143	143	143	^b	^b	^b	^b	^b
DE	144	144	144	144	144	144	144	^b	^b	^b	147
HI	153	153	153	153	153	153	153	^b	^b	^b	^b
IA	131	130	129	127	^b	^b	^b	^b	^b	^b	^b
ID	129	129	129	129	129	129	129	129	129	129	129
IN	^b	^b	^b	149	149	149	149	149	149	149	149
KS	141	141	141	141	141	141	141	141	141	141	141
KY	133	133	133	133	133	133	133	133	133	133	133
LA	141	141	141	141	141	141	141	141	141	141	^b
MD	143	143	143	143	143	143	143	143	143	143	143
ME	141	141	^a	^b	^b	^b	^b	^b	^b	^b	^b
MN	^b	^b	^b	^b	^b	^b	^b	137	137	137	137
MO	^b	^b	^b	141	141	141	141	141	141	141	141
MS	139	139	139	139	139	139	139	139	139	139	139
MT	135	135	135	135	135	135	135	^a	^b	^b	^b
NC	133	133	133	133	133	133	^b	^b	^b	^b	^b
ND	132	132	132	132	132	132	132	132	132	132	132
NE	131	131	^b	^b	^b	^b	^b	^b	^b	^b	^b

(Table continued)

TABLE IV. (Continued)

Minimum passing (cut) score (scaled score)											
NH	146	146	146	146	146	146	146	146	146	146	146
NJ	141	141	141	141	141	141	141	141	141	^b	^b
OH	^b	^b	^b	132	132	132	132	132	132	132	132
OR	^b	^b	^b	^b	^b	^b	^b	147	147	^b	^b
PA	140	140	140	140	140	140	140	140	140	140	140
RI	146	146	146	146	^b	^b	^b	^b	^b	^b	^b
SC	130	130	130	^b	^b	^b	^b	^b	^b	^b	^b
SD	130	130	130	130	130	130	130	130	130	130	133
TN	144	144	144	144	144	144	144	144	144	144	144
UT	136	136	136	136	136	136	136	136	136	136	^a
VA	147	147	147	147	147	147	147	147	147	147	147
VT	140	140	140	140	140	140	140	140	140	^b	^b
WA	^b	^b	^b	^b	^b	^b	^b	^b	140	140	140
WI	141	141	141	141	141	^b	^b	^b	^b	^b	^b
WV	126	126	126	126	126	126	126	126	126	126	126
WY	137	137	137	137	137	137	137	137	137	^a	^b

^aTest required, but cut score not available.^bTest not required.

TABLE V. Gender differential item functioning analysis of Praxis physics subject assessments test forms from 2006 to 2016. Item categories (ABC) were determined using form-specific scores for males (reference group) and females (focal group).

TestForm	Sample size	Category	Percent
0265055	3007	A	76.00
		B	14.00
		C	10.00
0265071	3162	A	52.53
		B	22.22
		C	25.25
0265080	3308	A	65.00
		B	21.00
		C	14.00
5265055	476	A	49.00
		B	19.00
		C	32.00
5265071	634	A	59.60
		B	18.18
		C	22.22
5265080	608	A	56.00
		B	16.00
		C	28.00
5265118	714	A	64.00
		B	23.00
		C	13.00
5265126	853	A	61.00
		B	20.00
		C	19.00
5265134	89	A	92.93
		B	0.00
		C	7.00
5265142	164	A	95.00
		B	0.00
		C	5.00

TABLE VI. Race or ethnicity differential item functioning analysis of Praxis physics subject assessments test forms from 2006 to 2016. Item categories (ABC) were determined using form-specific scores for White (reference group) and each listed ethnicity (focal group). Data is listed as “focal group vs reference group.” DIF analysis for test forms 5265134 and 5265142 are not included due to an insufficient number of test takers of each race or ethnicity for these most recently administered forms.

TestForm	Sample size	Effect	Category	Percent
0265055	2919	Black vs White	A	84.00
			B	2.00
			C	14.00
		Hispanic vs White	A	89.00
			B	4.00
			C	7.00
		Other vs White	A	88.00
			B	7.00
			C	5.00
0265071	3096	Black vs White	A	80.81
			B	3.03
			C	16.16
		Hispanic vs White	A	84.85
			B	3.03
			C	12.12
		Other vs White	A	83.84
			B	6.06
			C	10.10
0265080	3201	Black vs White	A	71.00
			B	3.00
			C	26.00
		Hispanic vs White	A	75.00
			B	2.00
			C	23.00
		Other vs White	A	79.00
			B	12.00
			C	9.00
5265055	436	Black vs White	A	83.00
			B	0.00
			C	17.00
		Hispanic vs White	A	91.00
			B	3.00
			C	6.00
		Other vs White	A	87.00
			B	2.00
			C	11.00
5265071	595	Black vs White	A	87.88
			B	0.00
			C	12.12
		Hispanic vs White	A	88.89
			B	0.00
			C	11.11
		Other vs White	A	96.97
			B	0.00
			C	3.03

(Table continued)

TABLE VI. (Continued)

TestForm	Sample size	Effect	Category	Percent
5265080	551	Black vs White	A	84.00
			B	0.00
			C	16.00
		Hispanic vs White	A	91.00
			B	2.00
			C	7.00
		Other vs White	A	93.00
			B	4.00
			C	3.00
5265118	634	Black vs White	A	91.00
			B	1.00
			C	8.00
		Hispanic vs White	A	92.00
			B	0.00
			C	8.00
		Other vs White	A	94.00
			B	1.00
			C	5.00
5265126	764	Black vs White	A	76.00
			B	1.00
			C	23.00
		Hispanic vs White	A	85.00
			B	5.00
			C	10.00
		Other vs White	A	84.00
			B	6.00
			C	10.00

- [1] L. Darling-Hammond, Teacher quality and student achievement, *Educ. Policy Anal. Arch.* **8**, 1 (2000).
- [2] M. Fetler, High school staff characteristics and mathematics test results, *Educ. Policy Anal. Arch.* **7**, 9 (1999).
- [3] C. T. Clotfelter, H. F. Ladd, and J. L. Vigdor, Teacher-student matching and the assessment of teacher effectiveness, *Journal of health and human resources administration* **41**, 4 (2007).
- [4] L. Shulman, Knowledge and teaching: foundations of the new reform, *Harv. Educ. Rev.* **57**, 1 (1987).
- [5] W. R. Veal, D. J. Tippins, and J. Bell, The evolution of pedagogical content knowledge in prospective secondary physics teachers, *Proceedings of the Annual Meeting of the National Association for Research in Science Teaching*, 1998 (San Diego, CA, 1998).
- [6] R. M. Sperandio-Mineo, C. Fazio, and G. Tarantino, Pedagogical content knowledge development and pre-service physics teacher education: A case study, *Res. Sci. Educ.* **36**, 235 (2006).
- [7] Educational Testing Service, The Praxis Test Information Bulletin, ETS (2017) https://www.ets.org/s/praxis/pdf/praxis_information_bulletin.pdf.
- [8] Educational Testing Service, The Praxis Study Companion Physics: Content Knowledge, ETS (2017) <https://www.ets.org/s/praxis/pdf/5265.pdf>.
- [9] Educational Testing Service, Technical Manual for The Praxis Series, and Related Assessments, ETS (2015) https://www.ets.org/s/praxis/pdf/technical_manual.pdf.
- [10] M. Z. Hashweh, Effects of subject matter knowledge in the teaching of biology and physics, *Teach. Teach. Educ.* **3**, 109 (1987).
- [11] V. F. Marion, An analysis of the relationship between teachers' acquisition of physics content knowledge and their level of science teaching efficacy, Ph.D. thesis, Harvard University, Cambridge, Massachusetts, 1998.
- [12] I. Ali, Exploring secondary school physics teachers' content knowledge of physics: A study from Gilgit-Baltistan, M. Ed. thesis, Aga Khan University, 2014.
- [13] A. D. Robertson, R. E. Scherr, L. M. Goodhew, A. R. Daane, K. E. Gray, and L. B. Aker, Identifying content knowledge for teaching energy: Examples from high school physics, *Phys. Rev. Phys. Educ. Res.* **13**, 1 (2017).
- [14] J. Johnston and M. Ahtee, Comparing primary student teachers' attitudes, subject knowledge and pedagogical

- content knowledge needs in a physics activity, *Teach. Teach. Educ.* **22**, 503 (2006).
- [15] R. Scherr, Never mind the gap: Gender-related research in Physical Review Physics Education Research, 2005–2016, *Phys. Rev. Phys. Educ. Res.* **12**, 2 (2016).
- [16] R. M. Lock and Z. Hazari, Discussing underrepresentation as a means to facilitating female students' physics identity development, *Phys. Rev. Phys. Educ. Res.* **12**, 020101 (2016).
- [17] G. White, Race and physics teaching, and the fair: A call to all physics educators for manuscripts on a rarely discussed topic, *Phys. Teach.* **54**, 70 (2016).
- [18] S. White and J. Tyler, Underrepresented minorities in high school physics, *Statistical Research Center* (American Institute of Physics, College Park, MD, 2015).
- [19] L. Merner, African American participation among bachelors in the physical sciences and engineering, *Statistical Research Center* (American Institute of Physics, College Park, MD, 2015).
- [20] P. J. Mulvey and S. Nicholson, Physics bachelor's degrees, *Statistical Research Center* (American Institute of Physics, College Park, MD, 2012).
- [21] S. White and J. Tyler, Who teaches high school physics? Results from the 2012–13 nationwide survey of high school physics teachers, *Statistical Research Center* (American Institute of Physics, College Park, MD, 2014).
- [22] G. T. Rushton, D. Rosengrant, A. Dewar, L. Shah, H. E. Ray, K. Sheppard, and L. Watanabe, Towards a high quality high school workforce: A longitudinal, demographic analysis of U.S. public school physics teachers, *Phys. Rev. Phys. Educ. Res.* **13**, 020122 (2017).
- [23] D. H. Gitomer, T. L. Brown, and J. Bonett, Useful signal or unnecessary obstacle? The role of basic skills tests in teacher preparation, *J. Teach. Educ.* **62**, 431 (2011).
- [24] H. Akaike, A new look at the statistical model identification, *IEEE Trans. Autom. Control* **19**, 6 (1974).
- [25] G. Schwarz, Estimating the dimension of a model, *Ann. Stat.* **6**, 461 (1978).
- [26] M. Zieky, A DIF primer, *Educational Testing Service Center for Education in Assessment*, (Educational Testing Service, Princeton, NJ, 2003).
- [27] R. Zwick, *A review of ETS differential item functioning assessment procedures: Flagging rules, minimum sample size requirements, and criterion refinement*, ETS Research Report Series (Educational Testing Service, Princeton, NJ, 2012).
- [28] H. J. Rogers and H. Swaminathan, A comparison of logistic regression and Mantel-Haenszel procedures for detecting differential item functioning, *Appl. Psychol. Meas.* **17**, 105 (1993).
- [29] U.S. Census Bureau, Population Estimates. Race and Hispanic Origin, <https://www.census.gov/quickfacts/table/PST045216/00> (accessed May 8, 2017).
- [30] M. J. Cahill, K. M. Hynes, R. Trousil, L. A. Brooks, M. A. McDaniel, M. Repice, J. Zhao, and R. F. Frey, Multiyear, multi-instructor evaluation of a large-class interactive engagement curriculum, *Phys. Rev. ST Phys. Educ. Res.* **10**, 020101 (2014).
- [31] L. Deslauriers, E. Schelew, and C. Wieman, Improved learning in a large-enrollment physics class, *Science* **332**, 862 (2011).
- [32] M. Milner-Bolotin, D. Egersdorfer, and M. Vinayagam, Investigating the effect of question-driven pedagogy on the development of physics teacher candidates' pedagogical content knowledge, *Phys. Rev. Phys. Educ. Res.* **12**, 020128 (2016).
- [33] K. E. Gray, D. C. Webb, and V. K. Otero, Effects of the learning assistant model on teacher practice, *Phys. Rev. Phys. Educ. Res.* **12**, 020126 (2016).
- [34] L. Schiebinger, *Has Feminism Changed Science?* (Harvard University Press, Cambridge, MA, London, 2001).
- [35] M. Ong, Body projects of young women of color in physics: Intersections of gender, race, and science, *Social problems* **52**, 593 (2005).
- [36] S. L. Li and M. E. Loverude, Identity and belonging: Are you a physicist (chemist)?, *AIP Conf. Proc.* **1513**, 246 (2013).
- [37] S. W. Brahmia, Improving learning for underrepresented groups in physics for engineering majors, *AIP Conf. Proc.* **1064**, 7 (2008).
- [38] E. Mazur, *Peer Instruction: A User's Manual* (Prentice Hall, Englewood Cliffs, NJ, 1997).
- [39] L. C. McDermott, and P. S. Shaffer, *Tutorials in Introductory Physics* (Prentice Hall, Englewood Cliffs, NJ, 1998).
- [40] P. Heller, R. Keith, and S. Anderson, Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving, *Am. J. Phys.* **60**, 627 (1992).
- [41] S. White and C. L. Tesfaye, Under-represented minorities in high school physics: Results from the 2008–09 Nationwide Survey of High School Physics Teachers, *Statistical Research Center* (American Institute of Physics, College Park, MD, 2011).
- [42] G. Potvin, D. Chari, and T. 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).
- [43] Educational Testing Service, The Praxis® Series Passing Scores by Test, and State, ETS (2006).
- [44] Educational Testing Service, The Praxis® Series Passing Scores by Test, and State, ETS (2007).
- [45] Educational Testing Service, The Praxis® Series Passing Scores by Test, and State, ETS (2008).
- [46] Educational Testing Service, The Praxis® Series Passing Scores by Test, and State, ETS (2009).
- [47] Educational Testing Service, The Praxis® Series Passing Scores by Test, and State, ETS (2010).
- [48] Educational Testing Service, The Praxis® Series Passing Scores by Test, and State, ETS (2011).
- [49] Educational Testing Service, The Praxis® Series Passing Scores by Test, and State, ETS (2012).
- [50] Educational Testing Service, The Praxis® Series Passing Scores by Test, and State, ETS (2013).
- [51] Educational Testing Service, The Praxis® Series Passing Scores by Test, and State, ETS (2014).
- [52] Educational Testing Service, The Praxis® Series Passing Scores by Test, and State, ETS (2015).
- [53] Educational Testing Service, The Praxis® Series Passing Scores by Test, and State, ETS (2016).