

Research Report
ETS RR-14-35

A Comparison of Achievement Gaps and Test-Taker Characteristics on Computer-Delivered and Paper-Delivered *Praxis I*® Tests

Jonathan Steinberg

Meghan Brenneman

Karen Castellano

Peng Lin

Susanne Miller

December 2014

ETS Research Report Series

EIGNOR EXECUTIVE EDITOR

James Carlson
Principal Psychometrician

ASSOCIATE EDITORS

Beata Beigman Klebanov
Research Scientist

Heather Buzick
Research Scientist

Brent Bridgeman
Distinguished Presidential Appointee

Keelan Evanini
Managing Research Scientist

Marna Golub-Smith
Principal Psychometrician

Shelby Haberman
Distinguished Presidential Appointee

Donald Powers
Managing Principal Research Scientist

Gautam Puhan
Senior Psychometrician

John Sabatini
Managing Principal Research Scientist

Matthias von Davier
Senior Research Director

Rebecca Zwick
Distinguished Presidential Appointee

PRODUCTION EDITORS

Kim Fryer
Manager, Editing Services

Ayleen Stellhorn
Editor

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Report series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Report series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

RESEARCH REPORT

A Comparison of Achievement Gaps and Test-Taker Characteristics on Computer-Delivered and Paper-Delivered *Praxis I*[®] Tests

Jonathan Steinberg, Meghan Brenneman, Karen Castellano, Peng Lin, & Susanne Miller

Educational Testing Service, Princeton, NJ

Test providers are increasingly moving toward exclusively administering assessments by computer. Computerized testing is becoming more desirable for test takers because of increased opportunities to test, faster turnaround of individual scores, or perhaps other factors, offering potential benefits for those who may be struggling to pass licensure examinations. This report extends previous research examining 4 years of paper-based *Praxis I*[®] test data to its corresponding computerized environment with these goals: (a) determine the extent to which achievement gaps exist on computer-based (CB) *Praxis I* tests, (b) examine which test-taker characteristics are most associated with mode, and (c) elicit opinions from previous *Praxis I* test takers to understand reasons why a particular testing mode was chosen. The results contribute to the literature about *Praxis*[®] examinations and reinforce the need to understand their dynamics in the context of evaluating performance and participation gaps by demographic characteristics.

Keywords Achievement gap; *Praxis I*; teacher licensure; paper-based testing; computer-based testing

doi:10.1002/ets2.12033

The existing comparison studies on large-scale assessments such as the *SAT*[®] (Gallagher, Bridgeman, & Cahalan, 2002), *GRE*[®] (Schaeffer, Steffen, Golub-Smith, Mills, & Durso, 1995), and *TOEFL*[®] (Taylor, Jamieson, Eignor, & Kirsch, 1998) tests were mostly conducted when computer infrastructure systems to support automated test delivery were in development and computer familiarity was not as prevalent as it is today. As computer-based (CB) testing generally becomes more prominent and paper-based (PB) testing becomes less prominent, it is important to discuss trends in test-taker characteristics and performance in a CB setting in relation to those in a PB setting. This is essential for the *Praxis*[®] test suite of teacher licensure examinations because of the high-stakes nature of the score results for test takers and the decisions made by states in determining minimum passing scores for these licensing exams. Additionally, differences in performance results and/or test-taking populations between testing modalities hinder the ability to clearly articulate achievement gaps if significant differences by modality exist.

The research contained in this report accomplished three objectives. The first objective was to examine mean performance differences and passing rates among test takers of CB *Praxis I*[®] exams in Reading, Writing, and Mathematics and compare those data to an independent sample of PB test takers during a similar time frame. The second goal was to determine what background characteristics were most associated with testing mode, comparing CB test-taker characteristics to the same independent PB sample. The final objective focused on evaluating opinions of additional independent samples of CB and PB *Praxis I* test takers as to motivations for choosing their respective modes of testing.

Literature Review

The relevant literature for this report can be summarized by three key themes: (a) previous studies about performance gaps for *Praxis I* tests, (b) quantitative methods and results utilizing large-scale testing program data to explore differences in performance between CB and PB testing, and (c) previous research regarding test-taker attitudes relevant to the findings presented in this report. The literature review was essential for framing the direction of the research presented here, as some comparisons will be made between CB and PB *Praxis I* testing.

Corresponding author: J. Steinberg, E-mail: JSteinberg@ets.org

Previous Research About Performance Gaps for Praxis I Tests

Tyler et al. (2011) showed that performance and passing rate gaps were largest with respect to performance comparisons of White test takers with those from other minority race/ethnicity populations, a finding consistent with research previously conducted by Camara and Schmidt (1999) concerning performance gaps on other standardized admissions tests. Nettles, Scatton, Steinberg, and Tyler (2011) conducted a comprehensive analysis of performance differences between African American and White prospective teachers on the PB version of Praxis I tests. The study made use of extensive information collected about test takers at the time of registration.¹ The performance differences were analyzed in aggregate and by specific subgroups based on previous research cited by these authors justifying the selection of certain demographic characteristics of test takers. Praxis I PB performance data were obtained from first-time test takers between November 2005 and November 2009 in 28 states administering the exams.²

The data presented in Nettles et al. (2011) showed that relative to Camara and Schmidt (1999), the Praxis I Reading score gap between African American and White candidates was larger than that for the SAT and more like that for the GMAT. The gap in Praxis I Writing was more comparable to that for the SAT, and the gap in Praxis I Mathematics was larger than that for the GRE Quantitative section. In terms of effect sizes (Cohen, 1988), all standardized score differences on Praxis I tests were considered to be large (i.e., greater than 0.80). The passing rates reflected the African American-White test taker passing rate gap. African American first-time test takers had a lower passing rate than White first-time test takers for each Praxis I exam. The Praxis I Mathematics exam had the largest gap in passing rate but was just slightly higher than the gap in Reading.

Nettles et al. (2011) also found that race/ethnicity, undergraduate grade point average (UGPA), undergraduate major, and the selectivity of the candidate's attending institution explained some portion of the variance in predicting scale score performance on each Praxis I test using regression analysis. The analyses indicated that being an African American candidate and majoring in education were associated with reduced Praxis I scores, whereas having a relatively high UGPA and attending a relatively selective college or university contributed to higher performance on Praxis I tests.

The regression models in the preceding study showed the significant predictors of Praxis I score performance to be background characteristics that were static or not easily changeable by the time an individual took the test. As suggested by the authors of that study, undergraduate major may not have been static; because African American candidates tended to take Praxis I tests later in their academic careers, they may have had a greater likelihood of switching to becoming education majors.

Quantitative Methods and Results Involving Large-Scale Testing Programs

The quantitative methods used to compare PB and CB test results have generally involved the direct calculation of mean differences (e.g., for GRE, Schaeffer et al., 1995, 1998) or the computation of standardized mean differences (SMD) or effect sizes (e.g., for Praxis, Parshall & Kromrey, 1993; Puhan, Boughton, & Kim, 2005). It is important to note the difference between *practically* significant results and *statistically* significant results. When statistically significant results were obtained using the direct calculation of mean differences, *t*-test results could be significant due to the large sample sizes used in these studies. Therefore, effect sizes were computed to standardize the differences, sometimes referred to as SMDs (Puhan et al., 2005). Generally, effect sizes greater than 0.20 in absolute value would be considered practically significant (Cohen, 1988). Gallagher et al. (2002) also used the terms *increased impact* and *decreased impact* to characterize the differences obtained using the SMD method. These authors also found small subgroup differences between testing modes in their study based on gender and race/ethnicity.

Previous Research Concerning Test-Taker Attitudes

The benefits of CB testing in general are positive for test takers due to a faster turnaround time for scoring and reporting of results and more flexibility in scheduling; for organizations, administrative and scoring costs can be significantly lower (Wang & Shin, 2009). Two previous studies discussed specific test-taker attitudes toward computerized testing relative to paper testing. Bridgeman, Lennon, and Jackenthal (2001) surveyed takers of the PSAT[®] test and asked which mode of testing they preferred. The results were that 44% preferred the computer version, 35% preferred the paper version, and 20% had no preference either way. In relation to expected test performance, 66% felt they would have done about the same

on paper as on computer, 20% would have done better on paper, and 14% would have done better on computer. Finally, 59% found CB testing less tiring than PB testing, 30% found it to be no less tiring, and 11% found the CB testing more tiring. Way, Davis, and Fitzpatrick (2006) in recent surveys found that students in a large state's K–12 testing program felt greater levels of comfort testing on computer and tended to prefer that mode of testing compared to PB testing.

Current Study

The research presented here focused on results from CB Praxis I testing, with some connections made to previous PB testing results (Nettles et al., 2011). In this study, we employed a mixed methods design comprising quantitative and qualitative aspects of examining Praxis I testing performance in addressing the following three research questions:

1. What are the achievement gaps between African American and White prospective teacher candidates under CB testing based on background characteristics?
2. What background characteristics of prospective African American and White teacher candidates are most associated with mode of testing?
3. What are some of the attitudes and opinions of prospective teacher candidates about choosing in which mode to take Praxis I?

Data Source and Sample

The Praxis I series measures basic skills in reading, writing, and mathematics. In each of the 28 selected Praxis I participating states, teacher candidates are required to pass each of the Praxis I exams to fulfill initial licensure requirements. The tests are offered in both PB and CB modes. For any administration, a test taker will choose only one mode. Although the number of items and time limits for each subject (e.g., Reading, Mathematics) are different between the two modes (see Table 1), the scale score ranges (150–190) are identical between modes. In addition, the minimum passing scores (also known as cut scores) in each participating state are also identical between modes, despite varying across the 28 participating states (Reading, 170–178; Writing, 171–176; Mathematics, 169–178).

As shown in Table 1, the average time per item is slightly higher for most parts of the CB exams. However, the essay specifications on the Writing test were identical. As shown in Table 2, the average internal consistency reliabilities based on a pool of active forms were similar between modes as well (Educational Testing Service [ETS], 2010).³

The tests selected for this research included the three CB Praxis I tests. The sample for analyses included people who tested between November 2005 and November 2009, spanning 51 test administrations. This sample represents a similar time frame as the PB analyses described in Nettles et al. (2011) and Tyler et al. (2011). A discussion of findings from the CB analyses relative to the PB analyses can be found toward the end of the report. The corresponding PB data comprised

Table 1 Summary of Praxis I Test Structure by Mode of Testing

Praxis I Test	Paper delivered			Computer delivered		
	Items	Timing (min)	Time/item (min)	Items	Timing (min)	Time/item (min)
Reading	40 MC	60	1.50	46 MC	75	1.63
Writing (2 sections)	38 MC	30	0.79	44 MC	38	0.86
	1 Essay	30	30.00	1 Essay	30	30.00
Mathematics	40 MC	60	1.50	46 MC	75	1.63

Note. MC = multiple choice.

Table 2 Internal Consistency Reliability Estimates for Praxis I Tests

Praxis I Test	Paper based	Computer based
Reading	0.87	0.87
Writing	0.72	0.68
Mathematics	0.87	0.88

only 20 administrations in this time period. It should be acknowledged that because the choice of testing mode for any administration is left to the individual test taker, there can be an inherent selection bias for choosing CB or PB testing. Therefore the CB and PB samples cannot be treated as randomly equivalent. The findings are based on aggregate observed performance, without direct comparisons of test takers who may have tested in both modes during the specified time period, more typical of a randomized control study. Therefore any differences in performance cannot be directly attributed to the mode of testing. The primary focus of this research was on first-time test takers, as there were too many factors to consider in evaluating repeat test takers given different possible test-taking patterns by mode.

Methodology

The methods used in this report contained some formal statistical tests but, for the most part, were descriptive in nature. In answering the first research question on achievement gaps by race/ethnicity and other relevant demographics, a similar descriptive approach was taken as in Nettles et al. (2011), showing percentages of CB test takers by race/ethnicity for each category of the following six demographic variables of interest: UGPA, teacher education program enrollment status, candidate educational attainment, parental educational attainment as a proxy for socioeconomic status (SES), undergraduate major, and candidate institutional selectivity. The rationale for selecting these variables can be found in the aforementioned report. Effect sizes (Cohen, 1988) were computed comparing Praxis I scale score performance among White test takers to African American test takers within the same category of each demographic variable. Some categories were recoded so that linear regression models could be constructed examining how demographic variables predict Praxis I performance (see Nettles et al., 2011, Appendix B).

In answering the second research question about which demographic characteristics were associated with testing mode, chi square tests of association would ordinarily be conducted to measure the association between background variables and choice of testing mode based on observed frequencies in the registration data provided in the Praxis background information questionnaire (BIQ). However, such tests are sensitive to sample size such that when large populations are used in these analyses, these are more likely to detect an association where perhaps one might not necessarily exist. It is also important to note that the categories for some background variables are ordered (UGPA, candidate educational attainment, parental educational attainment, and candidate institutional selectivity), whereas others are unordered (teacher education program enrollment status and undergraduate major). Therefore the summary statistics chosen for the forthcoming analyses reflected these considerations.

The primary statistic used for analyses of variables with ordered categories was Cramer's V (Cramer, 1946), a nominal measure of association. As Rea and Parker (1992) explained, the minimum value for any degree of association is 0.10. In this study, tests were conducted separately by race/ethnicity group to see if the association between demographic variables and mode of testing differed at all based on the observed frequencies. When the value of Cramer's V was at least 0.10, the adjusted residuals (Agresti, 1996) were also reported to examine the significant result(s) to uncover their potential sources, because there may have been one or more response categories for a BIQ item with a greater representation of test takers than other categories for that same BIQ item. Adjusted residual values of at least 1.96 in absolute value were considered to be significant. The interpretation is that a positive value indicates that more people than expected within the response category chose to take the test on computer compared to paper. The primary statistic used for unordered categories was the tau statistic (Goodman & Kruskal, 1954). Tau measures the proportional reduction in error in predicting the observed frequencies in a contingency table for each value of the dependent variable when there was no ordering of categories (Reynolds, 1984). Formulas for these statistics can be found in Appendix A.

For the final research question regarding attitudes and opinions of Praxis I based on mode of testing, in March 2011, an open-ended Internet survey question was presented to Praxis I examinees during the third quarter of 2010. Therefore it should be made clear that opinions are only reflective of this particular sample and are not generalizable to the sample used in the quantitative analyses. Due to shortened response length, uploading participant responses to a qualitative data analysis software program (e.g., NVivo 9) would not have been efficient. As a result, the coding team used Microsoft Excel to code responses. The first step was to determine interrater agreement within the sample between the two members of the coding team, who are two of the coauthors on this report. The coding team analyzed approximately 20% of the sample initially to determine agreement. According to Saldana (2012), no standard exists among qualitative researchers, but an 85–90% range seems to be a minimal benchmark. Once this threshold was reached, the remaining responses were coded. Analyses were generally limited to frequencies based on the coded responses.

Results

Summary Statistics

Table 3 presents the population of first-time Praxis I test takers in each prominent race/ethnicity group within the 28 participating states by mode spanning November 2005 to November 2009. The independent PB sample comes from Nettles et al. (2011).

The total group of first-time test takers on computer was about 10–15% lower than on paper. African American and White candidates (19% and 73% on average, respectively) made up the vast majority (over 90%) of the first-time CB test takers. The proportion of CB test takers from these two race/ethnicity groups represented a practically significant difference compared to the PB testing volumes. Participation among African American test takers was greater on computer compared to paper, whereas participation among White test takers was greater on paper compared to computer, yet the samples were still large enough to explore differences in performance by demographic variables. Proportions for other race/ethnicity groups were generally similar between modes.⁴

Table 4 presents the average scale scores and passing rates for candidates in the five predominant race/ethnicity groups across the selected 28 states on each of the three Praxis I exams taken on paper and on computer. The differences (gaps) between the subgroups and the White test-taker group are expressed in standard deviation units,⁵ also known as

Table 3 Sample Sizes and Percentages of All Praxis I First-Time Test Takers Between November 2005 and November 2009 by Race/Ethnicity and Test Mode

Race/ethnicity	Praxis I Reading				Praxis I Writing				Praxis I Mathematics			
	CB	%	PB	%	CB	%	PB	%	CB	%	PB	%
White	50,377	73	65,782	83	50,069	74	65,792	84	50,975	73	64,637	84
African American	13,413	19	8,408	11	12,816	19	8,213	10	13,631	20	8,117	11
Asian American	2,564	4	2,251	3	2,368	3	2,244	3	2,376	3	2,198	3
Hispanic	2,230	3	1,901	2	2,217	3	1,909	2	2,257	3	1,887	2
Native American	377	1	450	1	381	1	457	1	382	1	435	1
Total	68,961		78,792		67,851		78,615		69,621		77,274	

Note. CB = computer-based testing; PB = paper-based testing.

Table 4 Praxis I First-Time Test-Taker Performance and Pass Rate Differences by Race/Ethnicity and Testing Mode

Praxis I Test	Race/ethnicity	Paper-based testing						Computer-based testing						Between modes	
		N	M	SD	d	Pass %	Pass % diff.	N	M	SD	d	Pass %	Pass % diff.	d	Pass % diff.
RD	W	65,782	178.03	5.43		81.5		50,377	179.75	5.62		85.6		0.31	4.1
	AF	8,408	171.61	7.08	-1.14	40.7	-40.8	13,413	173.55	7.24	-1.03	53.8	-31.8	0.27	13.1
	AS	2,251	174.09	7.45	-0.71	57.2	-24.3	2,564	176.28	7.54	-0.61	67.3	-18.3	0.29	10.1
	H	1,901	175.06	7.08	-0.54	64.7	-16.8	2,230	175.95	7.49	-0.67	66.4	-19.2	0.12	1.7
	NA	450	175.33	7.06	-0.50	65.1	-16.4	377	177.16	6.71	-0.46	70.8	-14.8	0.27	5.7
WR	W	65,792	175.96	4.17		79.5		50,069	176.81	4.63		82.7		0.19	3.2
	AF	8,213	171.97	4.23	-0.95	44.2	-35.3	12,816	172.74	4.50	-0.89	53.7	-29.0	0.17	9.5
	AS	2,244	173.82	4.92	-0.51	63.2	-16.3	2,368	175.09	5.15	-0.37	72.4	-10.3	0.25	9.2
	H	1,909	173.71	5.03	-0.54	63.0	-16.5	2,217	173.93	5.45	-0.62	62.3	-20.4	0.04	-0.7
	NA	457	173.69	4.69	-0.54	57.3	-22.2	381	174.87	4.31	0.42	69.0	-13.7	0.26	11.7
MT	W	64,637	178.59	6.89		78.2		50,975	179.67	6.62		83.6		0.16	5.4
	AF	8,117	170.56	7.31	-1.16	36.8	-41.4	13,631	172.51	7.47	-1.05	49.5	-34.1	0.26	12.7
	AS	2,198	177.99	8.08	-0.09	71.2	-7.0	2,376	180.14	7.33	0.07	80.9	-2.7	0.28	9.7
	H	1,887	174.02	8.02	-0.66	57.2	-21.0	2,257	175.19	7.94	-0.67	62.2	-21.4	0.15	5.0
	NA	435	174.51	7.95	-0.59	59.5	-18.7	382	175.95	7.34	-0.56	66.8	-16.8	0.19	7.3

Note. AF = African American; AS = Asian American; d = the standardized difference in mean scores between PB testing and CB testing (CB – PB); H = Hispanic; MT = Mathematics; NA = Native American; RD = Reading; W = White; WR = Writing.

standardized differences or effect sizes, sometimes represented by d (Cohen, 1988).⁶ Values above 0.20 in absolute value are generally considered to be practically significant.

From Table 4, it was found that the mean score for CB test takers was *higher* than that for PB test takers across all race/ethnicity groups. The second finding was that standardized score differences between White test takers and those of other race/ethnicity groups were *smaller* on computer compared to paper, except for Hispanic test takers. In fact, Asian American test takers did slightly better than White test takers on the CB Praxis I Math test.

Using the 0.20 threshold for effect sizes, average test-taker performance for those choosing the CB version was higher compared to the average test-taker performance for those choosing the PB version. This was true for all race/ethnicity groups, except Hispanic test takers in Reading. In Writing, on average, performance by Asian American and Native American teacher candidates was significantly better for those choosing the CB version. In Mathematics, African American and Asian American teacher candidates choosing the CB version performed significantly better on average than those choosing the PB version. An investigation of possible sources of these findings will be touched upon in the following descriptive analyses of score performance controlling for selected background variables, but these may not represent the full array of possible sources for the findings. The reader is reminded that the PB and CB samples were not randomly equivalent.

Average passing rates for those choosing the CB version were generally higher and gaps in passing rates between race/ethnicity groups were smaller. The only exception was for Hispanic test takers, among whom passing rates were higher for those choosing the CB version, except for Writing. Yet the gaps in passing rates between Hispanic and White test takers on all Praxis I tests were not lower for those choosing the CB version compared to the PB version. African American teacher candidates on average generally had the greatest differences in passing rates on all Praxis I tests between those choosing the CB version and those choosing the PB version. The differences were 2–3 times that of White candidates, giving the appearance that the achievement gap is smaller for those choosing the CB version compared to those choosing the PB version for Praxis I. As mentioned earlier, because these were not random samples of test takers, no claims can be made as to whether the differences in the gaps are due to the mode of administration. Additionally, while the statistics appear to indicate that Praxis I represents a specific disadvantage to African American candidates, the reported gaps are generally in line with gaps on other large-scale achievement tests based on data from Camara and Schmidt (1999).

Summary Statistics Controlling for Background Variables

As with analyses of PB version testing data (Nettles et al., 2011), due to the smaller numbers of Hispanic, Asian American, and Native American test takers compared to African American and White test takers, the following analyses in this report on CB test takers based on background variables were focused on the latter two groups of test takers. The same six background variables in Nettles et al. (2011) were examined here: UGPA, teacher program enrollment, candidate educational attainment, parental educational attainment as a proxy for SES, major field of study, and institutional selectivity. The data were examined for test takers who answered those particular background questions and had a valid scale score for the particular Praxis test. Results from each Praxis I test are reported separately for each background variable when crossed with race/ethnicity.

Undergraduate Grade Point Average

Table 5 presents CB performance data based on self-reported UGPA as well as where the largest achievement gaps reside. Whereas 76% of White candidates had UGPAs of 3.0 or higher, only half of African American candidates had UGPAs in these ranges. This is one indicator of how well teacher candidates performed in their undergraduate courses. The score gaps were generally larger for Mathematics, except for the UGPA range 3.5–4.0, where the gaps for Reading and Mathematics were comparable. Across Reading, Writing, and Mathematics, the widest gaps occurred in the highest UGPA range (3.5–4.0). On all three tests, the mean score for African American test takers in the highest UGPA range was just above the White test-taker mean score in the lowest UGPA range. One exception to this trend was that in Writing, the average scale score for African American test takers with a UGPA at or above 3.5 was equal to that of White test takers with a UGPA of 2.0–2.49.

Enrollment Status in Teacher Education Programs

Table 6 focuses on the relationship between Praxis I performance and enrollment status in teacher education programs. While an equal proportion of White candidates (39%) were either currently or never enrolled in teacher education

Table 5 Praxis I Performance by Race/Ethnicity Group and Undergraduate Grade Point Average (UGPA) for First-Time Test Takers Choosing Computer Testing

UGPA	Frequency (%)		W test performance		AA test performance		<i>d</i>
	W	AA	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
RD							
1.5–1.99	<1	1	173	7	168	8	–0.63
2.0–2.49	4	11	177	7	172	7	–0.79
2.5–2.99	20	37	178	6	173	7	–0.76
3.0–3.49	39	35	179	6	174	7	–0.91
3.5–4.0	37	16	181	5	175	7	–1.19
WR							
1.5–1.99	<1	1	172	4	170	4	–0.59
2.0–2.49	4	12	174	5	171	4	–0.61
2.5–2.99	20	37	175	4	172	4	–0.69
3.0–3.49	39	35	176	4	173	4	–0.74
3.5–4.0	37	16	178	4	174	5	–0.90
MT							
1.5–1.99	<1	1	173	8	167	7	–0.84
2.0–2.49	4	11	177	7	171	7	–0.84
2.5–2.99	20	36	178	7	172	7	–0.81
3.0–3.49	39	35	179	7	173	7	–0.97
3.5–4.0	37	17	181	6	174	8	–1.18

Note. AA = African American; MT = Mathematics; RD = Reading; W = White; WR = Writing.

Table 6 Praxis I Performance by Race/Ethnicity Group and Enrollment Status in Teacher Education Programs for First-Time Test Takers Choosing Computer Testing

Status	Frequency (%)		W		AA		<i>d</i>
	W	AA	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
RD							
Currently	39	30	179	6	173	7	–0.92
Formerly	23	19	180	5	173	7	–1.40
Never	39	51	180	6	174	7	–1.01
WR							
Currently	39	30	176	5	173	4	–0.83
Formerly	22	18	177	5	172	5	–1.04
Never	39	52	177	5	173	5	–0.86
MT							
Currently	39	31	179	7	172	7	–0.99
Formerly	23	20	180	7	171	7	–1.27
Never	38	49	180	7	173	7	–1.02

Note. AA = African American; MT = Mathematics; RD = Reading; W = White; WR = Writing.

programs when first taking Praxis I, a larger proportion of African American candidates (49–52%) had never been enrolled in a teacher education program. The data presented in Table 6 appear to demonstrate that enrollment status did not have an impact on performance, but the gap between African American and White test takers appeared smallest for those who were currently enrolled in a teacher education program.

Candidate Educational Attainment Level

Table 7 presents the percentages of Praxis I test takers by educational attainment level and the corresponding mean scores and score gaps between African American and White test takers. Slightly more than one-third of the White test takers compared to about one-fifth of the African American test takers were undergraduates when they took the CB Praxis I tests. African American candidates were more likely to have a bachelor's degree or higher when first taking Praxis I (79%)

Table 7 Praxis I Performance by Race/Ethnicity Group and Educational Attainment for First-Time Test Takers Choosing Computer Testing

Educational level	Frequency (%)		W		AA		<i>d</i>
	W	AA	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
RD							
Freshman	4	1	178	6	171	7	-1.23
Sophomore	10	3	177	6	172	7	-0.80
Junior	9	7	177	6	172	7	-0.80
Senior	12	10	179	6	173	7	-0.90
Bachelor's degree	25	32	180	5	174	7	-1.11
Bachelor's degree +	26	30	181	5	174	7	-1.25
Master's degree	9	10	182	4	174	7	-1.43
Master's degree +	6	7	182	4	174	7	-1.64
WR							
Freshman	4	1	176	4	172	5	-1.08
Sophomore	9	3	176	4	172	4	-0.86
Junior	9	6	175	5	172	4	-0.77
Senior	12	10	176	5	173	4	-0.76
Bachelor's degree	25	32	177	5	173	5	-0.92
Bachelor's degree +	26	30	177	4	173	4	-0.95
Master's degree	9	10	178	4	173	4	-1.10
Master's degree +	6	7	178	4	173	5	-1.13
MT							
Freshman	4	1	180	6	173	7	-1.16
Sophomore	9	3	178	6	173	7	-0.92
Junior	9	6	177	7	172	7	-0.77
Senior	12	10	179	7	173	8	-0.87
Bachelor's degree	25	31	180	6	173	8	-1.06
Bachelor's degree +	26	31	180	7	172	7	-1.14
Master's degree	9	10	181	6	172	8	-1.27
Master's degree +	6	7	181	6	172	8	-1.32

Note. AA = African American; MT = Mathematics; RD = Reading; W = White; WR = Writing.

than White candidates (66%). The data presented in Table 7 for the three Praxis I tests indicate that the gap appears to widen as educational attainment level increases and was largest for the Reading test.

Parental Educational Attainment

The highest educational attainment of either parent of the test taker was used as a proxy for SES in this study. Table 8 presents African American and White test-taker mean scores for those choosing the CB version of Praxis I Reading, Writing, and Mathematics along with the gaps in the scores arrayed by level of highest parental educational attainment. As expected, White test takers were better represented among categories of higher parental educational attainment beyond the bachelor's degree than their African American counterparts. For example, around 27% of White test takers' parents attained a graduate or professional degree compared to just under 20% of African American test takers' parents. A larger share of the parents of African American test takers had completed a high school diploma or less than their White counterparts. The largest gaps occurred at the highest levels of parental educational attainment for each Praxis I exam.

Undergraduate Major Field

Table 9 presents the CB Praxis I Reading, Writing, and Mathematics mean scores and gaps between African American and White test takers in the sample arrayed by selected broad major field of study. The percentages of White and African American candidates who were science majors were somewhat comparable at 11–12% for White candidates and 13–14% for African American candidates. Just about half of White test takers (55–56%) and African American test takers (48–50%) majored in education. The gaps were smallest on the Writing test and were largest on the Reading test among humanities and science majors and on the Mathematics test among humanities majors. Education majors achieved the lowest mean

Table 8 Praxis I Performance by Race/Ethnicity Group and Parental Educational Attainment for First-Time Test Takers Choosing Computer Testing

Parental educational attainment	Frequency (%)		W		AA		<i>d</i>
	W	AA	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
RD							
<Some HS	2	8	178	6	171	7	-1.01
HS diploma	18	22	178	6	172	7	-0.98
Some post secondary	15	18	180	5	174	7	-0.90
Associate's degree	9	11	179	6	174	7	-0.87
Bachelor's degree	22	17	180	6	174	7	-0.99
Some graduate or professional school	6	5	180	5	175	7	-1.03
Graduate or professional degree	27	18	181	5	176	7	-1.07
WR							
<Some HS	2	8	174	4	171	4	-0.74
HS diploma	18	22	175	4	172	4	-0.82
Some post secondary	15	18	176	4	173	4	-0.80
Associate's degree	9	11	176	4	173	4	-0.79
Bachelor's degree	22	17	177	5	173	5	-0.87
Some graduate or professional school	6	5	177	5	173	4	-0.84
Graduate or professional degree	27	19	178	5	174	5	-0.89
MT							
<Some HS	2	8	176	7	170	7	-0.87
HS diploma	18	22	178	7	171	7	-0.99
Some post secondary	15	18	179	6	173	7	-0.92
Associate's degree	9	11	179	7	172	7	-0.99
Bachelor's degree	22	17	180	7	173	8	-1.04
Some graduate or professional school	6	5	180	6	173	8	-1.04
Graduate or professional degree	27	19	181	6	174	7	-1.10

Note. AA = African American; HS = high school; MT = Mathematics; RD = Reading; W = White; WR = Writing.

Table 9 Praxis I Performance by Race/Ethnicity Group and Undergraduate Broad Major Field Classification for First-Time Test Takers Choosing Computer Testing

Major	Frequency (%)		W		AA		<i>d</i>
	W	AA	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
RD							
Science	11	14	182	4	175	7	-1.30
Business	5	9	181	4	175	7	-1.26
Social sciences	15	18	182	4	176	7	-1.20
Education	56	50	178	6	171	7	-1.04
Humanities	12	9	183	4	177	7	-1.32
WR							
Science	12	14	178	4	174	4	-0.97
Business	5	10	177	4	173	4	-0.94
Social sciences	16	19	178	5	174	5	-0.96
Education	55	48	176	4	172	4	-0.94
Humanities	13	9	179	4	175	5	-0.94
MT							
Science	11	13	184	5	177	7	-1.18
Business	5	9	182	5	175	7	-1.17
Social sciences	16	19	180	6	173	7	-1.18
Education	56	49	178	7	171	7	-1.07
Humanities	13	10	181	6	173	7	-1.30

Note. Those majoring in technology-related disciplines, those who were undecided, or those whose majors did not fit the five major groupings displayed in this table were removed from this analysis. AA = African American; MT = Mathematics; RD = Reading; W = White; WR = Writing.

Table 10 Praxis I Performance by Race/Ethnicity Group and Institutional Selectivity for First-Time Test Takers Choosing Computer Testing

Selectivity	Frequency (%)		W		AA		<i>d</i>
	W	AA	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
RD							
Noncompetitive	5	7	179	6	172	7	-1.03
Less competitive	9	16	178	6	172	7	-0.86
Competitive	44	53	179	6	173	7	-0.95
Competitive +	3	5	180	6	175	7	-0.79
Very competitive	22	10	181	5	176	7	-0.99
Very competitive +	17	9	183	4	178	6	-0.98
WR							
Noncompetitive	5	7	176	4	172	4	-0.89
Less competitive	8	15	175	5	172	4	-0.79
Competitive	44	54	176	4	172	4	-0.83
Competitive +	3	5	177	5	173	5	-0.75
Very competitive	22	10	177	4	174	5	-0.82
Very competitive +	18	9	180	4	176	5	-0.87
MT							
Noncompetitive	5	8	179	7	171	7	-1.07
Less competitive	9	15	178	7	171	7	-0.86
Competitive	44	53	179	7	172	7	-0.98
Competitive +	3	5	180	6	173	8	-0.98
Very competitive	22	10	181	6	174	7	-1.10
Very competitive +	17	9	183	5	177	8	-1.16

Note. AA = African American; MT = Mathematics; RD = Reading; W = White; WR = Writing.

scores on each of the Praxis I tests, but the gaps across the three CB Praxis I tests, though still large, were generally smaller than those for other majors.

Selectivity of Colleges and Universities Attended

As Nettles et al. (2011) noted, it is important to consider the selectivity of the colleges and universities that teacher candidates attend to assess candidates' scores and compare race/ethnicity group performance on Praxis I. Table 10 shows that the majority of African American and White test takers testing on computer attended competitive institutions, a much smaller share of African American test takers (19%) attended colleges and universities in the two most selective categories than White test takers (39–40%), and that a larger share of African American test takers attended schools in the two least selective categories: 22–23% compared to 13–14% for White test takers. Table 10 also indicates that generally the largest overall gaps were in Mathematics and among candidates attending more selective colleges and universities.

Regression Model Summary

In an attempt to quantify the degree to which background variables predict CB Praxis I scale score performance, a linear stepwise regression analysis was performed. Nettles et al. (2011, Appendix B) describes the procedure for preparing the background data for this analysis based on PB testing data. The full model results based on CB testing data are displayed in Appendix B.

On the basis of a minimum threshold of 1% explained variance in Praxis I CB scale scores to be retained in the model, race/ethnicity, UGPA, undergraduate major, and institutional selectivity met this criterion for analyses of each Praxis I test. In addition, parental educational attainment (proxy for SES) met this criterion for the CB Writing test. These predictors explained 25–29% of the variance in the scores for each of the Praxis I tests based on the r^2 statistic. Table 11 displays the results for those variables retained in the final models.⁷

The regression analyses revealed that White test takers on average had a 3- to 6-point advantage over their African American counterparts on Praxis I, or based on the standardized coefficients that allow for comparisons across models, this range was 0.25–0.34. Having a UGPA at or above 3.0 as opposed to below 3.0 or having a major other than education

Table 11 Summary of Stepwise Regression Results for First-Time Praxis I Test Takers Choosing Computer Testing

	β	SE	Sig.	Std. β	r^2
RD					
Race/ethnicity	-5.25	.07	<.01	-.32	.14
UGPA	2.47	.06	<.01	.17	.03
Major	-3.66	.05	<.01	-.29	.09
Selectivity	2.07	.05	<.01	.16	.02
Par. ed. att.	NA	NA	NA	NA	NA
WR					
Race/ethnicity	-3.12	.06	<.01	-.25	.11
UGPA	2.24	.05	<.01	.21	.05
Major	-2.01	.05	<.01	-.21	.06
Selectivity	1.74	.05	<.01	.18	.03
Par. ed. att.	1.07	.05	<.01	.11	.01
MT					
Race/ethnicity	-6.23	.08	<.01	-.34	.14
UGPA	2.13	.07	<.01	.13	.02
Major	-3.03	.06	<.01	-.21	.06
Selectivity	2.55	.06	<.01	.17	.03
Par. ed. att.	NA	NA	NA	NA	NA

Note. MT = Mathematics; Par. ed. att. = parental educational attainment; RD = Reading; UGPA = undergraduate grade point average; WR = Writing.

gave test takers about a 2- to 4-point average increase in scores (range based on standardized coefficients was 0.13–0.21 for UGPA, 0.21–0.29 for major). Attending a selective college or university on average was associated with Praxis I scores that were 2–3 points higher than those attending a less selective college or university (range based on standardized coefficients was 0.16–0.18). For the Writing test specifically, having at least one parent with at least a bachelor's degree contributed to higher scores.

As the first research question identified those variables related to observed performance gaps, the next section describes those factors that are associated with mode of testing.

Demographic Patterns Among Praxis I Test Takers by Mode

The motivating factors for the second research question were to provide some additional context for the performance analyses reported earlier, to address the fact that the PB and CB groups were not randomly equivalent in nature, and that a candidate only chooses one mode of testing for a single administration, which may or may not be influenced by demographic characteristics. As described in the “Methodology” section, although the same six demographic variables were utilized as in the previous section, some have ordered categories (UGPA, candidate educational attainment level, parental educational attainment as a proxy for SES, and candidate institutional selectivity), whereas the others (teacher education program enrollment status and undergraduate major) have unordered categories. As noted, the statistics used to assess associations between background variables and mode of testing differ based on whether the categories are ordered. Table 12 shows the values of Cramer's V for those variables with ordered categories.

The results in Table 12 indicate that candidate educational attainment level and institutional selectivity demonstrated a statistically significant association ($V \geq 0.10$) with choice of testing mode both for White and African American teacher candidates. Regarding candidate educational attainment level, for African American candidates, the association is moderate based on guidelines in Rea and Parker (1992); for White candidates, the association is relatively strong. Regarding institutional selectivity, a weak association existed with Praxis I testing mode both for White and African American candidates. Therefore the adjusted residuals for educational attainment level (Table 13) and institutional selectivity (Table 14) were further analyzed to identify specific categories contributing to the statistically significant associations.

All adjusted residuals in Table 13 are greater than 1.96 in absolute value and therefore significant. The results demonstrate that those in their first 3 years of college were more likely than expected to take Praxis I on paper and that those with a bachelor's degree or higher were more likely than expected to take Praxis I on computer. White candidates who

Table 12 Summary of Cramer's *V* Statistics of Associative Strength Between Praxis I Testing Mode and Background Variables by Race/Ethnicity Group for First-Time Test Takers

Variable	Reading		Writing		Mathematics	
	W	AA	W	AA	W	AA
UGPA	0.028	0.026	0.029	0.024	0.031	0.026
Educational level	0.456	0.319	0.463	0.328	0.464	0.327
Parental educational attainment	0.075	0.073	0.079	0.071	0.077	0.066
Institutional selectivity	0.153	0.114	0.155	0.119	0.152	0.113

Note. AA = African American; UGPA = undergraduate grade point average; W = White.

Table 13 Summary of Adjusted Residuals Between Praxis I Testing Mode and Candidate Educational Attainment Level by Race/Ethnicity Group for First-Time Test Takers

Educational level	Reading		Writing		Mathematics	
	W	AA	W	AA	W	AA
Freshman	-68.6	-14.8	-69.7	-15.9	-70.3	-16.4
Sophomore	-60.7	-21.6	-62.3	-21.7	-62.7	-21.6
Junior	-26.1	-18.3	-26.2	-18.2	-26.3	-19.5
Senior	13.3	-8.3	13.7	-6.7	13.2	-7.9
Bachelor's degree	46.8	13.1	47.0	13.3	46.8	12.2
Bachelor's degree +	56.6	15.1	57.5	14.7	56.6	15.8
Master's degree	37.0	7.7	37.0	7.3	37.6	7.8
Master's degree +	33.0	8.9	33.7	8.6	33.6	8.1

Note. AA = African American; W = White.

Table 14 Summary of Adjusted Residuals Between Praxis I Testing Mode and Undergraduate Institutional Selectivity by Race/Ethnicity Group for First-Time Test Takers

Selectivity	Reading		Writing		Mathematics	
	W	AA	W	AA	W	AA
Noncompetitive	7.9	4.5	8.1	4.5	7.3	3.7
Less competitive	-9.1	-9.8	-8.7	-9.6	-8.6	-9.8
Competitive	-12.7	3.5	-13.1	3.3	-12.0	3.2
Competitive +	-17.4	-2.2	-18.0	-3.1	-18.3	-2.0
Very competitive	2.9	-0.2	3.0	0.0	2.9	0.1
Very competitive +	30.8	6.4	31.0	6.6	30.2	6.7

Note. AA = African American; W = White.

were seniors were more likely than expected to take Praxis I on computer, whereas African American candidates who were seniors were more likely than expected to take Praxis I on paper.

Table 14 shows that those candidates whose undergraduate institutions were of the highest selectivity were more likely than expected to choose CB testing. Interestingly, this finding was also true for those from undergraduate institutions of the lowest selectivity. A key difference, though, occurred in the medium level of selectivity (competitive), where African American candidates chose to test on computer more than expected, whereas White candidates chose to test more on paper than expected.

Table 15 displays the Goodman and Kruskal tau statistics for those background variables with unordered categories. The results in Table 15 indicate that there is no association with mode of choice for Praxis I for both racial/ethnic groups based on teacher education program enrollment status or undergraduate major.

In summary, two BIQ variables, candidate educational attainment level and institutional selectivity, showed associations with testing mode, but this does not preclude the possibility that other variables not captured at registration could also be associated with mode of testing. That allows for qualitative explorations, which are discussed in the next section.

Table 15 Summary of Goodman and Kruskal Tau Statistics Showing Associative Strength Between Praxis I Testing Mode and Selected Background Variables by Race/Ethnicity Group for First-Time Praxis I Test Takers

Variable	Reading		Writing		Mathematics	
	W	AA	W	AA	W	AA
Teacher education program enrollment status	0.017	0.001	0.017	0.001	0.018	0.002
Undergraduate major	0.013	0.012	0.013	0.011	0.013	0.011

Note: AA = African American; W = White.

Table 16 Summary of Emergent Themes in Open-Ended Survey Question

Theme	Description of theme	Total responses (%)	PB responses (%)	CB responses (%)
Certification requirement	This test was taken as a requirement to get certified as a teacher	947 (34)	273 (34)	674 (34)
Cost, timing, location	Choice was made based on cost, timing, and/or location of the test	598 (22)	126 (16)	472 (24)
Computer better	Test takers preferred the computerized version	618 (22)	36 (4)	582 (30)
Paper better	Test takers preferred paper version	314 (11)	269 (33)	45 (2)
N/A	Answer could not be coded	215 (8)	76 (9)	139 (7)
No choice/no preference/ only option offered	Test takers did not have a preference on test format	43 (2)	21 (3)	21 (1)
Other		42 (1)	4 (< 1)	39 (2)
Total		2,777 (100)	805 (100)	1,972 (100)

Note. CB = computer-based; PB = paper-based

Attitudes and Opinions of Teacher Candidates About Choosing Testing Modes

A discussion was included at the beginning of this report about attitudes related to large-scale testing on computer (e.g., Bridgeman et al., 2001; Gallagher et al., 2002; Way et al., 2006), including the generally positive attitudes among test takers toward CB testing (Wang & Shin, 2009). As demonstrated earlier, differences were evident in test-taker volumes by mode, but only a few background variables appeared to be associated with test mode. The goal of this section is to discuss results from a qualitative pilot study among prospective teachers specifically taking a PB or CB Praxis I exam who were asked about reasons for choosing their particular mode of testing, addressing the final research question in this study. A total of 2,777 responses were received for the open-ended question from the 30,765 Praxis I test takers who were invited to take the survey (9.0%). Since both previous PB and CB Praxis I test takers were included in this survey, it is worth reporting that 805 of 8,220 PB test takers responded (9.8%) and 1,972 of 22,545 CB test takers responded (8.7%). However, because there were no other predefined criteria for recruitment and participation was voluntary, the sample was completely one of convenience. Any differences in representation based on available demographic factors were treated as a random. No analysis was possible to detect the influence of demographic factors among those not responding.

The single open-ended question, “What influenced your decision to take the computerized or paper-and-pencil version of the Praxis I exam?” served as a means to understand the decision-making processes in which test takers engaged when selecting a testing platform (paper or computer), because, as pointed out earlier, a test taker can only choose one mode of testing for any one administration. As noted earlier, 2,777 valid responses were provided for this question, from which thematic coding took place. Participant responses to the open-ended question generally produced shorter answers than expected, even abbreviated phrases. The achieved rate of exact agreement was 88% across the initial set of scored responses. Once independent coding was complete, the team met again to compare results for the four primary categories that emerged from the data and to make coding decisions on two coding categories: other (42 responses) and not applicable (215 responses).

The four primary themes that emerged from the data were (a) certification requirement; (b) cost, timing, and location; (c) computer is better; and (d) paper is better. The third and fourth were directly related to this research question, whereas the other two (themes a and b) related more to general reasons for taking Praxis I independent of this research question. Table 16 displays the themes in more detail, along with the incidence percentages among the responses. Approximately

Table 17 Summary of Emergent Subthemes for Those Preferring Computer Praxis I Testing

Category	Description	Total responses (%)	PB responses (%)	CB responses (%)
Faster scores	Test takers receive score right after exam	217 (35)	12 (33)	205 (35)
Comfort	Test takers more comfortable with computerized format	98 (16)	6 (17)	92 (16)
Personal preference		78 (13)	4 (11)	74 (13)
More efficient process	Efficient sign-up process and test-taking process	76 (12)	2 (6)	74 (13)
Easier	Easier to read, easier to focus, easier to understand processes and procedures	64 (10)	4 (11)	60 (10)
N/A	Answer could not be coded	85 (14)	8 (22)	77 (13)
Total		618 (100)	36 (100)	582 (100)

Note. CB = computer-based; PB = paper-based

Table 18 Summary of Emergent Subthemes for Those Preferring Paper Praxis I Testing

Category	Description	Total responses (%)	PB responses (%)	CB responses (%)
Better to visualize/manipulate work	Wants to mark up test, go back and forth between problems	100 (32)	81 (30)	19 (42)
Comfort	Test takers more comfortable with paper-and-pencil format	72 (23)	63 (23)	9 (20)
Personal preference		49 (16)	48 (18)	1 (2)
Easier	Easier to read, focus, understand processes and procedures	38 (12)	30 (11)	8 (18)
Familiarity with paper and pencil	Used to this testing format	15 (5)	14 (5)	1 (2)
N/A	Answer could not be coded	40 (13)	33 (12)	7 (16)
Total		314 (100)	269 (100)	45 (100)

Note. CB = computer-based; PB = paper-based

89% of responses clustered under the four primary themes. A slightly greater proportion of CB test takers (24%) based their decision of testing mode on cost, timing, and location. Among total responses, one-third of these were about whether CB or PB testing (22% and 11%, respectively) was preferred for Praxis I, yet within the subsamples of those taking the PB or CB test, the distinctions in mode of preference are clear. The relevance of these two categories led to secondary coding to determine further underlying subthemes.

Table 17 displays the summary of emergent subthemes among those responding that testing on computer was better. Of the 618 respondents in this category based on Table 16, 582 (94%) had taken a CB test. A total of 533 (86%) across modes provided a subtheme that could be coded. The results in Table 17 show that receiving scores right away was by far the most popular response. This was one of the positive aspects of computerized testing cited by Wang and Shin (2009). The fact that comfort was cited as the second most popular response could possibly be reflective of how people have grown accustomed to taking these kinds of tests on a computer.

Table 18 displays the corresponding results among those preferring to take Praxis I on paper. Of the 314 respondents in this category based on Table 16, 269 (86%) had taken a PB test. A total of 274 (87%) across modes provided a subtheme that could be coded. The most popular category of responses was related to visualization and manipulation. It is worth mentioning that even with a small number of those preferring PB testing after taking a CB test ($n = 45$), a greater proportion of those having taken a CB test (42%) said the PB test format provided a better way to visualize the task compared to those having taken a PB test (30%). However, a slightly greater proportion of those preferring PB testing (23%) cited comfort as a reason for their choice compared to those preferring CB testing (20%). When examining the distribution of themes by Praxis I test mode, about 86% of those taking PB tests said paper testing was better and 94% of those taking CB tests said computer testing was better. What is interesting, then, is that 14% of those taking PB tests said computer testing was better, compared to 6% of those taking CB tests who said paper testing was better, which represents a potential advantage for CB delivery. This finding was consistent with findings from Bridgeman et al. (2001).

Summary and Discussion

This report has summarized findings on various aspects of research related to the Praxis I series of initial teacher licensure examinations focusing on the CB mode of testing. The study focused on quantitative findings related to those teacher candidates who tested from November 2005 to November 2009, as in Nettles *et al.* (2011) and Tyler *et al.* (2011), who focused on PB testing, with the primary purposes of describing achievement gaps for those choosing CB testing and determining whether selected test-taker characteristics were associated with PB and CB testing. Given that the CB sample could not be treated as randomly equivalent to the PB sample, the observed differences in performance and characteristics could not be adequately explained by testing mode. Additional qualitative findings were obtained from a completely independent sample of previous Praxis I test takers about reasons why they chose PB or CB Praxis I testing.

For the first research question, exploring achievement gaps in CB testing, this study found that the average scale scores in Reading of those choosing CB tests compared to those choosing PB tests were significantly higher beyond a 0.20 effect size for all race/ethnicity groups, except Hispanic candidates. Asian American and Native American candidates choosing the computer-delivered Writing test performed significantly better than those choosing the paper-delivered test. African American and Asian American candidates choosing the computer-delivered Mathematics test performed significantly better than those choosing the paper-delivered test.

Most important, with the exception of Hispanic test takers, all achievement gaps between White test takers and those from other race/ethnicity groups were smaller for those choosing CB testing compared to PB testing. Additionally, a higher percentage of candidates across races/ethnicities choosing the CB tests met the minimum passing scores on all Praxis I tests compared to those choosing the PB tests. The lone exception was for Hispanic candidates on the Writing test. The corresponding gaps in passing rates were also lower for minority test takers compared to White test takers, except in the case of Hispanic candidates. Asian test takers slightly outperformed White test takers on the Mathematics test, but this difference was quite small.

The performance results were then broken down for White and African American candidates by key demographic variables: UGPA, enrollment in a teacher preparation program, candidate education level, parental education attainment as a proxy for SES, undergraduate field of study, and undergraduate institutional selectivity. Compared to the PB results stated in Nettles *et al.* (2011), a larger percentage of White candidates taking the CB tests was formerly enrolled in a teacher education program, whereas a smaller percentage was currently enrolled. The achievement gap was smallest for those currently enrolled in such programs. The biggest difference occurred for the CB Mathematics exam compared to the PB Mathematics exam, where the gap significantly decreased for those currently enrolled in a program and the gap widened on computer for those formerly enrolled.

Relative to Nettles *et al.* (2011), there were noteworthy findings in this study based on candidates' individual and parental educational levels. First, a shift occurred in the percentages such that both White and African American candidates were more often beyond the baccalaureate degree at time of testing. Yet the pattern in achievement gaps on computer was similar to that on paper. There was also a shift for candidate parental education attainment as a proxy for SES. Around 27% of White test takers' parents attained a graduate or professional degree, compared to about 20% of African American test takers' parents. These percentages were slightly larger compared to those teacher candidates who took the test on paper. Gaps were largest for all CB Praxis I exams at the highest SES level, consistent with Nettles *et al.* (2011).

When looking at undergraduate field of study, the performance gaps were smallest only on the Writing test across fields, whereas gaps were largest among humanities majors for the Reading and Mathematics tests. Just about half of White test takers (55–56%) and African American test takers (48–50%) majored in education, smaller percentages than those reported by Nettles *et al.* (2011) from PB testing. Education majors achieved the lowest mean scores on each of the Praxis I tests, but the gaps across computerized Praxis I tests, though still large, were smaller than for other majors.

Compared to findings reported in Nettles *et al.* (2011), only slight differences in the results were present when looking at undergraduate institutional selectivity. The representation gap in selectivity for CB test takers at schools considered “very competitive” or better was larger compared to PB test takers. The pattern for African American test takers on each of the three Praxis I tests was that with increased selectivity, the mean scores were higher, as was true for PB testing.

Regression models were developed to determine the predictive power of the selected background variables on CB Praxis I performance. Consistent with the PB findings in Nettles *et al.* (2011), race/ethnicity, UGPA, undergraduate major, and undergraduate institutional selectivity predicted Praxis I scale score performance. For the Writing test, parental educational attainment was an additional significant predictor. When examining differences in predictive power on

computerized Praxis I tests relative to those previously reported for paper tests, undergraduate major explained 4–5% more variance in CB scores than in PB scores. However, the overall amount of explained variance in each mode was relatively small (25–30%).

The second research question explored the same demographic variables used to examine achievement gaps to explore whether these characteristics were associated with testing mode. Given the large sample sizes and the construction of category descriptors, Cramer's V and Goodman and Kruskal's tau were used as measures of association between the demographic variable of interest and testing mode for White and African American test takers. The greatest association was found based on candidate education level. An examination of the adjusted residuals clearly demonstrated that undergraduate students tended to take Praxis I more on paper, whereas those with a bachelor's degree or higher tended to take Praxis I more on computer. Weak associations with undergraduate institutional selectivity were found. Those candidates attending undergraduate institutions of the highest selectivity were more likely to test on computer than on paper. Interestingly, this finding was also true for those attending undergraduate institutions of the lowest selectivity. A key difference, though, occurred in the medium level of selectivity (competitive), where African American candidates tested on computer more than expected and White candidates tested on paper more than expected. There were no significant associations of UGPA or SES with testing mode.

This research benefited from the development of a survey to qualitatively explore reasons why Praxis I test takers chose PB or CB testing, which was the subject of the third research question. Four primary themes emerged from the data: (a) certification requirement (34%); (b) cost, timing, and location (22%); (c) computer is better (22%); and (d) paper is better (11%). Secondary coding of those saying that testing on computer was better revealed that receiving scores right away was by far the most popular response, consistent with Wang and Shin (2009). When examining the distribution of themes by Praxis I test mode, about 86% of those taking PB tests said paper tests were better and 94% of those taking CB tests said computer tests were better. What is interesting, then, is that 14% of those taking PB tests said testing on computer was better, compared to 6% of those taking CB tests, who said testing on paper was better. The survey results indicated that cost, timing, and location more often influenced test mode for CB test takers (24%) than for PB test takers (16%).

Implications and Future Directions

All of the research presented in this report provided insight into the computer-delivered testing world for the Praxis program as it continues a rollout sequence for its licensure tests. This research aimed to raise awareness about the presence of achievement gaps in this mode of testing. This was especially important because there are no apparent current benchmarks for evaluating achievement gaps on CB tests, and though achievement gaps for those taking CB Praxis I tests were smaller compared to those taking PB tests, the magnitudes of the gaps are still large. Therefore, it would be potentially useful to revisit the original study by Camara and Schmidt (1999) examining achievement gaps across large-scale testing programs by race/ethnicity to better reflect fundamental changes in some of the exams referenced in their study based on more widespread CB testing across exams.

The descriptive statistics in Tyler et al. (2011) and regression model results in Nettles et al. (2011) showing those background variables that explain the most variation in Praxis I performance are important. However, given the consistent low proportion of explained variance in Praxis I scores from the present set of predictors, additional background variables should be considered, such as age, gender, alternate route program participation, and desire to teach in the same state as where the teacher preparation program is located. Nonetheless, the findings lead to another reason for raising awareness, that being the need to continue fostering strong support systems for students of color in promoting student success, particularly academic support (Adelman, 1999) and particularly for teacher candidates of color. This is important given that Gitomer, Brown, and Bonett (2008) have emphasized the importance of passing Praxis I with relation to passing future teacher licensure exams across states, namely, the *Praxis II*[®] series of tests. However, given that adoption of Praxis II exams varies greatly across states, a more localized analysis may be more appropriate.

Finally, some of the participation differences between CB testing in this report and PB testing in Nettles et al. (2011) and Tyler et al. (2011) raise the need for further exploration of the underlying infrastructures and available resources for educator preparation programs and their students in general, whether these are traditional college/university programs or alternate route programs. Specifically, the differences in participation between CB and PB testing by educational attainment level suggest that the quality or capacity of college- or university-based testing facilities needs attention given the CB sample had fewer undergraduate students than in the PB data Nettles et al. (2011) reported. Additionally, the association

between institutional selectivity and mode of testing showing greater participation by both White and African American CB test takers at the most selective institutions could serve as a call for greater financial support to less competitive institutions to perhaps improve facilities to better accommodate CB testing.

In closing, policy makers and researchers should be encouraged to leverage the research presented in this report in grappling with difficult ongoing policy issues, such as the overall quality and makeup of the U.S. teaching workforce. Although there is a great deal of breadth in the self-reported registration data collected for teacher candidates, the data do not represent all possible factors influencing choice of test mode or resulting performance. Therefore, further quantitative and qualitative research on choice of test mode and resulting performance is encouraged.

Acknowledgments

We would like to thank many people for their contributions to this report. We recognize the efforts of many of our current and former ETS colleagues: Linda Tyler, Kevin Cureton, Seth Weiner, Katherine Bassett, and Sharon Temmer from the Praxis business unit for their assistance and funding of this research; Kevin Larkin and Danielle Siwek for support from Statistical Analysis; Brent Bridgeman and Yigal Attali for reviewing the original proposal; Nat Kuczma for support from Market Research; Shelby Haberman, Rick Tannenbaum, Jaime Cid, and Marna Golub-Smith for editorial assistance; and Ayleen Stellhorn, Kim Fryer, and Holly Monteith for their assistance in copyediting this report.

Notes

- 1 Please refer to http://www.ets.org/s/praxis/pdf/pdt_registration_form.pdf for a list of current questions.
- 2 Exceptions are states using composite scoring methodologies where no minimums are necessary on any test as long as the composite score across all three tests is met.
- 3 The lower reliability for the Writing test is due to the fact that the estimation of the reliability is based on a weighted composite score of the multiple-choice and essay sections, and the essay generally has a lower reliability than the multiple-choice section.
- 4 In tables where percentages are reported, totals may not add to 100% due to rounding.
- 5 The inclusion of information about standard errors (i.e., the ratio of the standard deviation to the square root of the sample size) can also be useful in interpreting descriptive data. This information is available from the authors on request.
- 6 Please see Nettles et al. (2011, Appendix C) for more details on the use and interpretation of this statistic.
- 7 When interactions of race/ethnicity with another BIQ variable were significant and the main effect of the BIQ variable was not significant, the main effect was not reinserted into the final model, as is typically done in regression analysis. This was true in the case of candidate educational level for Writing.

References

- Adelman, C. (1999). *Answers in the toolbox: Academic intensity, attendance patterns, and bachelor's degree attainment*. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.
- Agresti, A. (1996). *An introduction to categorical data analysis*. New York, NY: John Wiley.
- Bridgeman, B., Lennon, M. L., & Jackenthal, A. (2001). *Effects of screen size, screen resolution, and display rate on computer-based test performance* (Research Report No. RR-01-23). Princeton, NJ: Educational Testing Service. <http://dx.doi.org/10.1002/j.2333-8504.2001.tb01865.x>
- Camara, W. J., & Schmidt, A. E. (1999). *Group differences in standardized testing and social stratification* (College Board Report No. 99-5). New York, NY: College Entrance Examination Board.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Cramer, H. (1946). *Mathematical methods of statistics*. Princeton, NJ: Princeton University Press.
- Educational Testing Service. (2010). *Praxis™ technical manual*. Princeton, NJ: Author.
- Gallagher, A., Bridgeman, B., & Cahalan, C. (2002). The effect of computer-based tests on racial-ethnic and gender groups. *Journal of Educational Measurement*, 39, 133–147.
- Gitomer, D. H., Brown, T., & Bonett, J. (2008, November). *Useful signal or unnecessary obstacle? The role of basic skills tests in teacher preparation*. Paper prepared for the Association of Public Policy Analysis and Management Conference, Los Angeles, CA.
- Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross-classification. *Journal of the American Statistical Association*, 49, 732–764.
- Nettles, M. T., Scatton, L. H., Steinberg, J. H., & Tyler, L. L. (2011). *Performance and passing rate differences of African American and White prospective teachers on Praxis™ examinations* (Research Report No. RR-11-08). Princeton, NJ: Educational Testing Service. <http://dx.doi.org/10.1002/j.2333-8504.2011.tb02244.x>

- Parshall, C. P., & Kromrey, J. D. (1993, April). *Computer testing versus paper-and-pencil: An analysis of examinee characteristics associated with mode effect*. Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.
- Puhan, G., Boughton, K. A., & Kim, S. (2005). *Evaluating the comparability of paper-and-pencil and computerized versions of a large-scale certification test* (Research Report No. RR-05-21). Princeton, NJ: Educational Testing Service. <http://dx.doi.org/10.1002/j.2333-8504.2005.tb01998.x>
- Rea, L. M., & Parker, R. A. (1992). *Designing and conducting survey research*. San Francisco, CA: Jossey-Boss.
- Reynolds, H. T. (1984). *Analysis of nominal data* (Vol. 7). Thousand Oaks, CA: Sage.
- Saldana, J. (2012). *The coding manual for qualitative researchers* (2nd ed.). Los Angeles, CA: Sage.
- Schaeffer, G. A., Bridgeman, B., Golub-Smith, M. L., Lewis, C., Potenza, M. T., & Steffen, M. (1998). *Comparability of paper-and-pencil and computer adaptive test scores on the GRE General Test* (Research Report No. RR-98-38). Princeton, NJ: Educational Testing Service. <http://dx.doi.org/10.1002/j.2333-8504.1998.tb01787.x>
- Schaeffer, G. A., Steffen, M., Golub-Smith, M. L., Mills, C. N., & Durso, R. (1995). *The introduction and comparability of the computer adaptive GRE General Test* (Research Report No. RR-95-20). Princeton, NJ: Educational Testing Service. <http://dx.doi.org/10.1002/j.2333-8504.1995.tb01655.x>
- Taylor, C., Jamieson, J., Eignor, D., & Kirsch, I. (1998). *The relationship between computer familiarity and performance on computer-based TOEFL test tasks* (TOEFL Research Report No. 61). Princeton, NJ: Educational Testing Service. <http://dx.doi.org/10.1002/j.2333-8504.1998.tb01757.x>
- Tyler, L., Whiting, B., Ferguson, S., Eubanks, S., Steinberg, J., Scatton, L., & Bassett, K. (2011). *Toward increasing teacher diversity: Targeting support and intervention for teacher licensure candidates* (Research Report No. ETS-NEA-2011-01). Princeton, NJ: Educational Testing Service.
- Wang, H., & Shin, C. D. (2009). Computer-based and paper-pencil test comparability studies. *Test, Measurement & Research Services Bulletin*, 9, 1–6.
- Way, W. D., Davis, L. L., & Fitzpatrick, S. (2006, April). *Score comparability of online and paper administrations of Texas assessment of knowledge and skills*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.

Appendix A Statistical Formulas

Cramer's V (Cramer, 1946)

This is calculated as the square root of the chi square statistic (χ^2) divided by the product of the total sample size (N) and 1 less than the minimum of the number of rows and columns in the contingency table (k):

$$\varphi_c = \sqrt{\chi^2 / (N * (k - 1))}. \quad (\text{A1})$$

Adjusted Residuals (Agresti, 1996)

This formula standardizes the difference between observed and expected counts in each cell of a contingency table (n_{ij} and $\hat{\mu}_{ij}$, respectively) by dividing by the square root of the product of the expected cell count, the difference in the proportion of observations within that row (i) from 1, and the difference in the proportion of observations within that column (j) from 1:

$$\frac{n_{ij} - \hat{\mu}_{ij}}{\sqrt{\hat{\mu}_{ij} (1 - p_{i+}) (1 - p_{+j})}}. \quad (\text{A2})$$

Goodman and Kruskal's Tau (Goodman & Kruskal, 1954; Reynolds, 1984)

This formula computes the marginal reduction in error by predicting observed frequencies (O_{ij}) for each cell in the rows (R_i) and columns (C_j) of the contingency table where n is the total number of cases in the table:

$$\tau = \left(\sum_i ((n - R_i) / n) * R_i \right) / \left(\sum_i \left(\sum_j \left((C_j - O_{ij}) / C_j \right) * O_{ij} \right) \right). \quad (\text{A3})$$

Appendix B
Summary of Regression Analyses Using Computerized Praxis I Tests and Background Information Questionnaire (BIQ) Variables

Table B1 Summary of Full Regression Results for Computerized Praxis I Reading

Step	Predictor	Stepwise selection		Parameter estimates			
		r^2	Total r^2	β	SE	Sig.	Std. β
1	Race/ethnicity	0.14	0.14	-3.11	0.29	<.01	-0.19
2	Major	0.09	0.23	-2.84	0.07	<.01	-0.22
3	UGPA	0.03	0.26	2.64	0.08	<.01	0.19
4	Selectivity	0.02	0.29	1.83	0.07	<.01	0.14
5	Candidate educational level	<.01	0.30	1.70	0.09	<.01	0.11
6	Parental educational attainment	<.01	0.30	0.91	0.07	<.01	0.07
7	Race/ethnicity \times Candidate educational level	<.01	0.31	-2.02	0.24	<.01	-0.12
8	Race/ethnicity \times Major	<.01	0.31	-0.98	0.17	<.01	-0.05
9	Race/ethnicity \times Selectivity	<.01	0.31	0.90	0.18	<.01	0.03
10	Teacher education program enrollment	<.01	0.31	-0.33	0.07	<.01	-0.03
11	Race/ethnicity \times Parental educational attainment	<.01	0.31	0.62	0.16	<.01	0.03
12 ^a	Race/ethnicity \times UGPA	<.01	0.31	-0.55	0.16	<.01	-0.03

Note. UGPA = undergraduate grade point average.

^{aa}The interaction between race/ethnicity and teacher education program enrollment was not significant and was dropped from the model at Step 13.

Table B2 Summary of Full Regression Results for Computerized Praxis I Writing

Step	Predictor	Stepwise selection		Parameter estimates			
		r^2	Total r^2	β	SE	Sig.	Std. β
1	Race/ethnicity	0.11	0.11	-2.14	0.18	<.01	-0.17
2	Major	0.06	0.17	-1.95	0.05	<.01	-0.20
3	UGPA	0.05	0.22	2.43	0.06	<.01	0.23
4	Selectivity	0.03	0.25	1.73	0.05	<.01	0.18
5	Parental educational attainment	0.01	0.26	1.06	0.05	<.01	0.11
6	Race/ethnicity \times UGPA	<.01	0.27	-0.84	0.13	<.01	-0.05
7	Candidate educational level	<.01	0.27	0.27	0.07	<.01	0.02
8 ^a	Race/ethnicity \times Candidate educational level	<.01	0.27	-0.62	0.18	<.01	-0.05

Note. UGPA = undergraduate grade point average.

^{aa}Teacher education program enrollment and the interactions of ethnicity with enrollment in a teacher education program, parental educational attainment, major, and selectivity respectively were not significant and were dropped from the model at Step 9.

Table B3 Summary of Full Regression Results for Computerized Praxis I Mathematics

Step	Predictor	Stepwise selection		Parameter estimates			
		r^2	Total r^2	β	SE	Sig.	Std. β
1	Race/ethnicity	0.14	0.14	-3.06	0.31	<.01	-0.17
2	Major	0.06	0.20	-2.84	0.08	<.01	-0.20
3	Selectivity	0.03	0.23	2.42	0.08	<.01	0.16
4	UGPA	0.02	0.25	2.42	0.09	<.01	0.15
5	Parental educational attainment	0.01	0.26	1.31	0.07	<.01	0.09
6	Race/ethnicity \times Candidate educational level	<.01	0.26	-2.09	0.27	<.01	-0.11
7	Race/ethnicity \times UGPA	<.01	0.26	-1.14	0.19	<.01	-0.05
8	Race/ethnicity \times Major	<.01	0.26	-0.90	0.19	<.01	-0.04
9 ^a	Teacher education program enrollment	<.01	0.26	-0.30	0.08	<.01	-0.02

Note. UGPA = undergraduate grade point average.

^aThe interactions between race/ethnicity and parental educational attainment, selectivity, and teacher education program enrollment were not significant and were dropped from the model at Step 10.

Suggested citation:

Steinberg, J., Brenneman, M., Castellano, K., Lin, P., & Miller, S. (2014). *A comparison of achievement gaps and test-taker characteristics on computer-delivered and paper-delivered Praxis I® tests* (ETS Research Report No. RR-14-35). Princeton, NJ: Educational Testing Service. doi:10.1002/ets2.12033

Action Editor: Marna Golub-Smith

Reviewers: Richard Tannenbaum, Jaime Cid, Shelby Haberman, and Kevin Larkin

ETS, the ETS logo, GRE, LISTENING. LEARNING. LEADING., PRAXIS, PRAXIS I, PRAXIS II, and TOEFL are registered trademarks of Educational Testing Service (ETS). PSAT and SAT are registered trademarks of the College Board. All other trademarks are property of their respective owners.

Find other ETS-published reports by searching the ETS ReSEARCHER database at <http://search.ets.org/researcher/>