Research Report

Investigating Differences in Examinee Performance Between Computer-based and Handwritten Essays

Lei Yu

Samuel A. Livingston

Kevin C. Larkin

John Bonett

Research & Development

May 2004
RR-04-18

Investigating Differences in Examinee Performance Between Computer-based and Handwritten Essays

Lei Yu, Samuel A. Livingston, Kevin C. Larkin, and John Bonett

ETS, Princeton, NJ

May 2004

Abstract

This study compared essay scores from paper-based and computer-based versions of a writing test for prospective teachers. Scores for essays in the paper-based version averaged nearly half a standard deviation higher than those in the computer-based version, after applying a statistical control for demographic differences between the groups of examinees taking the two versions. The statistical control was implemented by means of a propensity score, applying weights to members of one group to match the propensity score distribution of the other group. The score difference between the groups did not change substantially when the analysis was restricted to examinees taking the same mix of essay topics or to examinees taking one particular essay.

Key words: Essay testing, score equivalence, computer-based testing

Acknowledgements

The authors thank Brent Bridgeman, Peter Cooper, Daniel Eignor, and David Anderson for their helpful comments and suggestions on an earlier draft. Any opinions expressed in this report are those of the authors and not necessarily of ETS. ETS Research and Development funded this research through the allocation fund.

Table of Contents

List of Tables

List of Figures

Page

Introduction

Purpose of the Study

The present study was designed to assess the equivalence of scores on the computerized and paper-delivered versions of the essay portion of a writing test. Some previous research (e.g., Powers, Fowles, Farnum, & Ramsey, 1992; Bridgeman & Cooper, 1998) has found that handwritten essays tend to be graded more leniently than essays with identical content in typewritten form. However, that effect may be countered by the greater ease, for many examinees, of composing their essays at the computer. This study was intended to determine the full effect (including both administration and scoring) of the mode of testing (paper or computer) on the examinees' essay scores.

The Tests

The Praxis™ Pre-Professional Skills Test is a battery of three tests that assess basic academic skills—in reading, mathematics, and writing—for individuals seeking to enter teacher training programs. Examinees may choose between a paper-and-pencil form of the test and a computerized alternative, introduced in January 2002. In this paper, the paper-and-pencil version will be referred to as the PPST®. The computerized version will be referred to as the CPPST. The CPPST is intended to be comparable to the PPST in terms of test length, difficulty, speededness, and scoring procedures. Both versions report scores on the same measurement scale, and scores on the two versions are interchangeable. Therefore, it is important that scores on the two versions be comparable.

Scaled scores on each test (reading, mathematics, writing) of the PPST and CPPST range from a minimum of 150 to a maximum of 190. Passing scores on these tests are not set by ETS, but by the states that have adopted the test for use. Thirty-two states currently use the PPST Writing exam. Most of those states accept scores on either the PPST or the CPPST. Their passing scores range from 170 to 176.

In both the PPST and the CPPST, the writing test assesses the examinee's ability to use correctly standard English language and grammar and to write effectively for communication. The PPST Writing test includes 45 multiple-choice items and a single essay. The multiple-choice and essay sections are timed separately. Thirty minutes are allowed for each section. The two sections contribute equally to the examinee's total raw score. The CPPST also includes 45

multiple-choice items (plus six unscored pretest items) and a single essay. Examinees are allowed 45 minutes for the multiple-choice section and 30 minutes for the essay section. Examinees taking the PPST write their essays by hand; examinees taking the CPPST type their essays into the computer.

The present study was designed to assess the equivalence between scores on the CPPST essay and scores on the PPST essay. It focuses on the performance of examinees taking these tests from January through September of 2002 (the first 9 months of testing with the CPPST).

The essay topics in the PPST and CPPST present situations or issues familiar to educated people. No topic requires specialized knowledge. Examinees are expected to write only on the topic assigned, to respond to all the points included in the topic, and to use specific examples drawn from their own knowledge or personal experience to support their arguments (Educational Testing Service, 2001).

Between January and September of 2002, 16 different forms of the PPST were administered, each form containing a different multiple-choice portion and a different essay topic. During this same period, a large number of different forms of the CPPST, containing many different essay topics, were administered. (Thirteen of those 37 essay topics were among the 16 topics used in the PPST.)

On both the PPST and the CPPST, the multiple-choice section is scored by awarding one point for each correct response. There is no penalty for incorrect responses. On both tests, the examinee's essay is read and evaluated holistically on a scale of 0 to 6 points by two raters, working independently. The examinee's raw essay score is the sum of the two ratings. To form a composite that gives the two sections of the exam equal weight, the essay score is multiplied by 3.75, so that the weighted essay score will have a maximum value of 45 points. The sum of the multiple-choice number-correct score and the weighted essay score is called the "raw weighted composite" (RWC).

The PPST essays and the CPPST essays were scored by two separate, nonoverlapping groups of raters, working under different conditions. The PPST essays were scored by raters who convened at ETS immediately following each test administration. The CPPST essays were rated by raters from an online essay-scoring service provided by another company. The professional qualifications required to become a rater were the same in the two groups.

Prior to the year 2000, PPST scores were equated through common items; each new form of the test incorporated a set of questions repeated from a previous form. Since then, PPST scores have been equated by "spiraling" (i.e., alternating) each new form of the test with a previously administered form, among the examinees at each test center, to produce highly similar groups of examinees taking the new form and the previous form.

In the development of the CPPST, an item response theory (IRT) pre-equating was used to derive a separate raw-to-scale conversion for the multiple-choice portion of each form. All the items available for use in any of the forms were calibrated together, using the three-parameter logistic response model. The estimated item response curves for the individual items were used to determine the test characteristic curve for each computer-based form and for a specified paper-based reference form. (The test characteristic curve estimates the expected number of items correct as a function of the examinee's ability.) The test characteristic curves were then used to translate each possible score on each computer-based form to a corresponding score on the reference form. This score was the equated multiple-choice score.

An examinee's scores on the multiple-choice and essay portions of a computer-based form of the test were converted to a scaled score by the following procedure:

1. Determine the examinee's equated multiple-choice score.
2. Multiply the examinee's raw score on the essay by the specified weight, and add it to the equated multiple-choice score, to form a raw weighted composite (RWC).
3. Transform the RWC to a scaled score by applying the raw-to-scale conversion for the paper-based reference form.

Three analyses were conducted in the study. The first analysis included all the examinees, with no control for possible differences in the difficulty of the different essay topics presented to the PPST examinees and the CPPST examinees. The second and third analyses were intended to reduce or remove the effect of differences in the difficulty of the topics. The second analysis was restricted to examinees who wrote in response to essay topics that were included in both the PPST and CPPST. The third analysis compared examinees who wrote in response to one specific essay topic.

Previous Research

Previous research has found that handwritten essays tend to be graded more leniently than essays with identical content in typewritten form. Arnold et al. (1990) compared 300

handwritten essays with word-processed versions of the same essays and found that the word-processed versions were consistently scored lower than the corresponding original handwritten ones. Raters were also found to show leniency when poor handwriting was encountered.

Powers et al. (1992) selected a representative sample of 32 out of 568 students who had taken the Praxis I: Academic Skills Assessment. The investigators had the students' essays converted to word-processed form, if originally handwritten, or to handwritten form, if originally word-processed, and scored in both versions. They found that the average score of the handwritten essays was higher than that of the word-processed essays, whether the essays were originally handwritten or word-processed.

In a second investigation of Praxis essays, Powers and Farnum (1997) observed the same tendency of handwritten essays to receive higher ratings than word-processed essays. The results did not depend on whether the essays were read on a computer screen or on paper.

Bridgeman and Cooper (1998) conducted a study of 3,470 examinees who took essays on the Graduate Management Admission Test® in both handwritten and word-processed formats. The handwritten essays received higher scores than the word-processed essays and this difference was not related to gender, ethnicity, or having learned English as a second language.

In a study of essays students wrote as the final exam for a writing course, Sweedler-Brown (1991) randomly selected 61 out of 700 handwritten essays, which were then typed verbatim on word processors. The better essays tended to receive higher scores in handwritten format than in typed format, but the poorer essays did not.

Two studies have compared the quality of essays written by hand on paper with essays composed at a computer keyboard, although the participants in those studies were quite different from the participants in the present study. Russell and Haney (1997) randomly assigned middle school students to write essays by hand or at a computer. The handwritten essays were input into the computer before being scored. The essays composed on the computer received scores an average of 0.9 SD higher than the handwritten essays. In a follow-up study, Russell (1999) found that this effect depended both on the subject tested (language arts, science, or math) and on the students' word processing ability.

Method

The present study is based on a comparison of the essay scores of the examinees taking the CPPST writing test from January through September of 2002 to the essay scores of the

examinees taking the PPST writing test during the same period. A simple comparison of the essay scores of the PPST and CPPST examinee groups would not be a sound basis for conclusions about the effect of the test format, because the groups are self-selected and differ on many variables. This study used a poststratification procedure based on demographic variables to create statistically matched samples of PPST and CPPST examinees.

Because many different demographic variables were available, it was not feasible to stratify on all possible combinations of these variables. Instead, the stratification was based on a *propensity score* (Rosenbaum & Rubin, 1985; Rubin & Thomas, 2000). The propensity score is the linear combination of the stratifying variables that best discriminates between the two groups—in this case, the examinees taking the PPST and those taking the CPPST. The procedure automatically assigns the greatest weight to the variables on which the groups are most different and the smallest weight to the variables on which the groups are most similar. Poststratification on the propensity score was implemented by dividing the propensity score range into 20 intervals. Each CPPST examinee was assigned a weight that depended on the examinee's propensity score interval; the weight was equal to the ratio of PPST examinees to CPPST examinees in that interval. Thus, the weighted sample of CPPST examinees closely resembled the group of PPST examinees in the distribution of the propensity score. Because the propensity score emphasizes the demographic variables on which the CPPST group differed most from the PPST group, the weighted CPPST sample was demographically similar to the PPST group.

Procedures

The demographic variables included in the analysis were empirically selected from the items in the background questionnaire by examining the distributions of the responses to each item in the two groups of examinees (the PPST group and the CPPST group). The items on which the two groups differed substantially were used as the independent variables in a logistic regression; the dependent variable was group membership (PPST or CPPST). The logistic regression model is

$$\log \frac{P}{1-P} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_n x_n$$

where P represents the probability that an examinee with a given set of demographic characteristics $(x_1, x_2, \ldots x_n)$ will take the PPST rather than the CPPST. The propensity score is

computed using

$$Y = b_1x_1 + b_2x_2 + b_3x_3 + \ldots + b_nx_n$$

where x_1, x_2, ... represent the examinee's values of the selected demographic variables and b_1, b_2, ... are the estimates of β_1, β_2, ... in the logistic regression model.

The range of the obtained propensity scores was then divided into 20 intervals, and the number of examinees in each group in each interval was determined. In each interval, the ratio of the number of PPST examinees to the number of CPPST examinees was computed. This ratio was used as the weight to be applied to each examinee in the CPPST group with a propensity score in that interval, in computing a distribution of essay scores. This weighted distribution of CPPST essay scores was then compared with the observed distribution of PPST essay scores.

Three such analyses were conducted. In all three analyses, the CPPST examinees were weighted to match the demographics of the PPST examinees. The first analysis included all the examinees taking each test (PPST or CPPST). This analysis treated all the different essay topics as equivalent in difficulty (as is done in the operational scoring of the CPPST). The second analysis was restricted to examinees taking the 13 essay topics that had been administered in both the PPST and the CPPST. The third analysis compared the scores of examinees who wrote on a single essay topic—the topic for which the numbers of PPST and CPPST examinees differed the least.

Results

Characteristics of the Examinee Groups

The numbers of examinees included in the study were 51,466 for PPST, and 42,912 for CPPST. The examinees' demographic characteristics, educational background, job-related information, and teaching-related experience are presented in Tables 1 to 4. The questionnaire is provided in the appendix. Examinees more likely to take the PPST tended to be (a) full-time students—freshmen and sophomores in particular, (b) students majoring in elementary and pre-elementary education, (c) those with a GPA of 3.5 or above, and (d) those who planned to enroll in a teaching program. Examinees more likely to take the CPPST tended to be (a) seniors or college graduates, (b) those majoring in social science, (c) those with a GPA between 2.50 and 2.99, and (d) those who had a full-time teaching job or some teaching experience.

Table 1

Gender and Ethnic Composition of the Examinee Groups

	PPST		CPPST	
	N	%	*N*	%
Gender				
Males	12,126	24	11,473	27
Females	39,302	76	31,358	73
No response	38	0	81	0
Total	51,466	100	42,912	100
Ethnicity				
Black	5,084	10	5,968	14
White	42,121	82	33,726	79
Other	4,261	8	3,218	7
Total	51,466	100	42,912	100

Note. The column totals in this table and the following tables may not add to exactly 100% because of rounding.

Table 2

Examinees' Educational Background

	PPST	CPPST
Educational level		
Freshman	14%	6%
Sophomore	23%	16%
Junior	16%	16%
Senior or higher	43%	62%
No response	3%	0%
Undergraduate major		
Education subject areas	19%	27%
Elementary and Pre-Elem. Education	40%	33%
Humanities	6%	8%
Math and Natural Sciences	6%	7%
Social Sciences	10%	15%
Special Education	5%	4%
Others	14%	7%
Undergraduate GPA		
3.5 – 4.0	29%	23%
3.0 – 3.49	38%	38%
2.5 – 2.99	27%	32%
2.49 and below	6%	7%

Table 3

Examinees' Job-related Information

	PPST	CPPST
Number of years since attending college or graduate school		
Currently attending	75%	63%
Less than 1 year	6%	9%
1–3 years	7%	12%
4–6 years	4%	6%
7–10 years	2%	3%
More than 10 years	5%	7%
No response	2%	0%
Most recent full-time occupation		
Student	51%	37%
Teacher	13%	18%
Professional/Executive	4%	7%
School aide	4%	5%
Clerical/Administrative support	3%	5%
Sales/retail	3%	5%
Managerial	2%	3%
Others	20%	20%

Table 4

Examinees' Teaching-related Experience

	PPST	CPPST
Ever enrolled in a teaching program		
Currently	52%	45%
Formerly	11%	14%
Never	35%	41%
No response	3%	0%
Teaching status		
Plan to enroll or currently enrolled in a teaching program	81%	77%
Recently graduated and expected to begin teaching in the near future	4%	5%
1–3 years teaching experience	7%	10%
More than 3 years teaching experience	5%	6%
Not planning to teach at this time	3%	2%

Table 5 shows the variables selected as independent variables for the logistic regression. The independent contribution of each variable to the regression was statistically significant ($Pr < 0.001$). Therefore, all of these variables were used in the calculation of the propensity scores.

Table 5

Categories/Variables Selected for the Logistic Regression and Their Parameter Estimates

Category	Variables
Undergraduate major	Education subject areas Elementary and Pre-Elementary Education Social Sciences
Certification field	Education subject areas Elementary and Pre-Elementary Education
Ethnicity	African American or Black
Educational level	Freshman Sophomore Junior Senior or above
Years since attending school	Currently attending
Undergraduate GPA	3.5 – 4.0
Most recent full-time occupation	Student
Teaching status	Planning to enroll or currently enrolled in a teacher education program

Comparison Based on All Examinees

Summary statistics of the propensity scores of the PPST group, the CPPST group, and the weighted CPPST sample are shown in Table 6. Figure 1 shows the propensity score distributions graphically. There is a substantial difference between the propensity scores of the PPST group and the CPPST group; weighting the examinees in the CPPST group removes this difference.

Table 6

Propensity Score Statistics of PPST Examinees, CPPST Examinees, and Weighted CPPST Sample

	PPST examinees	CPPST examinees	Weighted CPPST sample
Number of examinees	51,466	42,912	51,466
Mean	–1.81	–2.17	–1.81
Standard deviation	.64	0.56	.64

The distribution of the essay scores of the PPST group, the CPPST group, and the weighted CPPST sample are shown in Table 7 and Figure 2. Because the essay score is the sum of ratings by two independent raters, odd-numbered scores are unusual (even more so on the CPPST than on the PPST). There is a substantial difference between the essay scores of the PPST group and the CPPST group; the mean score of the PPST group is higher by half a standard deviation. Scores of 9 or higher are more common on the PPST, while scores of 6 or lower are more common on the CPPST. When the CPPST group is weighted to produce a sample demographically similar to the PPST group, the distribution of essay scores changes very little, becoming slightly more like the distribution in the PPST group.

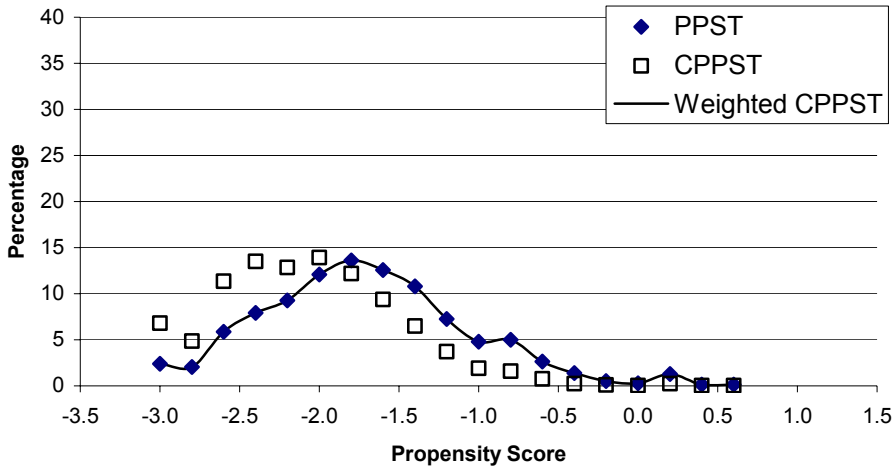

Figure 1. **Distributions of propensity scores in the PPST and CPPST groups.**

Table 7

Essay Score Statistics of PPST Examinees, CPPST Examinees, and Weighted CPPST Sample

	PPST examinees	CPPST examinees	Weighted CPPST sample
Number of examinees	51,466	42,912	51,466
Mean	8.11	7.36	7.46
Standard deviation	1.45	1.53	1.50
Percentage with scores of			
12	1	1	1
11	2	1	1
10	17	7	8
9	9	3	
8	48	52	53
7	7	1	1
6	13	30	28
5	1	2	1
4	1	4	3
3	0	0	0
2	0	1	0
0	0	0	0

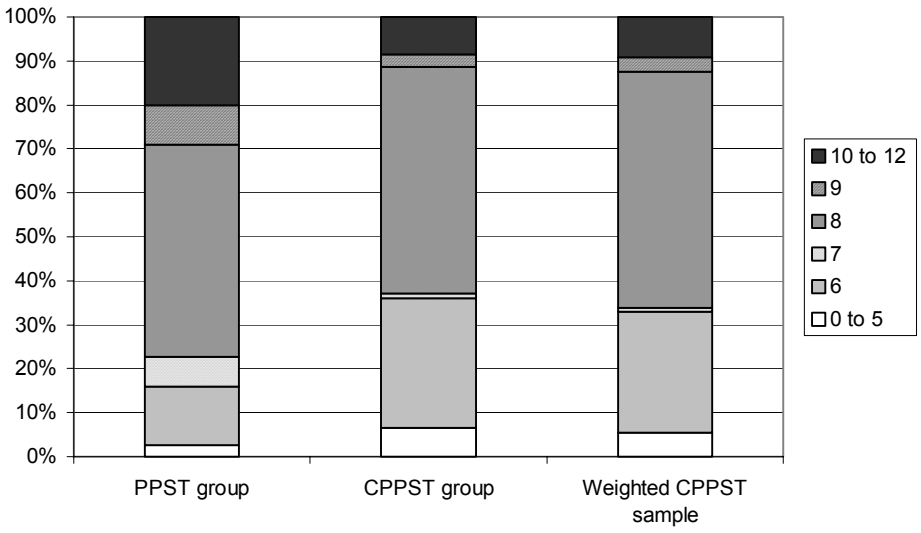

Figure 2. **Essay score distributions of the PPST group, CPPST group, and the weighted CPPST sample.**

Comparison Based on Examinees Taking Essay Topics Included in Both Tests

The second analysis was restricted to examinees who took any of the 13 topics included in both tests. This restriction reduces the PPST sample only slightly, but the CPPST sample is reduced by 65%, from more than 42,000 examinees to fewer than 15,000. Summary statistics of the propensity scores of the PPST group, the CPPST examinees who took PPST essays, and the weighted CPPST sample are shown in Table 8. Figure 3 shows the propensity score distributions graphically. The distributions are similar to those in Figure 1, indicating a systematic difference between the demographic characteristics of the groups. Table 8 also shows the percentage of each group taking each of the 13 essay topics that were included in both the PPST and the CPPST. The percentages of the examinees taking these 13 topics are very similar for the CPPST group but vary for the PPST group.

Table 8

Propensity Score Statistics and Essay Topics of PPST Examinees, CPPST Examinees Taking Essays Included in PPST, and Weighted CPPST Sample

	PPST examinees	CPPST examinees taking PPST essays	Weighted CPPST sample
Number of examinees	51,150	14,879	51,150
Propensity score mean	–1.71	–2.08	–1.71
Propensity score SD	.64	.57	.64
Percentage taking			
Essay topic 1	8	8	8
Essay topic 2	14	8	8
Essay topic 3	15	8	8
Essay topic 4	10	8	8
Essay topic 5	7	5	5
Essay topic 6	6	7	8
Essay topic 7	6	8	7
Essay topic 8	6	8	8
Essay topic 9	8	8	8
Essay topic 10	8	8	8
Essay topic 11	3	8	8
Essay topic 12	3	8	8
Essay topic 13	7	8	8

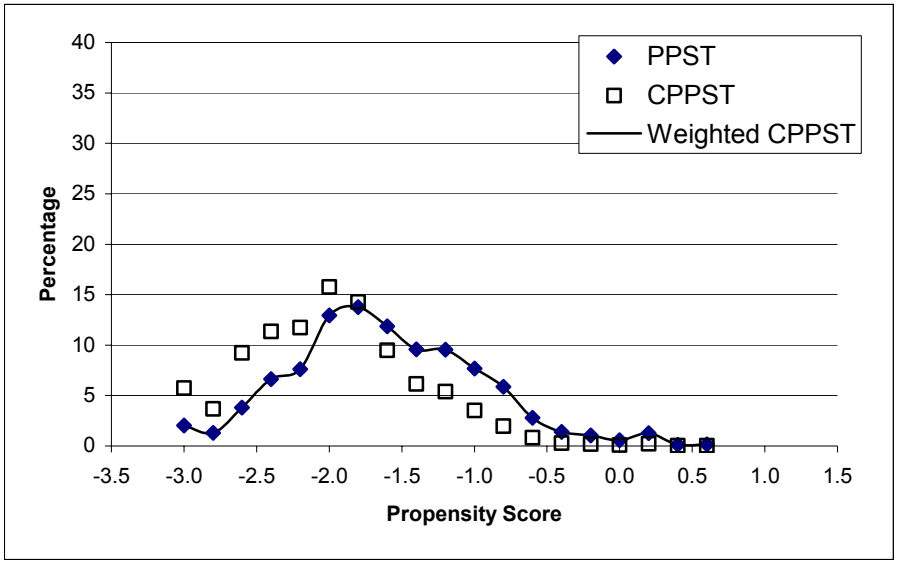

Figure 3. **Propensity score distributions of examinees taking the same mix of essays.**

The essay score distributions in the PPST group, the CPPST examinees taking PPST essays, and the weighted PPST sample are shown in Table 9 and Figure 4. The score distribution for the PPST group is the same as in Table 7 and Figure 4, because the group of examinees is the same. However, the score distributions for the CPPST group and the weighted CPPST sample are also very similar to those in Table 7 and Figure 4, even though nearly two thirds of the CPPST examinees have been removed from the analysis. Restricting the analysis to examinees who took the 13 essay topics included in both tests seems to make almost no difference in the results. The essay scores of the CPPST examinees are lower than those of the PPST examinees by half a standard deviation. Applying weights to the CPPST examinees to make their group demographically similar to the PPST group makes only a small change in the distribution of their essay scores.

Table 9

Essay Score Statistics of PPST Examinees, CPPST Examinees Taking PPST Essay Topics, and Weighted CPPST Sample

	PPST examinees	CPPST examinees taking PPST essay topics	Weighted CPPST sample
Number of examinees	51,150	14,879	51,150
Mean	8.11	7.35	7.45
Standard deviation	1.45	1.54	1.50
Percentage with scores of ...			
12	1	1	1
11	2	1	1
10	17	7	8
9	9	3	3
8	48	51	53
7	7	1	1
6	13	30	28
5	1	2	2
4	1	4	3
3	0	0	0
2	0	1	0
0	0	0	0

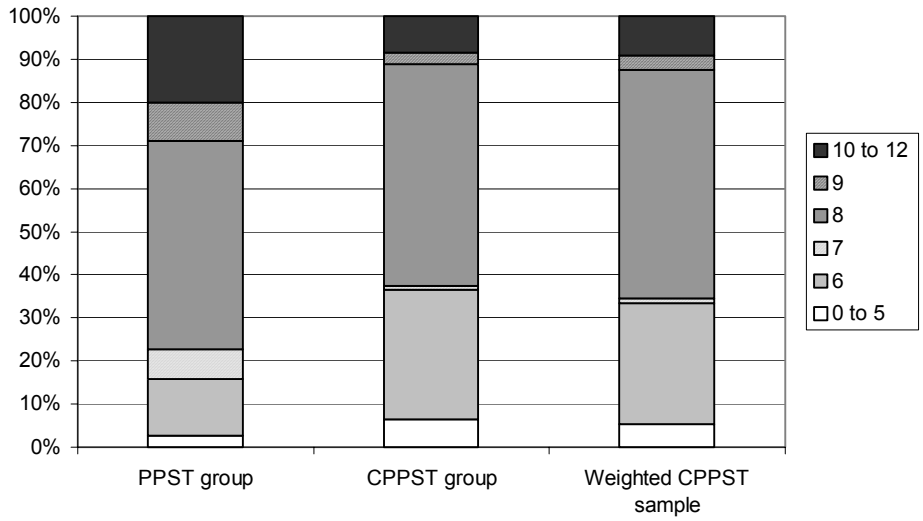

Figure 4. **Essay score distributions of CPPST examinees who took essays included in PPST and of weighted CPPST sample.**

Comparison Based on Examinees Taking a Single Essay Topic

The final comparison was restricted to examinees who wrote on one particular essay topic—the one with the smallest difference between the numbers of examinees in the PPST group (1,645) and the CPPST group (1,214). Table 10 and Figure 5 show the comparison of the propensity scores of the two groups of examinees. (The range of the propensity scores was somewhat smaller than in the previous analyses.)

Table 10

Propensity Score Statistics and Essay Topics of CPPST Examinees and PPST Examinees Taking a Single Essay Topic and Weighted CPPST Sample

	PPST examinees	CPPST examinees	Weighted CPPST sample
Number of examinees	1,645	1,214	1,645
Propensity score mean	–.41	–.64	–.41
Propensity score SD	.45	.50	.45

Table 11 and Figure 6 show the comparison of the essay scores of the PPST group, the CPPST group, and the weighted CPPST sample for the examinees taking this one essay topic. The results are essentially the same as those for the unrestricted groups of examinees. Again, the essay scores of the CPPST examinees are lower than those of the PPST examinees by about half a standard deviation. And again, applying weights to the CPPST examinees to make their group demographically similar to the PPST group makes only a small change in the distribution of their essay scores.

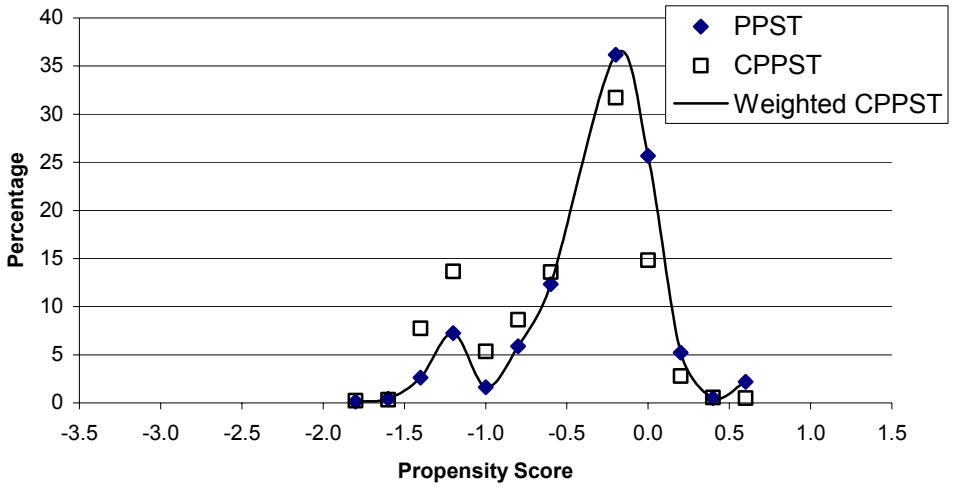

Figure 5. **Propensity score distributions of PPST and CPPST examinees taking the same essay.**

Table 11

Essay Score Statistics of PPST and CPPST Examinees Taking One Particular Essay Topic

	PPST examinees	CPPST examinees	Weighted CPPST sample
Number of examinees	1,645	1,214	1,645
Mean	8.18	7.39	7.43
Standard deviation	1.48	1.47	1.46
Percentage with scores of			
12	3	1	1
11	3	0	0
10	14	7	7
9	9	3	3
8	50	54	54
7	7	1	1
6	13	29	28
5	1	1	1
4	1	4	4
3	0	0	0
2	0	0	0

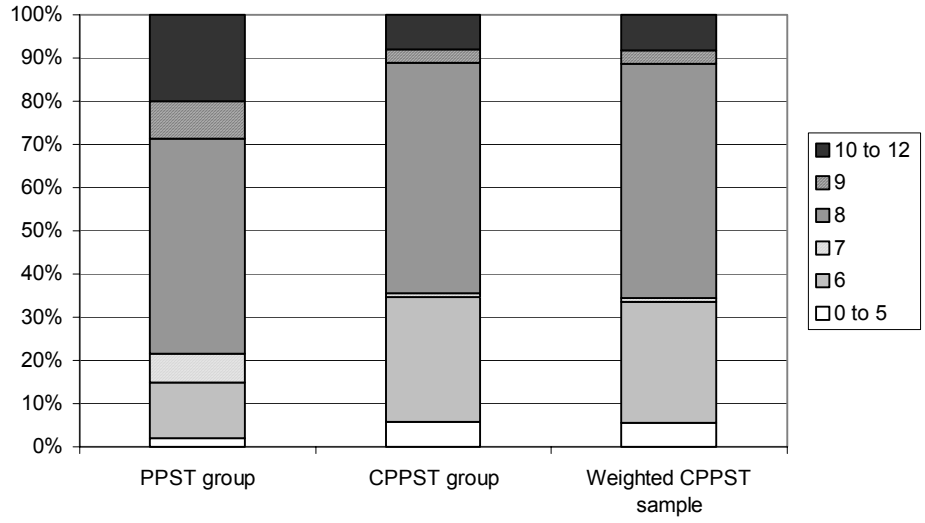

Figure 6. **Essay score distributions of the PPST group and the weighted CPPST group who wrote on the same essay.**

Discussion

The results of the study show that PPST examinees tended to receive higher essay scores than CPPST examinees, even after controlling for the available demographic variables and for the specific essay topics on which the examinees were tested. The difference in the essay score associated with taking the test on paper vs. computer was approximately three fourths of a point on the 2 to 12 essay score scale. That difference in the essay scores translates to a difference of approximately 1.25 points in the reported scaled scores. In the region of the score scale where most state-qualifying scores lie, an interval of one scaled-score point would include from 5% to 10% of the examinees.

There are at least four possible explanations for the results of this study:

1. Among examinees with similar demographic profiles, those who choose to take the PPST rather than the CPPST may have better writing skills.
2. The examinees may produce better essays when they write with a pencil or pen than when they type their essays directly into a computer.
3. Raters may be generally more lenient when rating handwritten essays than typed essays.

4. The particular raters who rate CPPST essays may be more lenient than those who rate PPST essays.

Previous research provides evidence to support the third explanation (e.g., Powers et al., 1992). Sweedler-Brown (1991) suggested two reasons for such a bias. First, scorers may have higher expectations of the typed essays. Typed essays may be more likely to be perceived as final drafts. Therefore, errors in typing and grammar may appear more prominent. Also, because typed essays look shorter and take less time to read, they may appear to be not well developed. Second, those who score the essays may unconsciously identify the individuality of each essay with its writer when scoring it. Handwritten essays make such personification easier. Raters may tend to judge handwritten essay quality through the personality they perceive and ignore some problems, while treating typed essays as a collection of writing skills.

The first reason—a tendency of the raters to evaluate essays written on computer as if they were final drafts—is a very plausible explanation for the score differences observed in this study. It may be possible to train raters to overcome this tendency.

The use of performance assessments for high-stakes decisions is increasing (Hollenbeck, Tindal, & Almond, 1999) and computer-based testing is becoming increasingly popular. According to "Technology Counts 2003" (Education Week, 2003), 12 states and the District of Columbia now deliver computer-based assessments. The techniques used in the present study can be applied to other testing situations involving paper and computer versions of the same test. If substantial demographic information about the examinees is available, these techniques make it possible to sharpen any comparisons of performance between examinees taking the paper and computer versions of the test. In that way, they may find wider application in educational measurement.

References

Arnold, V., Legas, J., Obler, S., Pacheco, M. A., Russell, C., & Umbdenstock, L. (1990). *Do students get higher scores on their word-processed papers? A study of bias in scoring hand-written vs. word-processed papers.* Unpublished paper, Rio Hondo College, Whittier, CA.

Bridgeman, B., & Cooper, P. (1998). *Comparability of scores on word-processed and handwritten essays on the Graduate Management Admissions Test.* Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA.

Education Week. (2003). Technology counts 2003. *Education Week, 22*(34). Retrieved on January 15, 2004, from http://www.edweek.org/ew/ewstory.cfm?slug=34tc_talkback.h22&keywords=Technology%20Counts

ETS. (2001). *Tests at a glance: Praxis I: Academic skills assessments.* Princeton, NJ: Author.

Hollenbeck, K., Tindal, G., & Almond, P. (1999). Reliability and decision consistency: An analysis of writing mode at two times on a statewide test. *Educational Assessment, 6*(1), 23–40.

Powers, D. E., & Farnum, M. (1997). Effects of mode of presentation on essay scores (ETS RM-97–8). Princeton, NJ: ETS.

Powers, D. E., Fowles, M. E., Farnum, M., & Ramsey, P. (1992). Will they think less of my handwritten essay if others word process theirs? *Journal of Educational Measurement, 31*, 220–233. Princeton, NJ: ETS.

Rosenbaum. P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician, 39*(1), 33–38.

Rubin, D. B., & Thomas, N. (2000). Combining propensity score matching with additional adjustments for prognostic covariates. *Journal of the American Statistical Association, 95*, 573–585.

Russell, M. (1999) Testing on computers: an follow-up study comparing performance on computer and on paper. *Education Policy Analysis Archives, 7*(20). Retrieved January 25, 2004, from http://epaa.asu.edu/epaa/v7n20/

Russell, M., & Haney, W. (1997) Testing writing on computers: An experiment comparing student performance on tests conducted via computer and via paper-and-pencil. *Education Policy Analysis Archives, 5*(3). Retrieved January 25, 2004, from http://epaa.asu.edu/epaa/v5n3.html

Sweedler-Brown, C. (1991). Computers and assessment: The effect of typing versus handwriting on the holistic scoring of essays. *Research & Teaching in Developmental Education, 8*(1), 5–14.

Appendix

Examinee Background Information Questionnaire on the Registration Form

18. BACKGROUND INFORMATION QUESTIONNAIRE
(Select one answer for each question below.)

***a. How do you describe yourself?**

1 ( ) African American or Black
2 ( ) Asian American/Asian (Ex.: Japanese, Chinese, Korean)
3 ( ) Southeast Asian American/Southeast Asian (Ex.: Cambodian, Hmong, Khmer, Laotian, Vietnamese)
4 ( ) Pacific Island American/Pacific Islander
5 ( ) Mexican, Mexican American, or Chicano
6 ( ) Puerto Rican
7 ( ) Other Hispanic, Latino, or Latin American
8 ( ) Native American, American Indian, or Alaskan Native
9 ( ) White
10 ( ) Other

b. What is your best language of communication?

1 ( ) English
2 ( ) Another language

c. Which language(s) did you first learn as a child?

1 ( ) English only
2 ( ) English and another language
3 ( ) Another language only

***d. What is the highest education level you have attained?**

1 ( ) Freshman (first year)
2 ( ) Sophomore (second year)
3 ( ) Junior (third year)
4 ( ) Senior (fourth or final year)
5 ( ) Earned bachelor's degree
6 ( ) Earned bachelor's degree plus additional credits
7 ( ) Earned master's degree
8 ( ) Earned master's degree plus additional credits
9 ( ) Earned doctoral degree

e. How many years has it been since you attended college or graduate school?

1 ( ) Currently attending college or graduate school
2 ( ) Less than 1 year
3 ( ) 1 - 3 years
4 ( ) 4 - 6 years
5 ( ) 7 - 10 years
6 ( ) More than 10 years

f. What is your cumulative undergraduate grade point average to date (based on a system where 4.0 = A)?

1 ( ) 3.5 - 4.0
2 ( ) 3.0 - 3.49
3 ( ) 2.5 - 2.99
4 ( ) 2.0 - 2.49
5 ( ) 1.5 - 1.99
6 ( ) below 1.5

g. Indicate the highest level of education completed by your father or male guardian.

1 ( ) Some high school or less
2 ( ) High school diploma
3 ( ) Some postsecondary education
4 ( ) Associate degree
5 ( ) Bachelor's degree
6 ( ) Some graduate or professional school
7 ( ) Graduate or professional degree
8 ( ) Unknown

h. Indicate the highest level of education completed by your mother or female guardian.

1 ( ) Some high school or less
2 ( ) High school diploma
3 ( ) Some postsecondary education
4 ( ) Associate degree
5 ( ) Bachelor's degree
6 ( ) Some graduate or professional school
7 ( ) Graduate or professional degree
8 ( ) Unknown

i. Is either of your parents in the education profession?

1 ( ) Yes 2 ( ) No

j. What was your most recent full-time occupation?

1 ( ) Student
2 ( ) Food service
3 ( ) Maintenance
4 ( ) Truck driver
5 ( ) Technician
6 ( ) Clerical/administrative support
7 ( ) Sales/retail
8 ( ) Managerial
9 ( ) Self-employed
10 ( ) School aide
11 ( ) Teacher
12 ( ) Professional/executive
13 ( ) Other
14 ( ) None

k. Are you or have you ever been enrolled in a teacher education program?

1 ( ) Currently 2 ( ) Formerly 3 ( ) Never

l. Your teaching status is:

1 ( ) Planning to enroll or currently enrolled in a teacher education program
2 ( ) Recently graduated and expect to begin teaching in the near future
3 ( ) 1 to 3 years teaching experience
4 ( ) More than 3 years teaching experience
5 ( ) Not planning to teach at this time

* Question 18a will be included for states or institutions that receive electronic reporting. Other questions with asterisks will be reported on the examinee score report. All other background questions are for research purposes only, and respondents will remain anonymous.