

COMPARING POSTSECONDARY MARKETING STUDENT PERFORMANCE ON COMPUTER-BASED AND HANDWRITTEN ESSAY TESTS

Allen D. Truell
Melody W. Alexander
Rodney E. Davis
Ball State University

ABSTRACT

The purpose of this study was to determine if there were differences in postsecondary marketing student performance on essay tests based on test format (i.e., computer-based or handwritten). Specifically, the variables of performance, test completion time, and gender were explored for differences based on essay test format. Results of the study indicate that there was no significant difference in postsecondary marketing student scores based on test format. There was, however, a significant difference in test completion time based on essay format. Postsecondary marketing students completed the computer-based essay test significantly faster than they did the handwritten essay test. Implications for postsecondary marketing educator are also discussed.

INTRODUCTION

The use of computer-based tests for assessing students has a long, established history. The use of these tests is likely influenced by many reported advantages. Goldberg and Pedulla (2002) supported this idea when they stated that “Moves toward computerized testing stem from the advantages it offers over the traditional paper-and-pencil format” (p. 1053). Some of the reported advantages of computer-based tests include immediate student feedback (Alderson, 2000; Barkley, 2002; Stevens, 2001), increased instructional time (Barkley, 2002; Truell & Davis, 2003), increased scoring accuracy (Stevens, 2001), increased test administration options (Alderson, 2000), more assessment opportunities (Barkley, 2002), records administration (Alderson, 2000; Stevens, 2001), and reduced testing costs (Barkley, 2002). All indicators point to not only the continued, but also the increased use of computer-based tests for student assessments (Barkley, 2002; Bugbee, 1996; Liefert, 2000; Shermis, Mzumara, & Bubnitz, 2001).

Despite the many advantages of using computer-based tests, there is a concern in the literature regarding student performance equivalence when compared with traditional paper and pencil tests. As noted by Davis and Gardner (2004) “A number of reviews regarding the statistical equivalence between paper-and-pencil based tests versus

computer-based tests have been documented, with mixed results” (p. 2). To further compound these mixed results, “Most of the literature regarding computer-based testing has focused on student performance on computer-based tests with objective type questions” (Truell & Davis, 2003, p. 29). Lee (2002) supported this contention by stating “. . . there has been a growing interest in the equivalence of computerized and paper-and-pencil multiple choice tests; however, little attention has been paid to open-ended tests such as writing assessments” (p. 136).

The results of the few studies comparing computer-based and handwritten essay performance have been mixed. For example, Bridgeman and Cooper (1998) investigated the comparability of computer-based and handwritten essay scores on the Graduate Management Admissions Test. Results of their analysis found that scores on the handwritten essays were higher than the scores on the computer-based essays. In addition, they indicated that this score difference did not interact with the English-as-a-Second-Language (ESL), ethnic, or gender variables. Manalo and Wolfe (2000) noted that the Test of English as a Foreign Language has been revised to include a writing component. They explained that individuals were given a choice of completing this written component in either a computer-based or handwritten format. They postulated that because of the variability of access to and comfort levels with computers that the results may not be comparable across formats. Their analysis found that handwritten essay scores were about 1/3 standard deviation higher than computer-based essay scores. Lee (2002) conducted a study to determine if differences existed in the computer-based and handwritten essay scores of ESL students. Although Lee (2002) noted that the computer-based essay responses contained more words and sentences than did the handwritten essay responses, the scores earned were not significantly higher based on format. MacCann, Eastment, and Pickering (2002) compared the essay scores of high school students between computer-based or handwritten test formats. The results of their study were inconclusive as the differences in scores were not consistent across the various essay tests administered. Lastly, Russell and Haney (1997) reported that student scores on the computer-based essays were significantly higher than were they were on the handwritten essays.

NEED FOR THE STUDY

While there has been considerable research on the topic of computer-based testing in general, few studies have compared student performance on the various essay test formats (i.e., computer-based or handwritten). Thus, the result of this study adds to the limited literature base regarding student performance on essays based on test format. In addition, this study builds upon the recommendation of Truell and Davis (2003) who suggested “A study of this type would provide additional information regarding student performance on computer-based essay questions” (p. 29). This additional insight is critical given the mixed results of the few studies that have examined student scores based on essay test format.

PURPOSE

The purpose of this study was twofold: (a) to determine if there were differences in postsecondary marketing student performance and time to essay test completion based on test format and (b) to determine if there were differences in postsecondary marketing student performance and time to essay test completion based on test format and gender. Specifically, the following research questions were explored.

1. Is there a significant difference in postsecondary marketing student performance between computer-based or handwritten essay tests?
2. Is there a significant difference in postsecondary marketing student completion time between computer-based or handwritten essay tests?
3. Is there a significant difference by gender in postsecondary marketing student performance between computer-based or handwritten essay tests?
4. Is there a significant difference by gender in postsecondary marketing student completion time between computer-based or handwritten essay tests?

METHODOLOGY

The research design, participants, data collection procedures, and data analysis are described in this section.

Research Design

A 2 X 2 Latin square quasi-experimental design was used for this study. Specifically, two intact postsecondary principles of marketing classes were the row factor, the two essay test formats were the column factor, and the method (i.e., computer-based or handwritten formats) was the treatment. This study design was used because “experimental control is achieved or precision enhanced by entering all respondents (or settings) into all treatments” (Campbell & Stanley, 1963, p. 50). As such, this design controls for most threats to internal validity. The Latin square design for this study is illustrated in Table 1.

Participants

Students enrolled in two intact postsecondary principles of marketing classes (32 students in each class) served as the study participants. Of the 64 postsecondary marketing students participating in the study, 33 were female and 31 were male (52 and 48 percent, respectively).

Table 1

Illustration of the 2 x 2 Latin Square Design

Row Factor	Column Factor	
	Essay Test 1	Essay Test 2
Class 1	A=Computer-Based Essay Test Format	B=Handwritten Essay Test Format
Class 2	B = Handwritten Essay Test Format	A = Computer-Based Essay Test Format

Data Collection Procedures

Students in both intact postsecondary marketing classes completed the same assignments, were taught by the same instructor, and were taught in the same classroom. The handwritten essay test formats were completed in the same classroom in which instruction took place. The computer-based essay tests were completed in a proctored computer-based testing lab located on campus. Participating postsecondary marketing students were notified in advance as to how the essay tests were going to be administered. Postsecondary marketing student test completion times for the computer-based essay tests were automatically recorded by the computer testing system. Student test completion times for handwritten essays were recorded by the proctor. For the handwritten essay tests participants were asked to keep all materials facedown until given the signal to begin by the proctor who noted the start time. Participating postsecondary marketing students were asked to submit their tests to the proctor immediately upon completion so the proctor could note accurately their respective ending times. To avoid potential essay test scoring bias based on essay test format, an analytical scoring procedure was used (Wang, 2000).

Data Analysis

To answer the research questions, MANOVA analyses were conducted. Post hoc ANOVAs were computed following each significant MANOVA analysis. All tests of significance were conducted at $\alpha = .05$. Power and effect size is reported for these analyses where appropriate. Omega squares (ω^2) are used to interpret effect size magnitude (Kirk, 1996).

FINDINGS

The following section presents the finding for each of the four research questions.

Research Question One

Research question one sought to determine if there was a significant difference in postsecondary marketing student scores on an essay test based on format. The

MANOVA analysis indicated that there was a significant difference in either score or time to test completion. A post hoc ANOVA analysis for differences on test scores $F(1, 127) = 0.676, p = 0.413$ indicated that there was not a significant difference in postsecondary marketing student test scores based on format. Table 2 presents the MANOVA and ANOVA analyses for research question one. Descriptive statistics for this analysis appear in Table 4.

Table 2

Analysis of Latin Square Design

Multivariate Tests: Effect	Model: (score time) = Class X Test Format X Test X Replication				
	Wilks' Lambda	<i>P</i>	Partial Eta ²	Observed Power	
Class	.775	<.001	.225	.997	
Test format	.881	.003	.119	.882	
Test	.888	.004	.112	.862	
Replications	.439	.018	.338	.999	
Univariate Tests: Effect	Type III SS	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>
<i>Dependent Variable (Score)</i>					
Class	.945	1	.945	.029	.865
Test Format	21.945	1	21.945	.676	.413
Test	25.383	1	25.383	.781	.379
Replications	1407.492	31	45.403	1.398	.112
Error	3030.977	93	32.484		
Total	4476.742	127			
<i>Dependent Variable (Time)</i>					
Class	962.508	1	962.508	19.617	<.001
Test Format	565.320	1	565.320	11.522	.001
Test	540.383	1	540.383	11.014	.001
Replications	2715.930	31	87.611	1.786	.018
Error	4563.039	93	49.065		
Total	9347.180	127			

Research Question Two

Research question two sought to determine if there was a significant difference in postsecondary marketing student completion time based on format. The MANOVA analysis indicated that there was a significant difference in either score or test completion time. A post hoc ANOVA analysis for differences on test scores $F(1, 127) = 11.522, p = 0.001$ indicated that there was a significant difference in student test completion time based on essay test format. Postsecondary marketing students completed the computer-based essay test format significantly faster than they did the handwritten essay test format. Table 2 presents the MANOVA and ANOVA analyses for research question

two. Descriptive statistics for this analysis appear in Table 4. The effect size for this analysis as interpreted by the ω^2 is 0.118, which is a medium effect size (Kirk, 1996).

Research Question Three

Research question three sought to determine if there was a significant difference between genders in postsecondary marketing student scores on an essay test based on format. The MANOVA analysis indicated that there was a significant difference in either test score or time to test completion based on gender. A post hoc ANOVA on test format score and gender difference, $F(1, 127) = 0.670, p = 0.451$ indicated that there was no significant difference in test score based on postsecondary marketing student gender and format. Table 3 presents the MANOVA and ANOVA analyses for research question three. Descriptive statistics for this analysis appear in Table 4.

Table 3

Analysis of Latin Square Design with Gender Added

Multivariate Tests: Effect	Model: (score time) = Class X Test Format X Test X Replication X Gender				
	Wilks' Lambda	<i>P</i>	Partial Eta ²	Observed Power	
Class	.774	<.001	.226	.997	
Test Format	.881	.003	.119	.879	
Test	.888	.004	.112	.858	
Replication	.439	.022	.337	.999	
Gender	.997	.876	.003	.070	
Univariate Tests: Effect	Type III SS	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>
<i>Dependent Variable (Score)</i>					
Class	1.191	1	1.191	.036	.849
Test Format	21.945	1	21.945	.670	.415
Test	25.383	1	25.383	.775	.381
Replication	1407.492	31	45.401	1.386	.118
Gender	6.912	1	6.912	.211	.647
Error	3014.064	92	32.762		
Total	4476.742	127			
<i>Dependent Variable (Time)</i>					
Class	960.126	1	960.126	19.358	<.001
Test Format	565.320	1	565.320	11.370	.001
Test	540.383	1	540.383	10.895	.001
Replication	2680.348	31	86.463	1.743	.220
Gender	.080	1	.018	.000	.985
Error	4563.021	92	49.598		
Total	9347.180	127			

Research Question Four

Research question four sought to determine if there was a significant difference by gender in postsecondary marketing student scores on an essay test based on format. The MANOVA analysis indicated that there was a significant difference in either test score or test completion time based on gender. A post hoc ANOVA analysis on test format score and gender difference, $F(1, 127) = 11.370, p = 0.001$ indicated that there was a significant difference in test scores based on postsecondary marketing student gender and format. Since the Latin square design does not allow for interaction analysis, it was not determined whether females were significantly faster than males when completing the computer-based, handwritten, or both test formats. Table 3 presents the MANOVA and ANOVA analyses for research question four. Descriptive statistics for this analysis appear in Table 4. The effect size for this analysis as interpreted by the ω^2 is 0.007. A ω^2 of 0.007 is a small effect size (Kirk, 1996).

Table 4

Descriptive Statistics for the Data in the Analysis

	N	Essay Test Score		Essay Test Time	
		M	SD	M	SD
Class					
First	64	21.32	6.29	20.61	7.94
Second	64	21.05	5.61	26.09	8.37
Test Format					
Computer based	64	21.55	4.88	25.45	7.13
Handwritten	64	20.72	6.85	21.25	9.41
Test					
First	64	21.58	6.05	25.41	8.56
Second	64	20.69	5.84	21.30	8.16
Gender					
Male	62	21.37	6.07	23.98	8.62
Female	66	20.91	5.85	22.76	8.57
Total	128	21.13	5.94	23.35	8.58

Note: Maximum possible tests score was 30 on all test versions; maximum possible time to complete each test version was 75 minutes.

CONCLUSIONS AND IMPLICATIONS

Based on the findings of this study a number of conclusions can be drawn. These conclusions, however, are put forward with caution and generalizing beyond the

participants is not possible. First, there is no difference in postsecondary marketing student scores based on format. Second, there is a difference in the essay test completion time based on test format. Postsecondary marketing students complete the computer-based format faster than postsecondary marketing students with the handwritten essay test format. Third, there is no difference in test scores based on test format and gender. Fourth, females completed the testing process faster than did the males. This difference, however, is too small to be of much practical importance. These conclusions translate into an important implication for practice. Chiefly, postsecondary marketing instructors are encouraged to adopt computer-based essay tests in their classes. No difference in postsecondary marketing student scores and faster essay test completion times offers advantages to both postsecondary marketing instructors and students. Both instructors and students benefit from the scheduling flexibility computer-based tests postsecondary marketing offer. In addition, both postsecondary marketing instructors and students benefit from faster essay test completion speeds.

RECOMMENDATIONS FOR FURTHER RESEARCH

Based on the findings of this study and a review of the relevant literature, the following recommendations for further research are offered:

1. This study should be replicated in other settings. Given that relatively few studies have examined student performance on computer-based and handwritten essay test formats, such a study would aid in growing the literature base regarding these student assessment options. This recommendation is especially important given the mixed results in the literature.
2. A study should be conducted to determine if postsecondary marketing students have a preference for completing essay tests using the computer-based or handwritten format. Such a study would provide additional insight into the use computer-based essay tests with postsecondary marketing students.
3. A study should be conducted to determine if students have a preference for other types of test question formats (i.e., matching, multiple choice, true/false, etc.). Such a study would provide additional insight in student testing preferences based on question format.

REFERENCES

- Alderson, J. C. (2000). Technology in testing: The present and the future. *System*, 28(4), 593-603.
- Barkley, A. P. (2002). An analysis of online examinations in college courses. *Journal of Agricultural and Applied Economics*, 34(4), 445-458.

- 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, April 13-17, 1998. (ERIC Document Reproduction Service No. ED421528)
- Bugbee, A. C., Jr. (1996). The equivalence of paper and pencil and computer-based tests. *Journal of Research on Computing in Education*, 28(3), 282-299.
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally & Company.
- Davis, J., & Gardner, T. (2004, April). *Effects of paper-based and computer-based administrations of high-stakes high-school graduation tests*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
- Goldberg, A. L., & Pedulla, J. J. (2002). Performance differences according to the test mode and computer familiarity on a practice graduate record exam. *Educational and Psychological Measurement*, 62(6), 1053-1067.
- Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, 56(5), 746-759.
- Lee, Y. (2002). A comparison of composing processes and written products in timed-essay tests across paper and pencil and computer modes. *Assessing Writing*, 8(2), 135-157.
- Liefert, J. (2000). Measurement and testing in a distance learning course. *Journal of Instructional Delivery Systems*, 14(2), 13-16.
- MacCann, R., Eastment, B., & Pickering, S. (2002). Responding to free response examination questions: Computer versus pen and paper. *British Journal of Educational Technology*, 33(2), 173-188.
- Manalo, J. P., & Wolfe, E. W. (2000). *The impact of composition medium on essay raters in foreign language testing*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA, April 24-28, 2000. (ERIC Document Reproduction Service No. ED443836)
- Russell, M., & Haney, W. (1997). Testing writing on computers: An experiment comparing student performance on tests conducted via computer and via paper-and-pencil. *Educational Policy Analysis Archives*, 5(3). Retrieved from <http://epaa.asu.edu/epaa/v5n3.html>
- Shermis, M. D., Mzumara, H. R., & Bubnitz, S. T. (2001). On test and computer anxiety: Test performance under CAT and SAT conditions. *Journal of Educational Computing Research*, 24(1), 57-75.