THE IMPACT OF SETTABLE TEST ITEM EXPOSURE CONTROL INTERFACE FORMAT ON POSTSECONDARY BUSINESS STUDENT TEST PERFORMANCE

Allen D. Truell
Jensen J. Zhao
Melody W. Alexander
Ball State University

ABSTRACT

The purposes of this study were to determine if there is a significant difference in postsecondary business student scores and test completion time based on settable test item exposure control interface format, and to determine if there is a significant difference in student scores and test completion time based on settable test item exposure control interface format by gender. Results of the study indicate that there is no significant difference in postsecondary business student scores or test completion times based on settable test item exposure control interface format. When the variable gender was added, female postsecondary business students were found to achieve significantly higher test scores and to have significantly faster test completion times. Effect size and descriptive statistic analysis suggests that these differences by gender are too small to be of much practical difference.

INTRODUCTION

A search of the *ERIC* database reveals a keen interest in computer-based testing by researchers over the past 35 years. Indeed, a focused search of the *ERIC* database using the descriptor "computer assisted testing" from 1970 through 2003 returned 1,954 citations. More than half (55.6%, n = 1,105) of these 1,954 citations were dated from 1990 through 2003. This research interest in computer-based testing is likely a result of the many advantages associated with its use (Goldberg & Pedulla, 2002). A number of researchers have reported on the advantages of computer-based testing (e.g., Alderson, 2000; Alexander, Bartlett, Truell, & Ouwenga, 2001; Barkley, 2002; Bocij & Greasley, 1999; DeSouza & Fleming, 2003; Goldberg & Pedulla, 2002; Greenberg, 1998; Shermis & Lombard, 1998; Shermis, Mzumara, & Bublitz, 2001; Song, 1998; Stephens, 2001; Truell & Davis, 2003). Often cited advantages of computer-based testing include decreased testing costs, effective records management, increased assessment options, improved scoring precision, instant feedback to students, more instructional time, more test administration choices, and reduced testing time. Despite the many advantages associated with computer-based tests for student assessment purposes, there are several

areas of concern associated with their use. Two areas of concern with computer-based test use are user interfaces and test item exposure control formats.

For example, a number of researchers have expressed concern with the potential impact of the user interface on student test performance (Booth, 1991, 1998; Huff & Sireci, 2001; Parshall, Spray, Kalohn, & Davey, 2002; Ricketts & Wilks, 2002). In addition, only a few researchers have investigated various test item exposure control features associated with computer-based testing use (e.g., Cheng & Loui, 2003; Davis, Pastor, Dodd, Chaing, & Fitzpatrick, 2003; Meijer & Nering, 1999; O'Neill, Lunz, & Thiede, 2000; Pastor, Dodd, & Chang, 2002; Ryan & Chiu, 2001; Stocking & Lewis, 1998; Stocking, Ward, & Potenza, 1998; van der Linden & Chang, 2003). The majority of the test item exposure control research focused on the impact of test items selected to be exposed to a test taker from large test item pools. Further, computer-based testing systems have caused some researchers to express concern that its equivalency with traditional testing techniques be confirmed (Alexander et al., 2001; Bugbee & Bernt, 1990; Bugbee, 1996; Truell & Joyner, 2003; Truell, 2005). Finally, Truell (2005) recommended that research was needed regarding the various settable interface formats available to faculty using computer-based testing systems.

NEED FOR THE STUDY

In recent years there has been a growing use of computer-based testing systems in postsecondary education. This increased growth is associated with the many advantages of their use for assessing student performance. Despite this growth and reported advantages, researches have noted several issues of concern. Specifically, this concern has focused on the user interface and test item exposure control formats. Thus, the results of this study fill a gap in the literature by addressing research recommendation put forward in the literature.

PURPOSE OF THE STUDY

The purposes of this study were (a) to determine if there is a significant difference in postsecondary business student test scores and test completion times based on settable test item exposure control interface format (i.e., all at once, one at a time—backing up, and one at a time—no backing up) and (b) to determine if there is a significant difference in postsecondary business student test score and test completion time based on settable test item exposure control interface format (i.e., all at once, one at a time—backing up, and one at a time—no backing up) by gender. Thus, the following research questions were investigated.

1. Is there a significant difference in postsecondary business student test scores based on settable test item exposure control interface format?
2. Is there a significant difference in postsecondary business student test completion time based on settable test item exposure control interface format?
3. Is there a significant difference in postsecondary business student test scores based on settable test item exposure control interface format by gender?

4. Is there a significant difference in postsecondary business student test completion time based on settable test item exposure control interface format by gender?

METHODOLOGY

Research Design

The counterbalanced, Latin square quasi-experimental design was used in this study. Specifically, the counterbalanced Latin square design was selected because ". . . experimental control is achieved or precision enhanced by entering all respondents (or setting) into all treatments" (Campbell & Stanley, 1963, p. 50). Additionally, this design controls for the majority of threats to internal validity (Campbell & Stanley, 1963). Treatment order was determined by random assignment. The specific counterbalanced, Latin square design used in this study is illustrated in Table 1.

Table 1. *Illustration of the 3 x 3 Counterbalanced, Latin Square Design*

Row Factor	Column Factor		
	Test 1	Test 2	Test 3
Class 1	All at Once	One at a Time—Backing Up	One at a Time—No Backing Up
Class 2	One at a Time—Backing Up	One at a Time—No Backing Up	All at Once
Class 3	One at a Time—No Backing Up	All at Once	One at a Time—Backing Up

Participants

Participants were those postsecondary business students enrolled in three, intact sections of the same college of business core course at a Midwestern university. More specifically, 90 students participated in the study. The number of students participating in each class was 34, 32, and 24, respectively.

Data Collection Procedures

The commercially available computer-based testing system used during this study automatically recorded postsecondary business student test score and test completion time data. The three classes were taught by the same instructor, met in the same classroom, and were provided with the same instructional materials. Classes met on a three day per week schedule. All computer-based tests were completed in a computer lab located near the classroom. All tests were proctored by the instructor. Students were allotted 50 minutes to complete each 50-item multiple choice test regardless which settable test item exposure control interface format.

Data Analysis

To answer research questions one, two, three, and four, MANOVAs and post hoc ANOVAs were used to analyze the data. There were 34, 32, and 24 postsecondary business students enrolled in the three, intact classes involved in this study, respectively. The Latin square design assumes an equal number of participants in each class so data from 24 postsecondary business students in each of the classes enrolling more than 24 students was randomly selected for inclusion in the data analysis. In order to form each of the 24 Latin squares, postsecondary business students were randomly matched across the three classes. Since each Latin square contained four observations and there were 24 replications, the data set had 72 observations. Effect size and observed power are reported in the findings section. As Kotrlik and Williams (2003) noted "It is almost always necessary to include some index of effect size or strength of relationship in your results section . . ." (p. 1). Effect size magnitude in this study was determined using Omega square ($\omega 2$) values. Kirk's (1996) procedure for interpreting $\omega 2$ effect size magnitude is used in this study. Tests of statistical significance were conducted at $\alpha = .05$.

FINDINGS

Research Question One

Research question one sought to determine if there was a significant difference in postsecondary business student scores based on settable test item exposure control interface format. Results of the MANOVA—Hotelling's Trace—analysis indicated that there was no significant difference in postsecondary business student test scores based on settable test item exposure control interface format. MANOVA and ANOVA analyses for research question one and their associated descriptive statistics appear in Tables 2 and 4, respectively.

Research Question Two

Research question two sought to determine if there was a significant difference in student test completion time based on settable test item exposure control interface format. MANOVA—Hotelling's Trace—analysis indicated there was no significant difference in postsecondary business student test completion time based on settable test item exposure control interface format. MANOVA and ANOVA analyses for research question two and their associated descriptive statistics appear in Tables 2 and 4, respectively.

Research Question Three

Research question three sought to determine if there was a significant difference by gender in student scores based on settable test item exposure control interface format. MANOVA—Hotelling's Trace—analysis indicated either a significant difference in postsecondary business student test score or test completion time by gender. Post-hoc

Table 2. *Analysis of Latin Square Design*

Model: (score time) = Class X Test Item Exposure Control Interface Format X Test X Replication

Multivariate Tests

Effect	Hotelling's Trace	*P*	Partial *Eta*^{2}	Observed Power
Class	0.055	0.040	0.027	0.717
Test Item Exposure Control Interface Format	0.055	0.913	0.003	0.103
Test	0.241	0.000	0.107	1.000
Replications	0.440	0.003	0.180	1.000

Univariate Tests

Effect	*Type III SS*	*df*	*MS*	*F*	*p*	ω^2
Dependent Variable (Score)						
Class	97.287	2	48.644	2.761	0.066	0.013
Test Item Exposure Control Interface Format	17.343	2	8.672	0.490	0.614	-0.004
Test	675.398	2	337.699	19.168	<0.001	0.134
Replications	656.218	23	28.531	1.619	0.043	0.052
Error	3294.861	186	17.714			
Total	4741.106	215				
Dependent Variable (Time)						
Class	484292.565	2	24146.283	2.232	0.110	0.010
Test Item Exposure Control Interface Format	899.287	2	449.644	0.004	0.996	-0.008
Test	898792.954	2	449396.477	4.143	0.017	0.026
Replications	4568370.204	23	198624.791	1.831	0.015	0.077
Error	20391083.639	186	109629.482			
Total	26343438.648	215				

ANOVA analysis $F(1, 185) = 11.164$, $p = 0.001$ indicated that there was a significant difference by gender in student scores based on settable test item exposure control interface format. Specifically, postsecondary business female students scored significantly higher than did male students based on settable test item exposure control interface format. The means and standard deviations for female and male postsecondary business students were 43.87 ($SD = 3.74$) and 41.56 ($SD = 4.85$), respectively. These means and standard deviation differences are too small to be of much practical significance, however. This lack of practical differences by gender in postsecondary business student scores is supported by the effect size for the analysis. The effect size for this analysis is $\omega 2 = 0.036$. A $\omega 2$ of <0.05 is considered a small effect size (Kirk, 1996).

Table 3. *Analysis of Latin Square Design with Gender Added*

Model: (score time) = Class X Test Item Exposure Control Format X Test X Replication X Gender

Multivariate Tests

Effect	Hotelling's Trace	*P*	Partial Eta^2	Observed Power
Class	0.060	0.028	0.029	0.756
Test Item Exposure Control Format	0.006	0.906	0.003	0.106
Test	0.255	0.000	0.113	1.000
Replication	0.423	0.005	0.174	0.999
Gender	0.076	0.001	0.071	0.924

Univariate Tests

Effect	*Type III SS*	*df*	*MS*	*F*	*p*	ω^2
Dependent Variable (Score)						
Class	117.116	2	58.558	3.503	0.032	0.018
Test Item Exposure Control Interface Format	17.343	2	8.672	0.516	0.598	-0.003
Test	675.398	2	337.699	20.204	<0.001	0.135
Replication	607.237	23	26.402	1.580	0.052	0.046
Gender	186.595	1	186.595	11.164	0.001	0.036
Error	3108.266	185	16.801			
Total	4741.106	215				
Dependent Variable (Time)						
Class	417674.624	2	208837.312	1.959	0.144	0.008
Test Item Exposure Control Interface Format	899.287	2	449.644	0.004	0.996	-0.008
Test	898792.954	2	449396.477	4.215	0.016	0.026
Replication	4232107.733	23	184004.684	1.726	0.026	0.066
Gender	452812.539	1	452812.539	4.247	0.041	0.013
Error	19938271.100	185	107774.438			
Total	26343438.648	215				

Research Question Four

Research question four sought to determine if there was a significant difference by gender in postsecondary business student test completion time based on settable test item exposure control interface format. MANOVA—Hotelling's Trace—analysis indicated either a significant difference in postsecondary business student scores or test completion times by gender based on settable test item exposure control interface format. Post-hoc ANOVA analysis $F(1, 185) = 4.247$, $p = 0.041$ indicated that there was a significant difference by gender in postsecondary business student test completion times based on

Table 4. *Descriptive Statistics for the Data in the Analysis*

Class	Frequency	Test Score		Test Time	
		M	*SD*	*M*	*SD*
First	72	42.78	4.74	1336.15	333.15
Second	72	42.42	5.17	1444.72	340.53
Third	72	41.21	4.03	1355.10	370.49
Test Item Exposure Control Interface Format					
All at Once	72	42.51	4.34	1377.29	415.12
One at a Time—Backing Up	72	42.06	4.70	1377.14	327.55
One at a Time—No Backing Up	72	41.83	5.06	1381.54	302.34
Test					
First	72	44.60	3.84	1346.75	328.27
Second	72	40.53	5.31	1468.63	402.18
Third	72	41.28	3.78	1320.60	298.10
Gender					
Male	162	41.56	4.85	1415.10	363.45
Female	54	43.87	3.74	1269.31	282.01
Total	216	42.13	4.70	1378.66	350.04

Note. Maximum possible test score was 50 regardless of settable test item exposure control interface format; maximum possible test completion time was 50 minutes regardless of settable test item exposure control interface format; time recorded and analyzed in seconds.

settable test item exposure control interface format. Specifically, female postsecondary business students achieved significantly faster test completion times than did male postsecondary business students based on settable test item exposure control interface format. The means and standard deviations for female and male postsecondary business students were 1269.31 (*SD* = 282.006) and 1415.10 (*SD* = 363.452) seconds, respectively. These means and standard deviation differences are too small to be of much practical significance, however. This lack of practical differences by gender in student scores is supported by the effect size for the analysis. The effect size for this analysis is $\omega 2 = 0.013$. A $\omega 2$ of <0.05 is considered a small effect size (Kirk, 1996). MANOVA and ANOVA analyses for research question four and their associated descriptive statistics appear in Tables 3 and 4, respectively.

CONCLUSIONS AND DISCUSSION

The results of this study offer several conclusions. These conclusions, however, are offered with the caveat that this study appears to be among the first to examine the impact of various settable test item exposure control interface formats and that additional investigation is needed. First, there is no significant difference in postsecondary business student performance based on settable test item exposure control interface format.

Specifically, postsecondary business student test scores and test completion times did not differ significantly regardless of settable test item exposure control interface format. Second, female postsecondary business student performance on both test score and test completion time were significantly different from their postsecondary business student male counterparts. This significant difference for both test scores and test completion time is likely of little practical difference. These conclusions are supported by data in Tables 1, 2, 3, and 4. The results of this study are consistent with the earlier work of Truell (2005) who examined if differences existed in student scores and test completion time based on two computer-based user interface and paper and pencil formats.

Truell (2005) reported that there was no significant difference in student scores based on test presentation format. In addition, there was no significant difference in test completion times between the two computer-based user interface test formats. Interesting, when gender was included in the analysis, female students scored significantly higher and achieved significantly faster test completion times than did their male counterparts. Truell (2005), after examining the effect size and descriptive statistics for each analysis, noted that these significant differences by gender were likely of little practical difference. The practice implication resulting from this study is that postsecondary business faculty can proceed with using the various settable test item exposure control interface formats. This use of various settable test item exposure control interface formats should be done with caution until more research has been conducted into their potential impact on test performance, however.

RECOMMENDATIONS FOR FURTHER RESEARCH

Based on a review of the literature and the findings of this study, the following recommendations for further research are put forward.

1. This study should be replicated. Given that relatively few studies have examined test item exposure control interface procedures, it would be prudent to conduct additional research in a variety of settings. Such studies would provide additional insight into the impact of settable test item exposure control interface features available with the various commercially available computer-based testing systems.
2. As new settable testing features become available, research should be conducted to determine their potential impact on postsecondary business student test performance. Such studies will provide insight as to the impact of evolving technology on postsecondary business student computer-based test performance.

REFERENCES

Alderson, J. C. (2000). Technology in testing: The present and the future. *System, 28*(4), 593-603.

Alexander, M. W., Bartlett, J. E., II, Truell, A. D., & Ouwenga, K. (2001). Testing in a computer technology course: An investigation of equivalency in performance

between online and paper and pencil methods. *Journal of Career and Technical Education, 18*(1), 69-80.

Barkley, A. P. (2002). An analysis of online examinations in college courses. *Journal of Agricultural and Applied Economics, 34*(4), 445-458.

Bocij, P., & Greasley, A. (1999). Can computer-based testing achieve quality and efficiency in assessment? *International Journal of Educational Technology, 1*(1), 1-18.

Booth, J. F. (1991). The key to valid computer-based testing: The user interface. *European Review of Applied Psychology, 41*(4), 281-293.

Booth, J. F. (1998). The user interface in computer-based selection and assessment: Applied and theoretical problematics of an evolving technology. *International Journal of Selection and Assessment, 6*(2), 61-82.

Bugbee, A. C., Jr. (1996). The equivalence of paper and pencil and computer-based testing. *Journal of Research on Computing in Education, 28*(3), 282-299.

Bugbee, A. C., Jr., & Bernt, F. M. (1990). Testing by computer: Findings in six years of use. *Journal of Research on Computing Education, 23,* 87-100.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally & Company.

Cheng, P. E., & Liou, M. (2003). Computerized adaptive testing using the nearest-neighbors criterion. *Applied Psychological Measurement, 27*(3), 204-216.

Davis, L. L., Pastor, D. A., Dodd, B. G., Chiang, C., & Fitzpatrick, S. J. (2003). An examination of exposure control and content balancing restrictions on item selection in CATs using the partial credit model. *Journal of Applied Measurement, 4*(1), 24-42.

DeSouza, E., & Fleming, M. (2003). A comparison of in-class and online quizzes on student exam performance. *Journal of Computing in Higher Education, 14*(2), 121-134.

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.

Greenberg, R. (1998). Online testing. *Techniques, 73*(3), 26-28.

Huff, K. L., & Sireci, S. G. (2001). Validity issues in computer-based testing. *Educational Measurement: Issues and Practices, 20*(3), 16-25.

Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement, 56*(5), 746-759.

Kotrlik, J. W., & Williams, H. A. (2003). The incorporation of effect size in information technology, learning, and performance research. *Information Technology, Learning, and Performance Journal, 21*(1), 1-7.

Meijer, R. R., & Nering, M. L. (1999). Computerized adaptive testing: Overview and introduction. *Applied Psychological Measurement, 23*(3), 187-194.

O'Neill, T., Lunz, M. E., & Thiede, K. (2000). The impact of receiving the same items on consecutive computer adaptive test administrations. *Journal of Applied Measurement, 1*(2), 131-151.

Parshall, C. G., Spray, J. A., Kalohn, J. C., & Davey, T. (2002). *Practical considerations in computer-based testing.* New York: Springer.

Pastor, D. A., Dodd, B. G., & Chang, H. H. (2002). A comparison of item selection techniques and exposure control mechanisms in CATs using the generalized partial credit model. *Applied Psychological Measurement, 26*(2), 147-163.

Ricketts, C., & Wilks, S. J. (2002). Improving student performance through computer-based assessment: Insights from recent research. *Assessment & Evaluation in Higher Education, 27*(5), 475-479.

Ryan, K. E., & Chiu, S. (2001). An examination of item context effects, DIF, and gender DIF. *Applied Measurement in Education, 14*(1), 73-90.

Shermis, M. D., & Lombard, D. (1998). Effects of computer-based test administration on test anxiety and performance. *Computers in Human Behavior, 14*(1), 111-123.

Shermis, M. D., Mzumara, H. R., & Bublitz, S. T. (2001). On test and computer anxiety: Test performance under CAT and SAT conditions. *Journal of Educational Computing Research, 24*(1), 57-75.

Song, J. K. (1998). *Using the World Wide Web in education and training.* Paper presented at the 1998 Information Technology and Training Conference. (ERIC Document Reproduction Service No. ED417703)

Stephens, D. (2001). Use of computer assisted assessment: Benefits to students and staff. *Education for Information, 19,* 265-275.

Stocking, M. L., & Lewis, C. (1998). Controlling item exposure conditional on ability in computerized adaptive testing. *Journal of Educational and Behavioral Statistics, 23*(1), 57-75.

Stocking, M. L., Ward, W. C., & Potenza, M. T. (1998). Simulating the use of disclosed items in computerized adaptive testing. *Journal of Educational Measurement, 35*(1), 48-68.

Truell, A. D. (2005). Comparing Student Performance on Two Computer-Based User Interfaces and Paper-and Pencil-Test Formats. *NABTE Review,32,* 29-35.

Truell, A. D., & Davis, R. E. (2003). Computer based testing: Adding value in the principles of marketing classroom. *The Ohio Business Technology Educator, 62*, 21-32.

Truell, A. D., & Joyner, R. L. (2003). Foundations of Business Communication Students' Performance on Online Computer-Based and Paper and Pencil Test Formats at a NABTE Institution, *NABTE Review, 30,* 42-47.

van der Linden, W. J., & Chang, H. H. (2003). Implementing content constraints in alpha-stratified adaptive testing using a shadow test approach. *Applied Psychological Measurement, 27*(2), 107-120.

THE AUTHORS

Allen D. Truell is an Associate Professor at Ball State University, Miller College of Business, ISOM Department, Muncie, IN 47306. Phone: (765) 285-5235. Email: atruell@bsu.edu.

Jensen J. Zhao is a Professor at Ball State University, Miller College of Business, ISOM Department, Muncie, IN 47306. Phone: (765) 285-5233. Email: jzhao@bsu.edu.

Melody W. Alexander is a Professor at Ball State University, Miller College of Business, ISOM Department, Muncie, IN 47306. Phone: (765) 285-5239. Email: malexand@bsu.edu.