Winter 1-16-2009

Proctored Versus Unproctored Online Exams: Studying the Impact of Exam Environment on Student Performance

Kimberly Hollister
Montclair State University, hollisterk@mail.montclair.edu

Mark L. Berenson
Montclair State University, berensonm@montclair.edu

Decision Sciences Journal of Innovative Education
Volume 7 Number 1
January 2009
Printed in the U.S.A.

EMPIRICAL RESEARCH

Proctored Versus Unproctored Online Exams: Studying the Impact of Exam Environment on Student Performance

Kimberly K. Hollister† and Mark L. Berenson
Department of Management and Information Systems, Montclair State University, Montclair, NJ 07043, 973-655-7521, e-mail: hollisterk@mail.montclair.edu, berensonm@mail.montclair.edu

ABSTRACT

Increasing numbers of universities are offering courses in online and hybrid formats. One challenge in online assessment is the maintenance of academic integrity. We present a thorough statistical analysis to uncover differences in student performance when online exams are administered in a proctored environment (i.e., in class) versus an unproctored environment (i.e., offsite). Controlling for student grade point average (GPA), no significant differences in mean overall course performance or exam performance between the two groups were found, nor were there any differences in the mean vectors of individual exam scores. The study reveals that the group taking online exams in the unproctored environment has significantly more variation in their performance results. In examining potential causes of the greater variation, analyses were performed to assess whether an increased level of possible cheating behavior could be observed from performance results for students in the unproctored section. No evidence of cheating behavior was found.

Subject Areas: Student Performance, Teaching Approaches, Online Assesment, Proctored, Offsite.

INTRODUCTION

At increasing rates, institutions of higher education are using technology in the presentation of course content. The availability and popularity of online learning is growing; 2 years ago the percentage of institutions offering Web-based courses in the United States reached 56% (Liu, 2006). As universities move toward the development of online/distance courses, they are faced with a number of challenges. One of these challenges is how to adequately assess student learning in an online environment.

The transition of a traditional course to either an online or hybrid format requires faculty to evaluate their assessment strategies. As with assessment in traditional courses, the development of an assessment portfolio within a course should be linked to the school's mission, degree-specific learning goals, and course-specific learning goals. Instructors need to match appropriate assessment

†Corresponding author.

techniques (e.g., homework, case projects, exams, presentations, participation, etc.) with the specific objectives of each course. Effective techniques to assess student learning in an online environment are addressed in the literature (Bryant, Kahle, & Schafer, 2005; Gaytan, 2005; Robles & Braathen, 2002).

Most of the work discussing performance in online and hybrid courses focuses on comparing student learning in an online course with a traditional course (Anstine & Skidmore, 2005; Liu, 2006; Weber & Lennon, 2007). Other works focus on the impact of learning styles in students' preference for various pedagogical techniques used in distance- or hybrid-based courses (Becker, Kehoe, & Tennent, 2007).

The most commonly reported challenge in online assessment is how to maintain academic integrity (which is a challenge in traditional courses as well); specifically, how do we catch cheating, verify the identity of the student, and curtail plagiarism (Byrd & Lott, 2003; Scanlon, 2003). Scanlon (2003) further asserts that no existing technology can ensure academic honesty. Commonly the literature "encourage[s] faculty to hold examinations on campus thereby ensuring a higher degree of academic honesty" (Alexander, Truell, & Bartlett, 2002; Gaytan, 2005). The assumption is that online testing will increase academic dishonesty and therefore result in inflated student performance. Peng (2007) attempts to evaluate the impact of cheating in an online assessment setting; his research uses univariate analysis in a multivariate setting that, according to Manly (1986), would fail to take into account interrelationships among variables.

Clearly, faculty are concerned with maintaining academic integrity in their classes. Overcoming the concern that students will commit more acts of academic dishonesty in online/distances courses is critical to obtaining buy-in from faculty asked to participate in an online/distance course. This research offers insight into quantifying the impact of varying the online exam environment on student performance.

In this article, we examine whether there is a difference in undergraduate student performance in an introductory computer literacy course at a state, comprehensive university when online exams are administered in a proctored environment (i.e., in class) versus an unproctored environment (i.e., offsite). The remainder of this article explores the effects of test environment on student performance, both on exams and over the course of the semester. The next section describes the course and the overall study groups; this section includes a descriptive and inferential evaluation of the overall study groups. In the section following, we describe the "performing" study groups that are used to study the effect of test environment on student performance; this section includes both descriptive and inferential evaluations for the "performing" study groups. We then present our findings on the effect of exam environment on student performance. In a later section, we examine exam performance patterns in the "performing" study groups to assess differences in academic integrity between the two groups. Our final section includes a discussion indicating the scope and limitations of the study and their impact on key findings and an agenda for further research.

THE OVERALL STUDY GROUPS

Subjects for this research were 217 students enrolled in INFO273, Introduction to the Computer in Business, the required computer literacy course for all

undergraduate business majors. The course was divided into three modules as follows: Module 1—Word and PowerPoint, Module 2—Excel, and Module 3—Access. The INFO273 course was taught using a hybrid course model (Koppel & Hollister, 2004), in which the primary method of instruction was through the use of online course tutorials. Assessment in the course had three main components—homework, case studies, and examinations. The focus in course assessments was twofold—to assess students' competency with basic computer literacy as well as their ability to apply computer skills to solve "real-world" problems in business environments. With the exception of a few mandatory lectures, all course meetings were optional for students. However, the instructor and/or graduate assistants were present during each class and acted as an instructional resource for all questions relating to course material or requirements.

The 217 student subjects for this research were enrolled in two sections of the course that met at similar times during the day and on comparable days of the week (Section 1 met on Tuesdays while Section 2 met on Thursdays). Additionally, both sections were identical in form, curriculum, and types of assessment (i.e., homework, case study projects, and exams) so that students in both sections should have had similar learning experiences. Students in both sections are traditional college-age students.

The only distinguishing learning variable difference between the groups was the environment in which their online exams would be taken; the intent of the study was to control for extraneous factors that might impact on pedagogy, learning, and student performance. One of the sections (i.e., the proctored group) was randomly assigned to take in-class exams for the three course-subject topics (Word with PowerPoint, Excel, and Access) in a proctored computer lab on a scheduled date and time, and the other section (i.e., the unproctored group) would take the same three exams outside the classroom environment within specified time blocks (4-day window of time). Both sections were limited to the same 60-minute span of time to complete each exam once it was launched. All exams administered in this course were hands-on, activity-based exams where students had to perform tasks in a simulated computer environment; exams were graded through the testing system.

Separate course syllabi were distributed to the sections on the first day of class that were identical except for the discussion regarding the dates and parameters for taking and submitting the three exams. The students in the two sections were not informed about the difference in the test-taking opportunities that was intended to be the basis for comparing learning and performance. As the semester progressed some of the students learned that one of the sections was taking the exams in the proctored computer lab while the other was not.

The random assignment of "testing environment" to the two sections required an examination of, and attempt at, controlling extraneous factors that might otherwise confound the desired analysis when comparing student performance. The student registration process was monitored and although both sections eventually filled to capacity it was clear that the Tuesday section (later designated as the proctored group) filled faster than the Thursday section. University policy provides for student registration privileges based on several factors. Students with seniority, in good-standing, who are in honors programs, on scholarship, involved in athletics, involved in student government leadership, or who are employed on campus

register earlier than those without these credentials. Three potential covariates or *concomitant* variables—cumulative grade point average (GPA), number of credits completed, and transfer status (i.e., whether or not the student transferred from another academic institution) were selected as possible filters of confounding effects owing to differences in the two student study groups.

Based on randomization, students enrolled in the Tuesday section were designated as Group 1 (i.e., proctored group) and those enrolled in the Thursday section were designated as Group 2 (i.e., unproctored group). Three of the 217 students formally withdrew from the course during the designated university drop period while 214 students—107 in each of the two groups—remained officially enrolled throughout the semester and completed at least some (if not all) assignments and/or exams and received final course letter grades ranging from A to F.

Overall, the purpose of this study was to determine what effect, if any, test environment had on student performance. Prior to focusing on student performance on the three exams, we developed a number of research hypotheses regarding differences in students' characteristics as well as possible differences in performance between the two groups.

Questions for Establishing Hypotheses for Testing in the Overall Study Groups

1. Do students in Group 1 (i.e., proctored group) have a different mean cumulative grade point average (GPA) than students in Group 2 (i.e., unproctored group)?
2. Have students in Group 1 completed a different mean number of credit hours than students in Group 2?
3. Do Groups 1 and 2 differ with respect to the proportion of transfer students enrolled in the classes?
4. Given that both sections are offered in a hybrid format that differs from the more traditional formally structured classroom environment with which students are familiar, will there be a difference in "course buy-in" between Groups 1 and 2 as reflected by:
   - Number of missed assignments (homework and case studies)?
   - Number of missed exams?
   - Number of attempts at completing online test preparation training tutorials?
5. Given that the overall course performance resulting in a letter grade of A to F is based on a weighted average of the 3 exams (Word with PowerPoint, Excel, and Access), 3 case studies (one for each of the aforementioned course topics), and 11 homework assignments (5 in Word with PowerPoint, 5 in Excel, and 1 in Access), will there be a difference between Groups 1 and 2 with respect to overall course average?

Descriptive and Inferential Evaluations of Overall Study Groups

Key descriptive statistical summary measures along with the results of the corresponding tests of hypotheses pertaining to the above questions 1 through 5 are displayed in Table 1.

Table 1: Summary of results of analysis of specific research hypotheses 1–5 for overall study groups.

Test	Specific Research Hypotheses	Summary Statistics Proctored Group (107) vs. Unproctored Group (107)	Tests	*p*-Values
1	GPA by group	$\overline{X_1} = 3.046$, $S_1 = .550$ $\overline{X_2} = 2.891$, $S_2 = .614$	Pooled variance *t*-test	.054
2	Mean credit hours by group	$\overline{X_1} = 66.97$, $S_1 = 29.56$ $\overline{X_2} = 58.12$, $S_2 = 27.07$	Pooled variance *t*-test	.024*
3	Proportion of transfer students by group	$P_1 = .336$ $P_2 = .308$	*Z*-test for differences in two independent proportions	.660
4a	Course "buy in": number of missed assignments by group	$\overline{X_1} = 1.16$, $S_1 = 2.303$ $\overline{X_2} = 2.33$, $S_2 = 3.605$	Separate variance *t*-test	.005**
4b	Course "buy in": proportion of students missing exams by group			
	• Missing at least 1 exam	$P_1 = .159$ $P_2 = .224$	*Z*-test for differences in two independent proportions	.224
	• Missing 2 or more exams	$P_1 = .028$ $P_2 = .131$	*Z*-test for differences in two independent proportions	.005**
	• Missing all 3 exams	$P_1 = .009$ $P_2 = .065$	*Z*-test for differences in two independent proportions	.031*
4c	Course "buy in": proportion of students attempting training by group			
	• Attempted Word/PowerPoint	$P_1 = .439$ $P_2 = .346$	*Z*-test for differences in two independent proportions	.162
	• Attempted Excel	$P_1 = .421$ $P_2 = .243$	*Z*-test for differences in two independent proportions	.006**
	• Attempted Access	$P_1 = .523$ $P_2 = .271$	*Z*-test for differences in two independent proportions	.000**
5	Overall course average by group	$\overline{X_1} = .845$, $S_1 = .144$ $\overline{X_2} = .759$, $S_2 = .246$	Separate variance *t*-test	.002**

*Results significant at the .05 level.
**Results significant at the .01 level.

The first three hypotheses presented in Table 1 concern the selected potential covariates for the study. Using the traditional .05 level of significance (two-tail), there was no evidence of a significant difference in the mean GPA between the two groups (p-value = .054) and there was no evidence of a significant difference in the proportion of transfer students in the two groups (p-value = .660). On the other hand, there was evidence that Group 1 students have, on average, completed more credits than their counterparts from Group 2 (p-value = .024). This latter result is interesting, if not surprising. Students registering earlier chose to take the Tuesday rather than Thursday classes offered in the same time periods and taught by the same instructor.

Hypothesis 4 and its respective subhypotheses all concern course "buy-in" between the two groups of students. From Hypothesis 4a, there was evidence that Group 1 students miss, on average, significantly fewer assignments than do their Group 2 counterparts (p-value = .005).

Given, however, that the three examination scores are the major response variables in this overall study it was particularly important to explore potential patterns in student absences on exams between the two groups (Hypothesis 4b). Although there was no evidence of a significant difference in the proportion of students over the two groups who missed one examination (p-value = .224), a "buy-in" effect was noted between the two groups when comparing the proportion of students who miss two or more of the exams. These proportions were significantly higher in the unproctored group with respect to missing two or more exams (p-value = .005) and with respect to missing all three exams (p-value = .031). Thus, it was clear that a significantly greater number of students in the unproctored Group 2 could not work more independently in a less structured exam environment and just "gave up" or "forgot" to take the test at the stipulated time periods.

Given that students in both groups were amply coaxed to complete online test preparation training tutorials on all three topics as preparation for the real exams, Hypothesis 4c indicates the proportion of students complying with this suggestion over the two groups. In comparing differences in the proportions between the two groups of students attempting training prior to their exams it was noted that for the Word/Powerpoint exam the differences were due only to chance (p-value = .162) but in the Excel and Access exams the differences were very highly significant with respective p-values of .006 and .000. This indicates that over the duration of the course there was greater "buy-in" with the proctored group and, in fact, the differences between the two study groups in the proportions of students attempting training magnified over time.

From Hypothesis 5, there is evidence that Group 1 students have a higher mean overall course performance average than their Group 2 counterparts (p-value = .002). There is, however, more variability in the overall course performance averages in Group 2 (p-value = .000) (Levene, 1960). Given that overall performance is a weighted average of exams, case studies, and homework assignments, this finding is not surprising because a greater number of Group 2 students "quit" (or "forgot") and received zero exam, homework, and case score grades for being absent. Such grades would pull down the Group 2 mean overall performance score and at the same time increase the variability in the distribution of overall performance scores.

DEFINING THE "PERFORMING" STUDY GROUPS

Although some substantial significant differences in the two study groups were observed, these findings should be viewed as global and do not pertain to those members of the groups that "performed" throughout the semester by taking all three exams and submitting homework and case study assignments. It is understandable that in a hybrid course there can be less "buy-in" from some students. Students who prefer or need structure would likely perform best in a traditional learning environment. Those who like to be challenged, have self-discipline, and like to work independently might flourish in a hybrid course atmosphere. If we filter out the "quitters" and compare the two "performing" study groups, that is, operationally defined as those students who took all three exams and thereby demonstrated a real effort to succeed in the course, we could better assess the effect that the two test-taking environments had on performance and draw more appropriate conclusions with respect to pedagogy and learning.

In these newly defined "performing" study groups there were 90 students in the proctored group (Group 1) and 83 students in the unproctored group (Group 2). Within each of these groups a correlation matrix was obtained to observe the bivariate Pearsonian product moment associations among the three main numerical outcome variables (i.e., the Word/PowerPoint, Excel, and Access exams), and other numerical outcome measures such as overall course performance average, exam average, Word/PowerPoint homework and case average, Word/PowerPoint case study grade, Excel homework and case average, Excel case study grade, Access homework and case average, Access case study grade, and the number of missed assignments. Also included in the correlation matrix were the two perspective numerical covariates, GPA, and number of credits completed. The third potential covariate, the categorical variable transfer status, was included along with three other dichotomous measures corresponding to whether or not students attempted each of the three online training tutorials. The pairwise associations between dichotomous categorical variables and numerical variables, referred to as point-biserial correlations, are equivalent to their respective Pearson product moment correlations. The pairwise correlations between dichotomous categorical variables, called the *phi* coefficients, are also equal to the Pearsonian product moment correlations between those variables (Glass & Stanley, 1967).

Given that a correlation matrix is a square symmetric matrix with all autocorrelations along the main diagonal equal to 1 and all the pairwise Pearsonian product moment correlations listed above the main diagonal equal to their corresponding row and column placements below the main diagonal, either half of the information in the matrix is superfluous. Therefore, in Table 2 the half correlation matrix below the main diagonal shows the aforementioned pairwise correlations for the 90 students in the proctored group (Group 1) while the half correlation matrix above the main diagonal shows the corresponding pairwise correlations for the 83 students in the unproctored group (Group 2). In this table, pairwise correlations that are significant at the .05 alpha level (two-tail) are indicated with an asterisk (*) and those significant at the .01 level are presented with a double asterisk (**).

From Table 2 we observe that GPA is significantly correlated with several outcome measures in each of the study groups and we know from Hypothesis 1

Table 2: Comparison of correlation matrices for group 1 and group 2 (group 1 below main diagonal, group 2 above).

	Group 2																	
	Word Exam	Excel Exam	Access Exam	Course Average	Exam Average	Word Hwk Case Average	Word Case	Excel Hwk Case Average	Excel Case	Access Hwk Case Average	Access Case	No. Missed Hwk/ Case	GPI	Credits Completed	Transfer	Word Training Attempted	Excel Training Attempted	Access Training Attempted
Group 1																		
Word exam		.317**	.308**	.428**	.692**	.269*	.288**	.225*	.178	.189	.081	−.125	.233*	.034	−.067	.120	.167	.249*
Excel exam	.452**		.334**	.363**	.789**	.049	.009	.289**	.207	.153	.036	−.081	.164	.148	.181	.250*	.281**	.316**
Access exam	.317**	.517**		.427**	.731**	.160	.182	.186	.202	.380**	.254*	−.086	.181	.209	.181	.136	.255*	.335**
Course average	.280**	.577**	.414**		.543**	.703**	.652**	.839**	.799**	.748**	.707**	−.762**	.603**	.106	.126	.237*	.248*	.293**
Exam average	.745**	.855**	.759**	.545**		.201	.197	.320**	.265*	.320**	.161	−.129	.255*	.181	.146	.236*	.323**	.408**
Word hwk case average	.040	.105	−.084	.643**	.033		.467**	.545**	.426**	.502**	.503**	−.850**	.481**	.013	−.059	.191	.096	.168
Word case	−.095	.080	.074	.578**	.027	.609**		.377**	.403**	.337**	.389**	−.432**	.494**	.158	.074	.147	.120	.162
Excel hwk case average	.062	.377**	.092	.790**	.238*	.510**	.315**		.869**	.590**	.491**	−.781**	.498**	−.041	.063	.139	.145	.148
Excel case	.104	.298**	.162	.691**	.245*	.348**	.235*	.814**		.575**	.577**	−.569**	.468**	−.059	.003	.095	.135	.098
Access hwk case average	.028	.227*	.309**	.388**	.238*	.089	.163	.007	−.022		.825**	−.629**	.404**	.119	.180	.128	.149	.216
Access case	−.043	.116	.293**	.391**	.150	−.005	.070	.067	.103	.717**		−.540**	.377**	.136	.162	.139	.164	.144
No. missed hwk/case	.035	−.183	.056	−.702**	−.050	−.870**	−.495**	−.732**	−.398**	−.150	−.054		−.476**	.075	−.031	−.159	−.056	−.131
GPI	.197	.385**	.360**	.618**	.401**	.380**	.311**	.397**	.328**	.373**	.371**	−.403**		.208	.278*	.329**	.313**	.370**
Credits completed	.050	.036	.049	.192	.057	.084	.071	.231*	.191	−.012	.101	−.103	.057		.499**	.200	.293**	.305**
Transfer	.005	−.097	−.066	−.011	−.069	.056	−.039	.077	.154	−.034	−.163	−.050	−.045	.297**		.308**	.313**	.400**
Word training attempted	.277**	.467**	.250*	.364**	.430**	.189	.262*	.186	.145	.151	−.011	−.173	.220*	.053	−.022		.628**	.623**
Excel training attempted	.192	.412**	.214*	.267*	.355**	.145	.142	.113	.066	.170	.025	−.127	.133	.077	.027	.800**		.786**
Access training attempted	.150	.403**	.358**	.298**	.389**	.161	.129	.168	.124	.039	.023	−.156	.147	.117	−.025	.652**	.638**	

*Correlation is significant at the .05 level (2-tailed).

**Correlation is significant at the .01 level (2-tailed).

that GPA is not significantly related to the two groups. These characteristics, correlation with a response variable and no significant association with a predictor variable make GPA a good covariate for the study (Berenson, Levine, & Goldstein, 1983). The same is not true for credits earned and transfer status. From Hypothesis 2 credits earned is significantly associated with the two study groups but not significantly correlated with a response variable. The dichotomous variable transfer status did not correlate well with the study groups and did not show any significant relationship with a response variable.

Based on the above, several research hypotheses were developed.

Questions for Establishing Hypotheses for Testing in the "Performing" Study Groups

6. Given that both sections are offered in a hybrid format that differs from the more traditional formally structured classroom environment with which students are familiar, using GPA as a covariate, will there be a difference in course "buy-in" between Groups 1 and 2 as reflected by:
   - Number of missed assignments (homework and case studies)?
   - Performance on the three exams based on whether the students attempted online test preparation training tutorials?
7. Given that the overall course performance resulting in a letter grade of A to F is based on a weighted average of the 3 exams, 3 case studies, and 11 homework assignments, using GPA as a covariate will there be a difference between Groups 1 and 2 with respect to overall course average?
8. Using GPA as a covariate, do students in Group 1 (i.e., proctored group) have a different mean average test grade over the three exams than students in Group 2 (i.e., unproctored group)?

Descriptive and Inferential Evaluation of "Performing" Study Groups

Key descriptive statistical summary measures along with the results of the corresponding tests of hypotheses pertaining to the above questions 6 through 8 are displayed in Table 3.

From Table 3 we note that analysis of covariance (ANACOVA) was employed to test various null hypotheses of no differences in the means between the two performing study groups while controlling for the effects of the covariate GPA.

Course "buy-in" was evaluated in several ways. First, from Hypothesis 6a there was no evidence of a difference in the adjusted mean number of missed assignments (i.e., homework and case studies) by students in Group 1 (i.e., proctored group) versus Group 2 (i.e., unproctored group). The p-value was .430.

On the other hand, for the six subhypotheses of Hypothesis 6b, where ANACOVA was used on each group for each of the three exams to determine whether controlling for GPA would result in a difference in the adjusted mean test performance based on whether or not the students attempted the corresponding online test preparation training tutorials, five of the six null hypotheses were rejected. Significantly higher adjusted mean test performances existed for students in Group 1 if they attempted the online test preparation training tutorials than if they did not make such attempt. The p-values for the Word/PowerPoint, Excel, and Access

Table 3: Summary of results of analysis of specific research hypotheses 6–8 for performing study groups.

Test	Specific Research Hypotheses	Groups and (Sample Sizes)	Summary Statistics	Tests	p-Values
6a	Course "buy in": number of missed assignments by group	Proctored (90) vs. unproctored (83)	$\overline{X_1} = .933, S_1 = 1.841$ $\overline{X_2} = 1.337, S_2 = 2.091$	ANACOVA	.430
6b	Course "buy in": exam average by group by training attempted				
	• Word exam proctored group	Training attempted (42) vs. not attempted (48)	$\overline{X_1} = 85.62, S_1 = 5.587$ $\overline{X_2} = 80.96, S_2 = 9.854$	ANACOVA	.021*
	• Word exam unproctored group	Training attempted (34) vs. not attempted (49)	$\overline{X_1} = 84.97, S_1 = 10.498$ $\overline{X_2} = 82.78, S_2 = 7.840$	ANACOVA	.672
	• Excel exam proctored group	Training attempted (39) vs. not attempted (51)	$\overline{X_1} = 89.15, S_1 = 6.869$ $\overline{X_2} = 81.18, S_2 = 10.095$	ANACOVA	.000**
	• Excel exam unproctored group	Training attempted (25) vs. not attempted (58)	$\overline{X_1} = 90.04, S_1 = 8.691$ $\overline{X_2} = 82.79, S_2 = 12.470$	ANACOVA	.026*
	• Access exam proctored group	Training attempted (48) vs. not attempted (42)	$\overline{X_1} = 88.92, S_1 = 7.437$ $\overline{X_2} = 83.24, S_2 = 7.509$	ANACOVA	.002**
	• Access exam unproctored group	Training attempted (29) vs. not attempted (54)	$\overline{X_1} = 92.69, S_1 = 6.767$ $\overline{X_2} = 85.70, S_2 = 10.643$	ANACOVA	.007**
7	Overall course average by group	Proctored (90) vs. unproctored (83)	$\overline{X_1} = .869, S_1 = .076$ $\overline{X_2} = .847, S_2 = .109$	ANACOVA	.380
8	Exam average	Proctored (90) vs. unproctored (83)	$\overline{X_1} = 84.678, S_1 = 6.857$ $\overline{X_2} = 85.598, S_2 = 7.647$	ANACOVA	.180

*Results significant at the .05 level.

**Results significant at the .01 level.

exams for Group 1 were, respectively, .021, .000, and .002. Similar findings occur for students in Group 2 taking the Excel and Access exams. The respective *p*-values were .026 and .007. Controlling for GPA, the only difference in adjusted mean test performance for those who attempted the preparatory/preliminary test and those who did not that was not significant at a .05 level were Group 2 students taking the Word/PowerPoint exam. The *p*-value was .672. Overall, the results demonstrate the importance of course "buy-in" and for each test a greater percentage of online training tutorials were attempted by students in Group 1.

The overall course performance averages for the students in the two groups were compared in Hypothesis 7. The overall course performance averages are weighted averages of 3 course examinations, 3 case studies, and 11 graded homework assignments. From a descriptive perspective, while there did not appear to be real differences in the mean overall course performance averages, Group 1 displayed more homogeneity in the results. The greater variability in Group 2 student performance, observed in the overall study groups, was seen here again in the "performing" study groups.

From a confirmatory perspective, ANACOVA was employed to test the null hypothesis of no differences in mean overall course performance by students in Group 1 (i.e., proctored group) and in Group 2 (i.e., unproctored group). Controlling for the effects of the covariate GPA, the null hypothesis could not be rejected at the .05 level of significance. The *p*-value was .380. There was no evidence of a difference in adjusted mean overall course performance by students in the two groups.

Given that everyone in the "performing" study groups took all three exams—the main outcome measures of this research—but may not have completed all 3 case studies and/or all 11 homework assignments, it was important to focus on the average test scores attained by the students in each group. Thus, for Hypothesis 8, ANACOVA was used to test the null hypothesis of no differences in mean average test grade over the three exams by students in the proctored group and in the unproctored group. Controlling for the effects of the covariate GPA, the null hypothesis could not be rejected at the .05 level of significance. The *p*-value was .180. There was no evidence of a difference in adjusted mean average test grades by students in the two groups.

Extending the Analysis: A Multivariate Approach for Comparing the Two "Performing" Study Groups

Given the multivariate nature of the key numerical outcome variables and having GPA as a covariate, it was important to investigate whether there was evidence of a difference among the three exam scores within and across the study groups and whether there were significant interaction effects between the two study groups and the sequence of the three exams (i.e., Word/Powerpoint, Excel, and Access). Specifically:

9. Is there evidence of a difference in the vector of mean exam scores between the two study groups? That is, does one group outperform the other?
10. Is there evidence of a difference in the effects—is there significant interaction in student performance over the three exams in the two study

Figure 1: Comparison of mean exam scores by group.

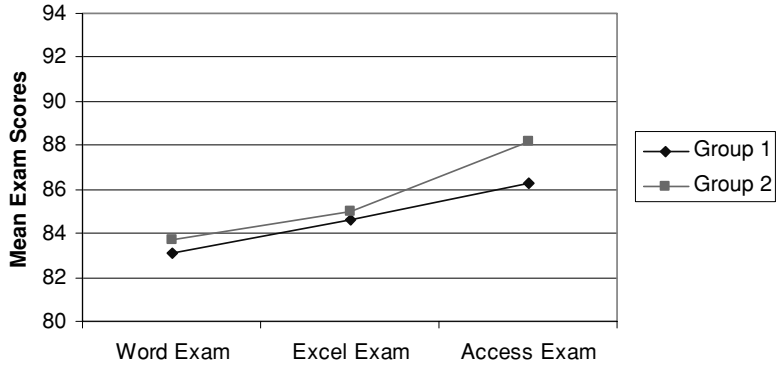
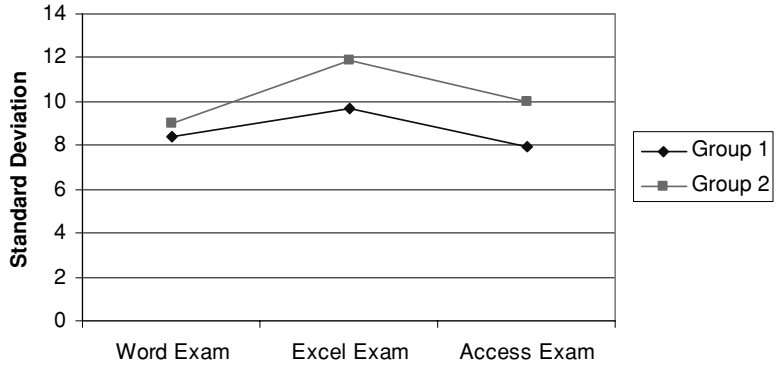


Figure 2: Comparison of standard deviation of exam scores by group.



groups or are differences in the mean exam scores between the two study groups similar or proportional over the three exams?

11. Is there evidence of a difference in the mean exam scores within the two study groups? That is, do students in each of the study groups perform significantly better (or worse) on at least one of the exams?

Figure 1 is a plot of the mean test scores over the three exams for each group. Figure 2 is the corresponding plot of the standard deviations.

A comparison of Figures 1 and 2 leaves the impression that there were much greater differences in average test performance across the three exams than there were within the two groups. Moreover, it appears that there were greater differences in the variability around the respective group means over the three examinations than there were just in the variability in the means of the two groups over the three examinations.

Given the multivariate nature of the key response variables, methodology employing individual inferential analysis was inappropriate (Hotelling, 1931, 1951;

Lattin, Carroll, & Green, 2003; Manly, 1986). Therefore, a repeated measures design containing a multivariate analysis of covariance (MANACOVA) was used (ACITS, 1997).

Although the repeated measures design is considered robust with respect to assumptions of multivariate normality and variability in the outcome measures across groups, the methodology is unreliable if the design is not "balanced," that is, if the sample sizes in the study groups are not equal. Given that the "performing" study groups are unbalanced, 90 students in the proctored group and 83 students in the unproctored group, four approaches were taken to achieve balance and the repeated measures design was then used on each. A sensitivity analysis evaluating the various results would be expected to uncover any strong differences in performance within and between the two "performing" study groups if such differences exist.

The four approaches taken to achieve balance in the repeated measures design and used for a sensitivity analysis of the results were:

Approach 1: The seven proctored students with the highest GPA covariate were removed and a balanced repeated measures design analysis on 166 subjects was performed.

Approach 2: The seven proctored students with the lowest GPA covariate were removed and a balanced repeated measures design analysis on 166 subjects was performed.

Approach 3: Seven proctored students were removed through the use of a table of random numbers and a balanced repeated measures design analysis on 166 subjects was performed.

Approach 4: Regression analysis of overall exam average with GPA in Group 1 was used to remove the seven proctored students with the largest residual/influential effects and then a balanced repeated measures design analysis on 166 subjects was performed.

Table 4 is a summary of the results from the four repeated measure design approaches. A sensitivity analysis showed no real differences across the four approaches so that the conclusions given below hold.

For Hypothesis 9, for each of the four approaches for obtaining a balanced repeated measures design containing MANACOVA, controlling for the effects of the covariate GPA, the null hypothesis of no difference in the vectors of mean exam scores between the two study groups could not be rejected at the .05 level of significance. There was no evidence of a "between subjects main effect" regarding the "performing" study groups—that is, there was no evidence of any real differences in student performance between Groups 1 and 2.

For Hypothesis 10, for each of the four approaches for obtaining a balanced repeated measures design containing MANACOVA, controlling for the effects of the covariate GPA, the null hypothesis of no interaction effect was not rejected at the .05 level of significance. There was no evidence of a "between subjects by within subjects interaction effect" of "performing" study groups and exam topic over time. Thus, there was no influence of class section on exam performance on

Table 4: A comparison of results from four approaches for obtaining repeated measures balanced designs.

Key Statistics	Approach 1: Remove 7 Gp. 1 Subjects with Highest GPA	Approach 2: Remove 7 Gp. 1 Subjects with Lowest GPA	Approach 3: Remove 7 Gp. 1 Subjects Randomly	Approach 4: Remove 7 Gp. 1 Subjects Based on Regression
Balanced study group sample sizes	83 and 83	83 and 83	83 and 83	83 and 83
Gp.1 covariate GPA mean and *SD*	3.041; .388	3.172; .366	3.117; .425	3.109; .429
Gp.2 covariate GPA mean and *SD*	2.995; .548	2.995; .548	2.995; .548	2.995; .548
Gp.1 Word/Ppt Mean and *SD*	82.904; 8.619	83.711; 8.016	83.590; 7.621	83.771; 6.885
Gp.2 Word/Ppt Mean and *SD*	83.675; 9.029	83.675; 9.029	83.675; 9.029	83.675; 9.029
Gp.1 Word/Ppt adjusted mean	82.815	83.430	83.349	83.534
Gp.2 Word/Ppt adjusted mean	83.763	83.956	83.916	83.912
Gp.1 Excel Mean and *SD*	84.205; 9.728	85.614; 8.960	84.783; 9.507	85.289; 8.672
Gp.2 Excel Mean and *SD*	84.976; 11.891	84.976; 11.891	84.976; 11.891	84.976; 11.891
Gp.1 Excel adjusted mean	84.078	85.232	84.448	84.954
Gp.2 Excel adjusted mean	85.102	85.358	85.311	85.311
Gp.1 access mean and *SD*	85.976; 8.009	86.940; 7.602	86.313; 7.979	86.651; 7.555
Gp.2 access mean and *SD*	88.145; 10.004	88.145; 10.004	88.145; 10.004	88.145; 10.004
Gp.1 access adjusted mean	85.869	86.592	86.017	86.387
Gp.2 access adjusted mean	88.252	88.493	88.440	88.408
p-Values via MANACOVA				
Between subjects main effects:				
• Group 1 vs. group 2	.182	.431	.227	.356
• Covariate GPA	.000**	.001**	.000**	.000**
Mauchly's Sphericity test	.747	.752	.389	.720
Within subjects main effects:				
• Interaction: group and exam topic	.634	.528	.499	.492
• Exam topic	.000**	.000**	.000**	.000**

*Results significant at the .05 level.

**Results significant at the .01 level.

the topics learned over time. The mean exam scores of the unproctored group did not significantly increase or change in a different pattern than did the mean exam scores of the proctored group as topics changed over time.

For Hypothesis 11, for each of the four approaches for obtaining a balanced repeated measures design containing MANACOVA, controlling for the effects of the covariate GPA, the null hypothesis of no difference in means among the three exams over the two "performing" study groups was rejected at the .05 level of significance. There was evidence of a "within subjects main effect" regarding the three exam topics (Word/PowerPoint, Excel, and Access).

Using this latter result and returning to the proctored group containing 90 students, based on Tukey's *a posteriori* comparisons it was observed that the mean Access exam score significantly exceeded the mean Word/PowerPoint exam score but differences in the means between the Word/PowerPoint and Excel exams and between the Excel and Access exams were due to chance. In the unproctored group, Tukey's *a posteriori* comparisons led to the conclusion that the mean Access exam score significantly exceeded both the mean Word/PowerPoint exam score and the mean Excel exam score but that differences in the mean Word/PowerPoint exam score and mean Excel exam score were due to chance. The significant difference in mean Access exam score could be due to increased familiarity with the computer-based testing environment, an easier Access exam resulting from a smaller pool of test questions from which to choose when creating the exam, or higher levels of motivation on the part of students late in the semester to improve their final course average.

FINDINGS ON IMPACT OF EXAM ENVIRONMENT ON STUDENT PERFORMANCE

A summary of the descriptive statistics provided in Table 3 along with Figures 1 and 2 demonstrates greater variability in performance in the unproctored group when compared to the proctored group. Controlling for the covariate GPA, which itself indicated greater variability in Group 2 than in Group 1, enabled a comparison of sets of adjusted response variable means in the two "performing" study groups; but no evidence of any real difference was found.

The overall question that can now be raised is whether or not there is indication of significantly greater variability in the set of performance measures by students in the unproctored group than in the proctored group.

Van Valen's test is the appropriate multivariate procedure to use to assess whether the vector of variances in one group exceeds the vector of variances in another group (Manly, 1986; Van Valen, 1978). It is most powerful in detecting significant differences in overall variability when each component of the vector in one group exceeds that of the other group. Controlling for the effects of the covariate GPA, the null hypothesis of equality of the vector of variances in the exam scores over the two study groups was rejected at the .05 level of significance (one-tailed). The p-value is .029.

Given that the main differences between the performance measures from the two groups were in overall variability, not central tendency, possible reasons for

such demonstrated effects must be considered. In the initial overall analysis containing 107 students in each of the study groups it would not have been surprising to find greater variability in the performance measures from the unproctored group. A significantly greater percentage of students in that group "gave up," did not take all of their exams, did not submit all of their other course assignments, and thus received grades of zeros on those exams and assignments. In these study groups of "performers," each of the 90 students in Group 1 and 83 students in Group 2 took all three exams.

Three possible reasons for observing differences in variability in performance between the two study groups are offered:

1. Environmental differences in taking the three exams.
2. Familiarity with and/or preference for particular learning styles.
3. The opportunity for cheating when taking the three exams.

In performing this research endeavor the only intended difference that was designed was the fact that one of the groups of students would be given their three online exams in a traditional computer lab classroom setting during a scheduled session with the instructor and proctors (graduate assistants) available while the other group of students would be taking these online exams on their own outside a formal classroom setting within a scheduled block of time. Once launching the exams, each group was provided the same amount of time (i.e., 60 minutes) in which to complete each exam.

Environmental differences in taking the three exams result in one potential reason for observing greater variability in performance in Group 2. Outside the classroom setting, students in Group 2 could have experienced problems with their computer or network connection, with extraneous noise factors, or other issues affecting concentration and, unlike their Group 1 counterparts, they could not easily access the instructor or the graduate students to help resolve these issues. Some of these same discouraging factors could further impact on course "buy in." In retrospect, unless more controls are added into a future research project little could be done to circumvent such issues.

Familiarity with and/or preference for particular learning styles (Campbell, 1991; Claxton & Murrell, 1987; Coffield, Moseley, Hall, & Ecclestone, 2004a, 2004b) is a second reason for potential differences in variability in performance. Typically, students are used to a more structured learning and testing environment. While students in both sections of the course were taught using the hybrid course model, students in the proctored group were afforded more structure in their testing environment that may have impacted their overall course "buy in" and potentially result in greater performance variability in the unproctored group. Future follow-up research could potentially control for this by employing a battery of tests regarding learning style preference (Hawk & Shah, 2007) and using the results as a covariate or even funneling students into particular class sections based on their learning style preferences and desire for more structured pedagogical approaches.

A third reason for observing greater variability in performance in the unproctored group is the greater opportunity for cheating to occur when students are assigned to take the exams "on their own," wherever they choose to do so,

during a variety of stipulated time periods (Sadler, 2007). Examining the results from the two "performing" study groups, with or without controlling for differences in GPA, it was surprising to note that on average Group 2 students did as well as Group 1 students on each of the three exams even though a significantly greater percentage of Group 1 students completed all their homework and case study assignments and, under the stress of the examinations, had the opportunity to have their anxieties immediately removed or reduced by asking questions from the instructor or proctor. Perhaps, then, a key reason for observing similarity in central tendency in exam performance between the two study groups coupled with significantly greater variability in performance in Group 2 is a result of students' academic integrity. There are far greater opportunities for cheating in Group 2. Efforts to uncover possible patterns of cheating in this study are discussed in the following section.

ASSESSMENT OF ACADEMIC INTEGRITY IN "PERFORMING" STUDY GROUPS

Given there was no discernable indications of cheating in Group 1 during the examination proctoring process, focus was placed on looking for potential cheating patterns in Group 2 where the online examinations were taken in uncontrolled environments. None of the 83 students in the unproctored group flagged based on any unsuspected behavior patterns. For example, those students in Group 2 who were observed to have scored very high on the exams did not have very low GPAs, nor did they fail to submit homework and case study assignments or not attempt taking the online training tutorials. Their behavior was consistent with their exam performance.

Four approaches were taken to find differences in the two "performing" study groups that may signal cheating in Group 2 where students took their online exams in unproctored environments.

1: The Word/PowerPoint exam scores and the corresponding homework/case average averages were used as "baselines" to develop an ANACOVA to measure the difference in the average of the Excel and Access exam scores with the Word/PowerPoint exam score "baseline" as a function of the study group (proctored vs. unproctored) and the covariate (the difference in the average of the Excel and Access homework/case averages with the Word/PowerPoint homework/case average "baseline.") The *p*-value is .833 and no significant difference was observed in the adjusted means between the two groups. If substantial cheating were to exist in Group 2, say, we would expect a significant difference in the adjusted means.

2: Tables 5 and 6 are two-way tables for each "performing" study group where the two rows represent "increase" versus "no increase" in the average of the Excel and Access homework/case averages over the Word/PowerPoint homework/case average "baseline" and the two columns represent "increase" versus "no increase" in the average of the

Table 5: Proctored group 1—changes in performance trend.

	Excel/Access Exam vs. Word Exam		
	Decreasing Exam	Increasing Exam	Grand Total
Excel/Access Hwk vs.Word Hwk			
Decreasing Hwk	30	50	80
Increasing Hwk	5	5	10
Grand total	35	55	90

Table 6: Unproctored group 2—changes in performance trend.

	Excel/Access Exam vs. Word Exam		
	Decreasing Exam	Increasing Exam	Grand Total
Excel/Access Hwk vs. Word Hwk			
Decreasing Hwk	30	40	70
Increasing Hwk	5	8	13
Grand total	35	48	83

Excel and Access exam scores over the Word/PowerPoint exam score "baseline." McNemar's test for significance of changes (Berenson & Koppel, 2007) was then applied to the results in both tables and the two obtained McNemar statistics were compared to each other to determine if there were differences in the two groups.

From the Group 1 table it is observed that 55 of 90 students displayed "inconsistent behavior"—50 students had increasing exam scores with decreasing homework/case averages while only 5 students had decreasing exam scores coupled with increasing homework/case averages. The *p*-value for the McNemar test measuring the significance in performance changes in Group 1 was .000. Similarly, from the Group 2 table it is seen that 45 of the 83 students displayed "inconsistent behavior." From this "inconsistent" group of students, 40 had declining homework/case averages in conjunction with increasing exam scores and 5 had increasing homework/case averages along with declining exam scores. The *p*-value for the McNemar test measuring the significance in performance changes in Group 2 was also .000.

One might expect such "inconsistent behavior" from a group in which prevalent cheating is occurring but one could not possibly expect such behavior in Group 1 where the exams were monitored and cheating is likely to be minimal. One should have expected a higher level of "consistent" performance in Group 1.

It must be stressed that information is lost by dichotomizing differences in numerical variables as "increasing" or "not increasing," as shown above. Nevertheless, it is clear from Tables 5 and 6 that there were no differences in performance across the two groups. Perhaps the conclusion to be reached here is that the homework and case studies did not prepare and enhance exam performance on the corresponding topics of Word/PowerPoint, Excel, and Access.

Table 7: Wolfe *rho* correlation coefficient statistics for each group.

Comparison	Gp 1 *rho*	Gp 2 *rho*	*Z*	*p* Value (One-Tail)
Word/Ppt exam	+ .009	−.023	+ 0.21	.4168
Excel exam	−.070	+.003	+ 0.47	.3192
Access exam	−.037	+.072	+ 0.70	.2420

A possible reason for this is the differing assessment goals of exams versus case projects and homework. Homework is designed for practice, exams are designed to assess students' abilities to perform skills in various software programs, and cases are designed to assess students' abilities to apply software skills in the solution of business problems. Students who demonstrated competency with skills were not always able to demonstrate the same level of mastery in the applied cases.

3: Using Williams' (1959) *t* test for trivariate relationships we can separately study the correlations between each of the three exam scores and the corresponding homework averages and case study grades for both Groups 1 and 2. If substantial cheating were to exist in Group 2, say, we would expect Group 1 to have much higher correlations between exam score and the corresponding homework average and case study that serve as proxies for exam preparation.

None of the six Williams' *t* tests that were conducted on data transformed to expected normal scores in order to enhance the power of detecting correlation effects (Boyer, Palachek, & Schucany, 1983) were statistically significant at the .05 level (one-tail).

We can conclude that for each group, for each of the three exams, neither homework average nor case study project correlated more significantly with exam performance and neither is a significantly better preparatory predictor of exam result.

The Williams' *t*-statistic can be reformulated into Wolfe's (1976) *rho* correlation coefficient statistic. Three *Z* tests for the differences in the corresponding correlations from the two independent groups (Glass & Stanley, 1967) would then signal possible cheating in Group 2.

The six Wolfe *rho* correlation coefficient statistics for the relationship of exam score with homework average and case study grade were obtained and the results of the three *Z* tests are provided in Table 7.

For each of the three exams, we can conclude there were no significant differences between the two groups with respect to the Wolfe *rho* correlation coefficients. Again, no evidence of cheating was found.

4: Using Schucany's L^* statistic we may ascertain whether or not there was concordance both within and across the study groups with respect to student performance on the three exams (Beckett & Schucany, 1979; Li & Schucany, 1975). If, for example, students in Group 2 realize over the

course of the semester that they can take advantage of the unproctored exam situation and choose to cheat we should expect significant increases in exam scores over what might occur in Group 1. A positive and statistically significant L^* statistic would indicate that performance over the three tests were in agreement within the two groups and between the two groups. A negative and statistically significant L^* statistic would indicate concordance within the two groups but disagreement in the direction between them. Cheating explanations would be considered. If the L^* statistic is not significant it is an indication of discordance both within and across the groups.

In comparing Groups 1 and 2, an L^* statistic of $+3.83$ was obtained and this is statistically significant at the .05 level. Its *p*-value was .000. We may conclude that there was a significant amount of concordance in the students' exam performance both within the groups and across the groups. There was no indication of any cheating pattern.

Using the four different approaches no patterns of cheating were uncovered. It may be concluded that there was no discernable evidence of cheating.

CONCLUSIONS AND IMPLICATIONS FOR FUTURE RESEARCH

Summary

In this article, we investigate the effect of a change in exam environment on student performance in an undergraduate information systems course. A thorough statistical analysis was undertaken to uncover differences in student performance between the group taking the online exams in a proctored environment (i.e., in-class) versus the group taking exams in an unproctored environment (i.e., offsite).

Controlling for student GPA, no differences in central tendency of performance measures were found. There were no significant differences in mean overall course performance or exam performance between the two groups nor were there any differences in the mean vectors of individual exam scores. The study did reveal that the group taking exams in the unproctored environment did have significantly more variation in their performance results. In examining potential causes of the greater variation, analyses were performed to assess whether an increased level of possible cheating behavior could be observed from performance results for students in the unproctored section. No evidence of cheating behavior was found.

As discussed in the Introduction, the expectation of increased levels of cheating when course assessments are given outside the classroom is seen as a negative element in offering distance and hybrid courses. This research points in the direction that an off-site test environment does not necessarily result in a significantly increased level of cheating. However, the results in this study are not applicable to all forms of exam assessment. The exams administered in this course are hands-on, activity-based exams where students must perform tasks in a simulated computer environment; exams are graded automatically by computer. We see differences in the groups on the range of performances but not in their mean scores.

Future Research

We are interested in extending this research to study and evaluate multiple types of course assessments (i.e., multiple choice exams, short answer exams, mathematical computation exams, individual computation-based exams with each student having a different set of data) as well as to study the effect that exam environment has on student performance. The expectation would be that some forms of assessment are more prone to student cheating in an unproctored setting.

Another avenue for future research is a detailed study of the "buy-in" effect for students in hybrid and distance-based courses. Although the only changed element in this study was exam location, this one structural element seemed to impact "buy-in" in the overall study groups.

In the section above, we identify possible reasons for increased variation in the unproctored group. In this study we did not have data to determine specific exam environment factors that may contribute to greater variation in student performance. A third direction for future research is to quantify factors in exam environment that contribute to variation in student performance.

Another path for future work will be to study the impact of student learning style on performance in hybrid courses. We are interested in answering three overall questions:

1. How does learning style affect performance in hybrid courses?
2. What is the impact on student performance when students are provided a course map based on their preassessed learning style?
3. Can student performance be improved by placing students into course sections (i.e., traditional, hybrid) based on preassessed learning style?

REFERENCES

ACITS. (July 31, 1997). Repeated Measures ANOVA Using SPSS Manova.

Alexander, M. W., Truell, A. D., & Bartlett, J. E. (2002). Students' perceptions of online testing. *The Delta Pi Epsilon Journal, 44*(1), 59–68.

Anstine, J., & Skidmore, M. (2005). A small sample study of traditional and online courses with sample selection adjustment. *Journal of Economic Education, 36*(2), 107–128.

Becker, K., Kehoe, J., & Tennent, B. (2007). Impact of personalized learning styles on online delivery and assessment. *Campus-Wide Information Systems, 24*(2), 105–119.

Beckett, J., & Schucany, W. R. (1979). Concordance among categorized groups of judges. *Journal of Educational Statistics, 4*, 125–137.

Berenson, M. L., & Koppel, N. B. (2007). McNemar test for significance of changes. In N. J. Salkind (Ed.), *The encyclopedia of measurement and statistics, vol. 2* (pp. 576–580). Thousand Oaks, CA: Sage.

Berenson, M. L., Levine, D. M., & Goldstein, M. (1983). *Intermediate statistical methods and applications: A computer package approach.* Englewood Cliffs, NJ: Prentice-Hall.

Boyer, J. E., Palachek, A. D., & Schucany, W. R. (1983). An empirical study of related correlation coefficients. *Journal of Educational Statistics, 8*, 75–86.

Bryant, S. M., Kahle, J. B., & Schafer, B. A. (2005). Distance education: A review of the contemporary literature. *Issues in Accounting Education, 20*(3), 255–273.

Byrd, B., & Lott, K. (2003). Evaluation in online courses. *Business Education Forum, 58*(1), 48–50.

Campbell, B. J. (1991). Planning for a student learning style. *Journal of Education for Business, 66*(6), 356–359.

Claxton, C. S., & Murrell, P. H. (1987). *Learning styles*. Washington, DC: George Washington University (ERIC).

Coffield, F. J., Moseley, D. V., Hall, E., & Ecclestone, K. (2004a). *Learning styles and pedagogy in post-16 learning: A systematic and critical review*. London: Learning and Skills Research Centre.

Coffield, F. J., Moseley, D. V., Hall, E., & Ecclestone, K. (2004b). *Learning styles: What research has to say to practice*. London: Learning and Skills Research Centre.

Gaytan, J. (2005). Effective assessment techniques for online instruction. *Information Technology, Learning, and Performance Journal, 23*(1), 25–33.

Glass, G. V., & Stanley J. C. (1967). *Statistical methods in education and psychology*. Englewood Cliffs, NJ: Prentice-Hall.

Hawk, T. F., & Shah, A. J. (2007). Using learning style instruments to enhance student learning. *Decision Sciences Journal of Innovative Education, 5*(1), 1–19.

Hotelling, H. (1931). The generalization of Student's ratio. *Annals of Mathematical Statistics, 2*, 360–378.

Hotelling, H. (1951). A generalized t-test and measure of multivariate dispersion. *Proceedings of the Second Berkeley Symposium of Mathematical Statistics and Probability, 2*, 23–41.

Koppel, N., & Hollister, K. (2004). Teaching business computer applications: Comparison between the traditional delivery and a hybrid approach. *Review of Business Research Journal, IV*(1), 95–101.

Lattin, J., Carroll, J. D., & Green, P. E. (2003). *Analyzing multivariate data*. Pacific Grove, CA: Brooks/Cole.

Levene, H. (1960). Robust tests for equality of variance. In I. Olkin, S. G. Ghurge, W. Hoeffding, W. G. Madow, & H. B. Mann (Eds.), *Contributions to probability and statistics* (pp. 278–292). Palo Alto, CA: Stanford University Press.

Li, L., & Schucany, W. R. (1975). Some properties of a test for concordance of two groups of rankings. *Biometrika, 62*, 417–423.

Liu, Y. (2006). Effects of online instruction vs. traditional instruction on students' learning. *International Journal of Instructional Technology & Distance Learning, 2*(3), 57–64.

Manly, B. F. J. (1986). *Multivariate statistical methods: A primer.* London: Chapman and Hall.

Peng, Z. (2007). Giving online quizzes in corporate finance and investments for a better use of seat time. *The Journal of Educators Online, 4*(2), 1–18.

Robles, M., & Braathen, S. (2002). Online assessment techniques. *The Delta Pi Epsilon Journal, 44*(1), 39–49.

Sadler, B. J. (2007). The wrongs of plagiarism: Ten quick arguments. *Teaching Philosophy, 30*(3) 283–291.

Scanlon, P. M. (2003). Student online plagiarism: How do we respond? *College Teaching, 51*(4), 161–165.

Van Valen, L. (1978). The statistics of variation. *Evolutionary Theory, 4*, 33–43.

Weber, J., & Lennon, R. (2007). Multi-course comparison of traditional versus Web-based course delivery systems. *The Journal of Educators Online, 4*(2), 1–19.

Williams, E. J. (1959). The comparison of regression variables. *Journal of the Royal Statistical Society, Series B, 21*, 396–399.

Wolfe, D. A. (1976). On testing equality of related correlation coefficients. *Biometrika, 63*, 214–215.

Kimberly Killmer Hollister is an interim associate dean of the School of Business at Montclair State University. She received her BSE, MSE, and PhD in Systems Engineering from the University of Pennylvania. Prior to joining Montclair State, she held a Lecturer position at the Wharton School of Business, University of Pennsylvania. She teaches in the areas of Operations Management and Management Information Systems. Her professional experience includes consulting in both the Management Information Systems area and in Systems Engineering. At Montclair State University, she has been chair of the Curriculum Committee and is a member of the Assessment Team in the School of Business. Her research focuses on educational program assessment and applied operations management. Recently, her work on assessment was presented to both the 2007 MAACBA and AACSB conferences. Her recent research has appeared in *European Journal of Operational Research, Journal of Informatics and Education Research, Journal of Accountancy, Journal of Business & Economic Studies, Review of Business Research, Journal of Academy of Business and Economics*, and the *Journal of Business and Economics Research*.

Mark L. Berenson is a professor of Management and Information Systems in the School of Business at Montclair State University. He is also Professor Emeritus of Statistics in the Zicklin School of Business, Baruch College, City University of New York (CUNY). He received his BA in Economic Statistics and his MBA in Business Statistics from CCNY, and his PhD in quantitative methods in business from CUNY. Over the years he has received numerous awards for teaching, innovative contributions to statistics education, and for service. At Montclair State he has been chair of the School of Business Learning Goals and Assessment Committee. He is a coauthor of 11 statistics texts published by Prentice Hall

and his research has appeared in *Decision Sciences Journal of Innovative Education, Review of Business Research, The American Statistician, Communications in Statistics, Psychometrika, Education and Psychological Measurement, Journal of Health Administration Research*, *Encyclopedia of Measurement and Statistics,* and in the *Encyclopedia of Statistical Sciences.*