

The Management Of Testing In Distance Learning Environments

Marian C., Schultz (E-mail: mschultz@uwf.edu), University of West Florida
James T. Schultz, (E-mail: james.schultz@erau.edu), Embry-Riddle Aeronautical University
James Gallogly, (E-mail: james.gallogly@erau.edu), Embry-Riddle Aeronautical University

ABSTRACT

In 2004 Embry-Riddle Aeronautical University transitioned from proctored examinations for distance learning courses to online examinations that are not proctored. The purpose of this study was to determine if there is a significant difference in the midterm and final examination grades between proctored and non-proctored online examinations. Three courses, MGMT 311, MGMT 210 and ASCI 254, were selected for the study,. The study found that there was a significant increase in the mean scores between the midterm and final examinations. What this study did not examine was the impact of the overall course grades. This needs to be considered in future studies.

INTRODUCTION

In 1995, only a third of higher learning institutions offered some form of virtual learning, with roughly 754,000 students' enrolled ("National Center", n.d.). Virtual learning in higher education has become almost as common as traditional face-to-face classes. In 2005, almost 65% of institutions that offer face-to-face undergraduate and graduate classes now also offer online courses, with enrollment at an all-time high of 2.35 million students. Virtual learning is an integral part of most institutions of higher learning. In fact, at 56% of those institutions, administration considers online learning as a critical long-term strategy (Sloan Consortium, 2005).

Numerous universities, including Embry-Riddle Aeronautical University, have instituted on-line testing in their Distance Learning programs. Previously, Embry-Riddle's procedure entailed the mailing of paper copies of both the midterm and final examinations from the Distance Learning Department, located in Daytona Beach, Florida, to testing facilities located in various countries throughout the world. The examinations would be sent to one of the approximately 100 Embry-Riddle Extended Campus center locations, where the student would go to complete the examination. For those students where there was no center located nearby, a pre-approved proctor would administer the examination. The completed examination was then mailed to the professor for grading. While this approach guaranteed the confidentiality of the examinations, while ensuring the right test taker; it also included numerous disadvantages. One of these disadvantages includes the delivery of the examinations through the mail system to the testing location, and then to the professor when completed. While generally dependable, the U.S. mail is not 100% reliable; there were numerous occasions when students or professor did not receive the examinations on time, and sometimes not at all. This resulted in having to forward additional examinations, as well as subsequent delays in processing the course grades. Frustration and dissatisfaction among the students and staff as well as professors was apparent in the Distance Learning program.

To remedy this situation, the Distance Learning Department transitioned to on-line testing in 2004. This format allows students to access the examinations on-line at a previously scheduled timeframe representing mid-course and end of the course. The scheduled exams were intended to occur at a convenient time, as well as at a time when the students were most likely prepared to be tested on the material. The students complete the examination in a single attempt with a predetermined time restriction. The online exam, without a proctor present in the testing room, suggests the possibility that students will have access to their textbooks, notes, and possibly other students while completing the examination. With the popularity of the laptop increasing, it is not impossible to identify identical IP addresses associated with consecutive exams being completed.

The purpose of this study was to determine if there is a significant difference in scores between examinations taken before and after on-line testing was initiated. If a difference does exist, and the grades are significantly higher after on-line testing was established, then perhaps there is a need to consider increasing the level of difficulty associated with the examinations by integrating application questions (those which are not directly accessible in the text, but must be derived from a compilation of information) or by addressing other variables, such as content access restrictions and time restrictions. If, however, the grades are significantly lower than they were before on-line testing was instituted, then perhaps the authors of the examinations had already considered this possibility, and compensated by making the examinations much more difficult. If this is the case, adjustments in on-line examinations may be appropriate. If there is not a significant difference in grades after online testing was initiated, then it can be surmised that the authors had made adjustments to the examinations in such a way to insure that answers to questions could not be easily accessed by the students.

ONLINE EDUCATION

There is limited literature pertaining directly to online proctored versus non-proctored testing. There have been many studies which have indicated that online students often outperform their traditional classroom counterparts (Schulman & Sims, 1999). However, many of the studies do not reveal how the examinations were administered, for example, did the distance learning students take the examination online while the traditional students took theirs in a proctored classroom environment. Another factor to contemplate is the enforcement of the time restriction for the exam and the integration of randomness throughout the exam choices, as well as the exam questions themselves. In a poll of college calculus students, 56.6% of students stated they would prefer online testing to 21.7% paper testing with 21.7% indicating no preference. Students' reasons for choosing online testing included: absence of examination stress, pressure resulting from other individuals in the same room, immediate feedback, flexibility, using modern technology, and no partial-credit. Professors, however, perceive that in accommodating the new-age student, standards have been sacrificed (Engelbrecht & Harding, 2004).

A study to determine the effectiveness of online learning was conducted at California State University at Northridge. This study evaluated the use of online testing in relation to traditional testing in to identify to most effective preference. Schutte (1997) found that students who were tested using an online process in a virtual environment scored 20 percent better across the board than the students who tested in a traditional background. Schutte split one of his classes into two sections: the first section would take the class in a traditional classroom setting with lectures and written assignments, and the second section would learn completely through online lectures, email, and electronic assignments. No major demographic differences existed between the two groups. The study found that students who were enrolled in the online version adapted quickly to the environment and even formed groups among themselves to help sort through the material. In fact, students who were in the online version spent roughly 50% more time working together than the students in the traditional classroom. Perhaps the team effort affected or brought about the higher testing scores. Virtual students may have to be more attentive and expend more time actually learning and communicating the material with each other than those in a traditional classroom (Schutte, 1997).

A study by Bocij and Greasley (2003) concerned the development and implementation of a general purpose computer-based assessment (CBA) system. More than 6000 individual computer-based assessments were recorded and observed. There were two stages to this study. The first stage was a formative test prior to an examination, a post-test questionnaire and a group discussion. The second was a "mock" test and a formal summative test. Questionnaires were developed for collection of data about the students to determine learning styles. The data collected were prepared for interpretation by the use of spreadsheets and statistical packages. The opinion of computer-based assessment went from neutral to favorable after the post-test. Students found computer-based tests to be less threatening and a more accurate means of testing a wider range of skills.

According to Parker and Gemino (2001) students performed equally as well in either place-based or virtual format classes. Their study found that there was no significant difference between the Asynchronous Learning Network (ANL) and place-based students in overall scores. The material in the course and the tests were the same for both formats. This gave the researchers a chance to observe the differences, if any, between the two types of learning.

The degree of learning differed however, between conceptual and technique-based material. The study was based upon a business administration course which was chosen because many of the students who enrolled in the class were co-operative education students who didn't want to spend time commuting. There were two sections offered per year, one place-based and one online. There were two objectives of the class. The conceptual part of the class was to understand the role of a systems analyst, while the second objective was technical and involved understanding diagrams and languages that a systems analyst would use.

The researchers studied the final exam grades over a period of five semesters. It was found that the group of students in the ALN course scored higher in the conceptual component than place-based students. They scored lower, however, in the technical component versus the place-based students. The researchers hypothesized that the higher conceptual scores for the online based classes could be due to the increased time allotted for discussions which was apparent once the classroom time constraints were eliminated. Overall, the study confirmed that traditional place-based classes and online classes provide similar outcomes when assessments are analyzed and statistically tested for differences (Parker & Gemino, 2001)

In a study by Johnson, Aragon, Shaik & Palma-Rivas (2000), it was found that online learning and place-based instruction were almost equal in learning outcomes. The main purpose of the study was to compare an online course with a traditional face-to-face course. The ratings included student perceptions of professors, course quality, course interaction, and learning outcomes. The study involved two versions of a graduate level instructional design course. One was online, while the other was face-to-face. Both courses were taught by the same professor using the same curriculum with groups similar in academic background. To determine if any differences existed in the learning outcomes between the two groups, data were collected from two sources. The evaluation included a major course project and a self-assessment instrument. The difference in quality between the work presented by the online learners vs. the face-to-face learners was not significant. The mean rating for the online class was 3.40 out of 4. The mean rating for the face-to-face learners was 3.47 out of 4. The course grades were evenly distributed between the online and face-to-face learners. Overall, the results from this study support "no significant difference" between online and face-to-face learning.

A study of a college level math class that tracked students for two years, and reviewed the results of paper and pen test scores, as compared to online scores found that, "Students do seem to perform better in the online section..." (Engelbrecht & Harding 2004, p. 221). Another study found that no significant difference in grades or grade point averages between the traditional and non-traditional (distance learning) students in the same business class existed. In the study by Jennings (2003), online students raved about the flexibility, the self-pacing, lower costs of travel and sitters, and the use of their higher level communication skills honed by the professor's use of discussions. The online students felt confident about how they were able to consider the discussion board assignments and develop a quality response through brainstorming the topic prior to posting a final response. The "online students scored significantly higher in content examinations and experienced greater levels of instructor-student interaction than did students in the traditional on-campus course" (Jennings, 2003, p. 184). In comparison, the online students' grade point averages were a 2.68, while the traditional students maintained a 2.65 (Jennings, 2003).

Some studies have addressed the problems associated with developing online tests. One of the problems addressed is the need to link questions to course objectives and learning outcomes, while another is the inadequate number of questions that assess the student's mastery of the content (Driscoll, 2001). Yet another problem pertains to determining the identity of the test-taker. There are programs which have been developed to address this challenge, such as JonesKnowledge.com's e-education proctored tests. Students are not allowed to begin a test until a pre-designated proctor has logged in his or her password, integrating a pre-set time limit with an automatic submission at the conclusion of the testing period (Chronicles, 2000). Other measures developed for testing security include Pretty Good Privacy (PGP) encryption software, which provides forge-proof digital signatures on e-mail communications between students and professors to allow authentication of essay-type tests. Veridicom (a venture partnership with Lucent Technologies) developed a fingerprint scanner which can be used to authenticate students (Galambos, 1999).

One method universities rely on to show the effectiveness of online learning is through proctored exams. A proctored exam is considered “a secure testing in distance education” (Pace University, 2003, p. 1). In a proctored exam, the online student is verified by the exam proctor (Pace University, 2003). For instance, at Pace University, the University reserves the right to approve or deny a proctor for a student and have functional databases that track students’ proctors. In this proctored setting, it ensures that students do not use their study materials (or other students in the class) for assistance. The restriction includes cell phones, text messaging, memory keys and other technologically advanced equipment that could carry information necessary to pass the exam.

Research has been conducted in regard to proctored versus un-proctored exams. In a study undertaken at a major university, it was revealed that students learned more (scored better) on exams that were proctored versus students who took the same classes (and content) but were not proctored (Wellman, 2005). At this university, the main concern was that allowing students to be in an un-proctored setting would allow them to “collaborate without the consent of the instructor” and “...influence cheating behavior...” (Wellman, 2005, pp.1-2). Although both groups prospered under this study and showed major gains in their pre/post tests, the proctored group of students showed more gains than the un-proctored group of students (Wellman, 2005).

In the proctored testing environment, the professor maintains the supervision and influences the examination honesty, thereby reducing cheating. In addition, a proctored student has the structure of a classroom and the face-to-face interaction of their professor. Furthermore, the proctored students have the environment of a “time and place” to be at the exam and take it versus an un-proctored student who could take it at his/her leisure within the 24 hour exam period. It is also possible that the classroom environment could cause proctored students to become more competitive with each other and work harder to complete their exam. Lastly, the intra-test feedback secures the proctored student into knowing there is someone readily available if something goes technically wrong with the exam whereas the un-proctored students lacked the security of having a technical expert in the room when the exam was taken (Wellman, 2005).

Although proctored examinees scored higher than the un-proctored students on both exams and had higher pre/post test results, test anxiety was much higher on the proctored students. On the other hand, the proctored students had higher motivation, time management, concentration, information processing main idea, techniques of self-testing, and on testing preparation (versus the un-proctored students). Additionally, the top half of the proctored group scored an average of 3.62 points higher than the un-proctored testing group where the bottom proctored group scored 2.32 points higher than the un-proctored group (Wellman, 2005).

Although this study suggests that students in a proctored setting motivate themselves and that the proctored setting is more challenging, it disputes that the un-proctored students used each other to complete the exam. If the un-proctored students had utilized outside resources (including each other), why were their scores lower? Although the researchers presumed that these un-proctored students did not study as hard as the proctored students and therefore did not learn more – the grades on the proctored students ranged from the high of an 18.93 to the low of 8.8. In contrast, in the un-proctored group, the high was a 15.31 and the low was a 6.48. If the outcome of the studies had subsequently been reversed, would the researchers have assumed that these un-proctored students had cheated? Wellman concluded that un-proctored testing, paired with on-line content delivery, was less effective in promoting overall learning and reduced time-on-task for some students (Wellman, 2005).

Unproctored online tests are in effect open-book exams. Test-takers typically have reference materials available to them during the course of the exam. Does this provide an unfair advantage to online test-takers over their brick-and-mortar counterparts? It may, and then it may not. It depends on the specific questions that the administrator chooses to create for the exam. To prohibit the test-taker’s reference materials from being too much of a benefit, the professor may ask some questions that require interpretation and or application of one or more of the concepts contained within the testable material, or questions regarding one or more of the more obscure concepts contained within the testable material. However, the real question may be, “Does it matter?” The response to this question is up to the test administrator to decide.

Online testing presents several new challenges to the reliability and validity of the test given and the testing process. Trochim (2001) defines reliability as the consistency or repeatability of your measures. This may be asked in another way, if the test-taker expected to be taking the test retook the test, would he/she respond with the same answers to the questions? Trochim defines validity as the best available approximation to the truth of a given proposition, inference, or conclusion (Trochim, 2001). Several questions must be answered which are of much less concern in a traditional fixed-location-administered pen and paper test.

Who exactly is taking the online test? Has test-prepping taken place? Has there been collaboration between students, or other forms of cheating by students in the test-taking process? Can new biometrics aimed at verifying test-taker identities be effective? Are open-book online exams unfair to brick-and-mortar students? What are the benefits of online testing?

In a virtual environment how can the absolute identity of the test taker who is entering the inputs into a computer terminal in a remote location be known (Irele, 2006)? At a computer terminal in a remote location, an authorized test-taker could be making all the test inputs, but deriving the question answers with the aid of one or more assistants who provide the correct answers during the test. Biometrics may prevent alternate test-takers from sitting at the terminal and making the inputs; however, they cannot prevent unauthorized persons from “feeding” the correct answers to the test-taker during the test. This action would not be very different from passing question answers around during a proctored exam in a physical location –which has occurred in a variety of test settings for decades. For that matter, how do we know for sure that a research paper submitted in a traditional classroom was authored by the person receiving credit for the work? Jeanne M. Wilson, president of the Center for Academic Integrity, which is affiliated with Duke University, believes the growth of on-line education makes it harder to be sure that the student who gets the credit is the student who did the work (Carnevale, 1999). Again, the integrity of the testing falls on the personal sense of honor of the test-taker.

STATEMENT OF THE HYPOTHESIS

The research hypothesis for this study states that there will be a significant difference in examination scores between examinations taken before on-line testing was initiated, as compared to those taken after the advent of on-line testing. The null hypothesis states that there will be no significant difference in the scores of examinations taken before on-line testing was initiated, as compared to those after on-line testing was initiated, as measured at the $\alpha = .05$ level of significance.

RESEARCH METHODOLOGY

This research project utilized the causal-comparative research methodology. The authors gathered examination scores from the courses which were taught before and after the advent of on-line testing. The courses selected for this study were MGMT 311, Marketing; MGMT 210, Financial Accounting, and ASCI 254, Aviation Legislation. These courses were selected because they were some of the initial courses to transition into online testing, at least one section is offered on a monthly basis, and there is historically a large enrollment of students for these courses. These data were extracted from student records located in the Distance Learning Department. A t-test for independent samples was utilized to determine if there was a significant difference in examination scores between those taken before and after on-line testing was instituted. The researchers realize that combining examination scores for different courses would not be appropriate since there differences may exist, so courses were considered separately in this study. The study also gathered information to determine if there was a significant difference between grade point average of courses taken before and after the initiation of on-line testing. It is possible that while there is a significant difference in examination grade, there may not be a difference in overall course grades since the course monitors have adjusted other requirements associated with the course.

RESULTS

Not all of the information furnished from the Department of Distance Learning could be utilized in this study. In some cases the grades for the midterm and final examinations were letter grades, and no percentages were available for quantification. In other cases there were no letter grades available, just point totals, while in others only the grades for one of the examinations was listed. The data for the proctored examinations were extracted from the classes held during the months of June through November of 2005. The data for the non-proctored online examination were from classes conducted during the months of January through June of 2006. Some Embry-Riddle courses are offered 11 times a year, starting every month except December, while others are offered less frequently, depending upon the demand.

Table 1
MGMT 311 Midterm Examination Grades – t-Test Results

Proctored Paper Examination		Online Non-Proctored Examination	
N	210	N	290
Mean	81.63	Mean	85.42
df	498		
t-value	4.44		
Significance	.05		
Critical t	1.960		

Table 2
MGMT 311 Final Examination Grades – t-Test Results

Proctored Paper Examination		Online Non-Proctored Examination	
N	208	N	285
Mean	77.63	Mean	82.00
df	492		
t-value	4.71		
Significance	.05		
Critical t	1.960		

Table 3
MGMT 210 Midterm Examination Grades – t-Test Results

Proctored Paper Examination		Online Non-Proctored Examination	
N	290	N	289
Mean	72.75	Mean	77.38
df	577		
t-value	3.98		
Significance	.05		
Critical t	1.96		

Table 4
MGMT 210 Final Examination Grades – t-Test Results

Proctored Paper Examination		Online Non-Proctored Examination	
N	289	N	288
Mean	73.57	Mean	77.45
df	575		
t-value	3.48		
Significance	.05		
Critical t	1.96		

Table 5
ASCI 254 Midterm Examination Grades – t-Test Results

Proctored Paper Examination		Online Non-Proctored Examination	
N	242	N	282
Mean	76.06	Mean	80.34
df	522		
t-value	3.023		
Significance	.05		
Critical t	1.960		

Table 6
ASCI 254 Final Examination Grades – t-Test Results

Proctored Paper Examination		Online Non-Proctored Examination	
N	258	N	263
Mean	74.07	Mean	77.47
df	519		
t-value	2.319		
p.	.05		
Critical t	1.960		

DISCUSSION

Results from the t-tests revealed a significance difference in both the midterm and final grades between the proctored and online non-proctored examinations. While the scores were significant, they did not appear to be overly excessive.

CONCLUSIONS

The t-test results for the three courses revealed a significant difference for both the midterm and final examinations between the proctored examinations and online non-proctored examinations. The hypothesis for the study, that there would be a significance difference between the examination scores between proctored and online non-proctored examinations, was supported and the null hypothesis rejected. While the authors hypothesized that there would be a significant difference in the examination scores since the online test was not proctored, thus giving student access to the text and notes, the results were much closer than expected by the authors. The difference between

the midterm and final examination means for the MGMT 311 examinations were 3.79 and 4.37 respectively; for MGMT 210 they were 4.63 and 3.88; and for ASCI 254 they were 4.28 and 3.40.

While the research revealed significance between the proctored and non proctored online examinations, the increase in points was less than five in three of the examinations, less than four in the remaining three. To truly determine the affect of non proctored online testing, further study needs to be undertaken to determine if this increase in points has led to a significant increase in the overall grades for the courses. A follow-on study is recommended to analyze the final grades for the courses to determine if there is significance between proctored and non proctored online testing.

REFERENCES

1. Bocij, P. & Greasley, A. (2003). Can computer-based testing achieve quality and efficiency in assessment? *International Journal of Educational Technology*, 31(4).
2. Carnevale, D. (1999). How to proctor from a distance. *The Chronicle of Higher Education*, A47. Retrieved March 31, 2006, from: <http://chronicle.com>
3. *Chronicles*. (2000, February). Retrieved March 1, 2006 from http://JonesKnowledge.com/news/chronicles_feb_00.html#1
4. Driscoll, M. (2001, June). Building Better E-Assessments. Retrieved February 2, 2006 from <http://www.learningcircuits.org/2001/jun2001/driscoll.html>
5. Engelbrecht, J. & Harding, A. (2004). Combining online and paper assessment in a web-based course in undergraduate mathematics. *The Journal of Computers in Mathematics and Science Teaching* 23(3) 217-31.
6. Galambos, D. (1999). WebCT: Info and Resources. Retrieved March 10, 2006 from <http://www.sheridanc.on.ca/academic/edserv/WCTHome.htm>
7. Irele, M. (n.d.) The World Campus. On-line Testing and Assessment: Effectiveness and Security Issues. Retrieved January 10, 2006 from http://www.ed.psu.edu/acsde/online_testing_assess.pdf
8. Johnson, S., Aragon, S., Shaik, N., & Palma-Rivas, N. (2000) Comparative Analysis of Learner Satisfaction and Learning Outcomes in Online and Face-to-face Learning Environments, *Journal of Interactive Learning Research* 13, (3.)
9. National Center for Education Statistics. (n.d.). *Distance Education in Higher Education Institutions*. Retrieved March 29, 2006, from: <http://nces.ed.gov/surveys/pegis/publications/98062/>
10. Pace University (2003). *Secure Testing for Online Learners* [Electronic version] Retrieved March 26, 2006, From http://www.sloan-c.org/effective/details3.asp?LE_ID=42
11. Parker, D., & Gemino, A. (2001) Inside Online Learning: Comparing Conceptual and Technique Learning Performance in Place-based and ALN Formats, *Journal of Asynchronous Learning Networks*.
12. Schulman, A. H., & Sims, R. L. (1999, June), Learning in an online format verses an in-class format. *T.H.E. Journal*, 26, 11; Educational Module, p. 54.
13. Schutte, J. G. (1997). Virtual Teaching in Higher Education: The New Intellectual Super Highway or Just Another Traffic Jam? Retrieved March 22, 2006 from <http://www.csun.edu/sociology/virexp.htm>.
14. Sloan Consortium. (2005). *Growing by Degrees: Online Education in the United States*, 2005. Needham, MA: Sloan-C. Retrieved March 29, 2006 from: http://www.sloan-c.org/resources/growing_by_degrees.pdf
15. Trochim, W. M. K., 2001. *The Research Methods Knowledge Base*, 2nd ed. Cincinnati, OH: Atomic Dog Publishing.
16. Wellman, G. S. (2005). Comparing learning style to performance in on-line teaching: Impact of proctored versus un-proctored testing. *Journal of Interactive Online Learning* 4(1) 20-39.
17. Woit, D., & Mason, D. (2003). *SIGCSE '03*, February 19-23.