

A COMPARISON OF ONLINE VS. PROCTORED FINAL EXAMS IN ONLINE CLASSES

By

ABRAHAM(AVI) MOZES-CARMEL *

STUART S. GOLD **

**Professor of Research Methods, Statistics and Conflict Management, Nova Southeastern University .*

***Adjunct Professor, Management and MIS courses, Nova Southeastern University.*

ABSTRACT

This study examined the question of the relationship between the modality used to administer a final exam and the student outcomes achieved on the final exam in online university courses. This question directly addressed the issues of 1) the ongoing and dynamic growth of online university offerings and 2) the need for online universities to employ processes that will scale to allow for effective management of large numbers of online course takers. The study was a large-scale study that incorporated data from 100 online courses and over 1800 students across the full range of undergraduate course offering at the institution studied. The results indicate that in the university environment studied there was not a statistically significant difference between the average final exam score obtained irrespective of the modality used to administer the exam. The results of the study indicated that the use of sophisticated online exam software is a viable alternative to the use of proctored final exams. However, it is not necessarily sufficient to merely provide an online final exam. The environment studied also chose to take a number of measures to assure the academic rigor of the online exam and to minimize the ability of the student to access other online resources while taking the exam.

Keywords: Distance Education, Hybrid Courses, Online Education, Online Courses, Student Retention, Student Satisfaction, Online Final Exams, and Proctored Final Exams.

INTRODUCTION

The purpose of this study was to address the question. Are student outcomes achieved in an online final exam when compared to the outcomes of proctored final exams for online university classes? As stated by James, McInnis and Devlin (2002) the question is: does on-line assessment have any influence on the quality of learning. This study directly addresses the issue of the need for online universities to employ processes that allow for effective management of large numbers of online courses.

Specifically, the research project examined the relationship between student outcomes on final exams and the modality of final exam delivery among working adult students in online courses. Through an analysis of archival course records and student final exam grades, the researchers were able to conduct statistical analyses of the data for a sample of 50 courses in each group (online and proctored final exams).

Statement of the Problem

Are the student outcomes achieved when administering an entirely online final exam comparable to the outcomes achieved when administering proctored final exams for online university classes?

Importance of the Project

Enrollment in online courses has been growing at an extremely fast rate for the past several years and is projected to continue this growth for the foreseeable future. One of the issues this dynamic growth has created is the scalability of the internal management processes and systems within the university. Processes that were time-tested and worked well for a few hundred students tend to encounter problems when the student population increases to thousands and tens of thousands of course takers per term. The university being studied is a classic example of such rapid growth.

As working business professionals participate in distance

learning at increasingly higher rates, it is important to identify specific instructional technology that can scale readily to support this increasing population of course takers and provide positive outcomes for these students. Institutions of higher education are actively expanding or implementing online education programs to meet this burgeoning trend. The need to identify instructional technology that supports the increasing number of online course takers becomes increasingly important to the success of such programs.

This study focused on one of the key administrative and educational issues, the scalability of the final exam process. Wellman and Marcinkiewicz (2004) state that "as educators adopt online instructional techniques, one of the challenges they face is assessing learner mastery of course content." James, McInnis and Devlin (2002) stress that if lower-order learning becomes the result of online assessment, then the gains made in efficiency, staffing and cost savings may be offset by a drop in the quality of the outcomes achieved.

Until late 2003 all final exams for online students, at the institution being studied, were delivered in a proctored setting. This required the student to obtain an approved proctor. Faculty submitted final exams to the university administration, copied manually and then a copy was sent to the proctor for each student. After the student completed the exam, the proctor returned the completed and properly validated final exam packet to the university to be copied and filed. The completed exam copies were subsequently mailed to the faculty for grading. Once the grading was complete, faculty members made copies of all exams for their own records and returned the graded final exams to the university along with an end of term grading package. In late 2003, the university administration decided, that the copying, express delivery, temporary workforce requirements and inherently time consuming and error prone nature of this process required it to be changed.

Traditional assessment techniques are costly and time consuming efforts, which an online course management system should be able to alleviate, if the results of the online process can be trusted (Rowe, 2004). The issue of

trust has been a significant factor in slowing the implementation of online testing.

While plagiarism has been a focus of many online programs there has been much less attention paid to other problems related to the issue of dishonesty in online assessment (Rowe, 2004). It is important to remember that cheating on final exams is far from a new phenomenon and certainly not an online exam only situation. Bushweller (1999) cites statistics stating that 70% of American high school seniors admit to having cheated on an in-class exam. Further 95% of the students who did admit to having cheated said they were never caught. Numerous other authors support the perspective that cheating on exams is not a phenomenon unique to the online environment, including Cizek (1999) who makes the point that cheating increases with student age. This is a significant issue for online programs, which focus on the adult learner population.

In spite of the risks, the decision was made in 2003 to move to a purely online form of final exam for all classes, eliminating the proctor and the perceived safeguards that a proctor may provide in terms of academic integrity. In place of the proctor, the university took several important steps to address the integrity and quality of the final exam process. All final exams were rewritten and subsequently reviewed by the academic management group and departmental Deans as appropriate. The use of question pools was encouraged to randomize the questions and the order in which questions appeared to the students. To reduce the possibility of students engaging in various forms of online activity, which could degrade the integrity of the process, special software, was implemented within the course management system. This software prevented the student from using the internet or other sources while taking the exam, though the course text was available as a hard copy reference. The primary steps taken by the university to address the broader range of issues were as follows:

Length of Exam = 3 hours. In developing their final exams, faculty members will need to keep in mind that the average student in their courses will be able to complete their final exams in 3 hours. This may mean that a few

students may not finish the exam in the 3 hour time period. Faculty members can add up to 30 minutes of additional time beyond the 3 hours for technical issues. After 3.5 hours, all students will be locked out of their final exams with their answers saved.

The question types are objective and essay. For Undergraduate minimum of 30% essay, and for Graduate no more than 20% objective.

Number of points and number of questions will remain open and up to the discretion of the faculty member, as long as it can be supported that all Terminal Course Objectives are sufficiently covered. Each exam will be reviewed to ensure that final exams assess the mastery of each terminal course objective.

Benjamin Bloom's Taxonomy will be used as the guide to writing questions that more effectively measure students' ability to use versus memorization of information. The Taxonomy of Educational Objectives is divided into six major levels: Knowledge, Comprehension, Application, Analysis, Synthesis, and Evaluation.

Undergraduate final exams will contain test items that target the objectives within the following percentages (note: higher levels of questions are preferred, considering that these exams are open book and open notes):

- *Knowledge and Comprehension = 20% to 30%*
- *Application and Analysis = 50% to 60%*
- *Synthesis and Evaluation = 20% to 30%*

Graduate final exams will contain test items that target the objectives within the following percentages (note: higher levels of questions are preferred, considering that these exams are open book and open notes):

- *Knowledge and Comprehension = 20% to 30%*
- *Application and Analysis = 30% to 40%*
- *Synthesis and Evaluation = 40% to 50%*

Wellman and Marcinkiewicz (2004) state "that there is paucity of research examining the impact of proctored versus un-proctored testing..." Quilter and Chester (2001) emphasize that few formal research studies have been conducted to examine the relationships between online communication technologies and teaching and learning

and reaffirm that "research with empirical documentation of the use of communication technologies is lacking".

This project analyzed the results at the end of the first year of the utilization of online exams by the subject institution. It advanced the research by providing an objective comparison of two of the more commonly used modalities for administering final exams in online university courses and by utilization of data obtained directly from the course management software database. This combination added a new dimension to the body of knowledge in this area.

Methodology

t-test for independent samples

Purpose and Assumptions

The *t*-test is the most commonly used method to evaluate the differences in means between two groups. In this study, the *t*-test can be used to test for a difference in test scores between a Online vs. Proctored testing. Theoretically, the *t*-test can be used even if the sample sizes are very small, as long as the variables are normally distributed within each group and the variation of scores in the two groups is not reliably different. As mentioned before, the normality assumption can be evaluated by looking at the distribution of the data by performing a normality test.

The *p*-level reported with a *t*-test represents the probability of error involved in accepting our research hypothesis about the existence of a difference. Technically speaking, this is the probability of error associated with rejecting the hypothesis of no difference between the two categories of observations (corresponding to the groups) in the population when, in fact, the hypothesis is true.

Arrangement of Data

In order to perform the *t*-test for independent samples, one independent (*grouping*) variable (e.g., Modality) and at least one dependent variable (e.g., a test score) are required (Table 1). The means of the dependent variable was compared between selected groups based on the specified values (e.g., *modalities*) of the independent variable. The data set can be analyzed with a *t*-test

| | Modality | Score |
|--------------------------|-----------|-------|
| case 1 | | |
| case 2 | Online | 90 |
| case 3 | Proctored | 80 |
| case 4 | Etc... | 95 |
| case 5 | | |
| mean Grade Online = X | | |
| mean Grade Proctored = Y | | |

Table 1. Test Scores with Modality

comparing the Test scores with modality (Table 1).

Sample

The research project included a total of 100 courses and approximately 1800 students that together provide a detailed analysis of the topic. Courses were selected using random sampling techniques.

Archival data collected from course management software administration statistics included: (a) student final exam grades and (b) final exam delivery modality. The database recorded actual grades only and did not include any information related to end of course student surveys or student satisfaction. Therefore, it was not possible from the available data to determine the level of student satisfaction with a course or to relate the level of student satisfaction to the method of final exam administration employed. Data analysis was performed using appropriate statistical techniques.

The online learning program evaluated in this study is a regionally accredited, university offering a range of undergraduate and graduate degree programs to students in both online and face-to-face formats. The university offers programs in business, management, and technology specifically directed toward working adult professionals. Online courses from each of these programs were included in the study.

The final exam process design was designed to assure that final exams were administrated consistently and included comprehensive coverage of the entire course. All final exams were developed with a wide range of questions, covering both lower level and higher-level cognitive skills as defined in Bloom's Taxonomy (Bloom et al, 1964). The consistent nature of the final exam structure and process throughout the university and the wide range

of disciplines and number of classes included in the study served to increase internal validity.

Summary of Findings

There was not a statistically significant difference in the average final exam grade achieved for students in courses utilizing proctored final exams vs. the average final exam grade achieved for students in courses utilizing online final exams. The analysis was performed using student's 't'-test for comparison of the group means of the final exam raw scores. The analysis was not significant at the .01 or .05 levels. This result was verified by an Analysis of Variance test which yielded similar results.

The conclusion was that the method of final exam administration was not a significant factor in determining the average grade achieved on the final exam.

The average values for the final exam scores are shown in Table 2. The 't' scores, ANOVA results and related significance levels for the comparison of the group mean values are shown in Table 3a and b.

The averages were surprisingly close, almost identical, and clearly there is no statistically significant difference between the two groups in this respect. However, the average does not tell the entire story. Upon further analysis the nature of the grade distributions proved to be somewhat different.

To provide another picture of the data, box plots (Figure 1)

| Modality | Average Score |
|-----------|---------------|
| Online | 73.8070 |
| Proctored | 73.8838 |

Table 2. Average Final Exam Score

| 't' | Sig. (2-tailed) |
|-------|-----------------|
| -.029 | .977 |

Table 3a. Statistical Analysis (Student's 't' - test)

| | Sum of Squares | df | Mean Square | F | Sig. |
|----------------|----------------|----|-------------|------|------|
| Between Groups | .147 | 1 | .147 | .001 | .977 |
| Within Groups | 17008.013 | 98 | 173.551 | | |
| Total | 17008.160 | 99 | | | |

Table 3b. Analysis of Variance

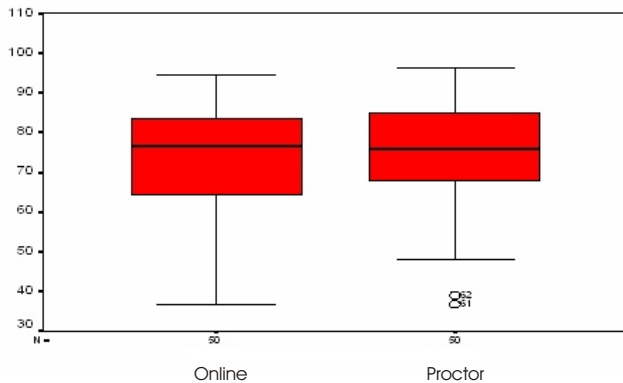


Figure 1. Box plots for each sample

were calculated for each sample. As can be readily observed from the box plots, the proctored final exam grades exhibited a more tightly grouped distribution in comparison to those recorded from the online final exams. The proctored exam group had fewer courses with final exam scores toward the bottom of the range. The result of this tighter grouping is that the very low scores do not fit in statistically with the rest of the proctored exam sample and are considered to be outliers in that group even though they would fit readily into the online exam sample. (Trochim, 2001)

As a next step, the outliers were removed from the data and the analysis was redone. When the outliers were removed from the proctored final exam group that changed the sample average from 73.8838 to 75.3842. This revised average value was still not statistically different from the mean value for the online sample. The conclusion is that the online final exams appear to be allowing for a somewhat greater degree of variation in the average class scores than was present under the prior system. This is an area which may merit further research, however that research was outside the scope of this study.

In both groups the range of final exam class average scores was extremely broad with class averages ranging from approximately 36% to 94% for online exams and from approximately 48% to 96% for proctored exams (excluding the outlier values). If we include the outliers in this analysis the data for the proctored exams ranges from approximately 37% to 96% which is much closer to the values seen for the online courses.

The results relate well to earlier research by Smith and Dillon (1999) who refer to the media/method confound, a concept stating that the technology alone does not cause the effect, rather it is the combination of the technology and the way the technology is employed that impacts student outcomes.

Implications

The research study results suggest that the university was able to construct a set of online final exams which were generally equivalent to the prior proctored final exams in terms of student outcome achieved on the test. The intent was not to validate the existing proctored exams, or to determine their efficacy in a pure online mode. The university instead decided to focus on creating a revised set of online exams that would yield similar results.

The analysis indicates an opportunity for further research by gathering a larger sample of classes in specific disciplines or courses. This would allow for a more granular analysis incorporating course specific and instructor specific variation. Another opportunity for research is in the analysis of the distribution of grades in an online vs. proctored exam environment.

Limitations of the Study

The use of packaged course management software is a relatively recent development in the history of instructional technology. Faculty and student use of software facilitated communication tools will continue to evolve over time and new software features for managing online testing will be developed. This factor represents a potential limit on the external validity of the study to generalize these findings to different course management systems or to future upgraded releases of the course management system used in the study.

The course management system database contained data focused solely on the final exam grades and did not incorporate any information related to end of course surveys or student satisfaction. The use of archival data from the course management system database limited the opportunity to study factors such as the level of student satisfaction with the course.

Similarly, the experience level of the student with the

technology may have impacted the internal validity of the results. Incorporating a wide mixture of courses ranging from beginning to advanced levels served to mitigate this effect.

As a single university setting was used in the research, there was no control group against which to measure the results of the research. This may limit the external validity of the study and the generalization of findings to other institutions and other forms of course design and use. This issue was mitigated by the fact that the collected data is similar to data provided by the course management systems in use at many universities.

There may be indirect relationships supported by multiple factors impacting the final exam outcomes including student perceptions and attitudes toward the final exam modality employed. The analysis of these indirect relationships was outside the scope of the proposed study. Additional comparisons and extending the research to include more data should be a focus for future research, as it will solidify the assertions made in this research.

References

- [1]. Bushweller, K. (1999, April). Generation of cheaters. *The American School Board Journal*, 186(4), pp. 24-32.
- [2]. Bloom, B., & Masia, B. Krathwohl, D., (1964). *Taxonomy of Educational Objectives the Classification of Educational Goals: Handbook II Affective Domain*. New York: David McKay Company Inc.
- [3]. Cizek, G. J. (1999). *Cheating on tests: how to do it, detect it, and prevent it*. Mahwah, NJ: Lawrence Erlbaum.
- [4]. James, R., McInnis, C., & Devlin, M. (2002). *On-line assessment*. Retrieved April 25, 2005, from <http://www.Adelaide.edu.au/myuni/instruct/planning/assessment.html>
- [5]. Quilter, S. M., & Chester, C. (2001). *The relationship between web-based conferencing and instructional outcomes*. Retrieved July 31, 2003, from Wilson SelectPlus database.
- [6]. Rowe, N. C. (2004, Summer). Cheating in online student assessment: beyond plagiarism. *Online Journal of Distance Learning Administration*, VII, Retrieved April, 15, 2005, from <http://www.westga.edu/~distance/ojdla/summer04/olt72.html>.
- [7]. Smith, P. & Dillon, C. (1999). Comparing distance learning and classroom learning: conceptual considerations. *The American Journal of Distance Education* 13(2), pp. 6-23. Retrieved February 23, 2003, http://www.ajde.com/contents/vol13_2.html.
- [8]. Trochim, W. M. (2001). *The research methods knowledge base* (2nd ed.). Cincinnati, OH: Atomic Dog Publishing.
- [9]. Wellman, G. S., & Marcinkiewicz, H. (2004, December). Online learning and time-on-task: impact of proctored vs. un-proctored testing. *Journal of Asynchronous Learning Networks*, 8, Retrieved April 26,, 2005, from <http://www.sloan-c.org/publications/jalnv8n4/>

ABOUT THE AUTHORS

Abraham (Avi) Mozes Carmel, Ph.D., is a published author with more than 15 years experience in traditional and on-line instructional delivery. He is a Professor of Research Methods, Statistics and Conflict Management. He has a Bachelor in Business Administration degree from Temple University; Master of Business Administration from University of Phoenix; Jurist Doctor from University of Florida, and a Masters of Science and a Doctor of Philosophy from the Graduate School of Humanities and Social Studies at Nova Southeastern University. He can be reached at acarmel@bellsouth.net.



Stuart S. GOLD, Ph.D., is an experienced Professor teaching both on line and face-to-face classes. As an adjunct professor he has taught Management and MIS courses in both the United States and the Caribbean during the past several years. He holds a B.A. in Physics and Mathematics from Northeastern Illinois University, M.B.A. from Loyola University of Chicago, and PhD from Northcentral University. He can be reached at sgold1@bellsouth.net.

