

Integrity of Online Testing in E-Learning: Empirical Study

Tom Wielicki

Craig School of Business, California State University, Fresno

e-mail: tomaszw@csufresno.edu

Abstract

An online upper division undergraduate MIS course has been designed and delivered over consecutive five semesters to a sample of over 200 students. The same course was also delivered in a web-enhanced version (paperless course with regular classroom lectures). This paper reports on a study, which attempts to answer numerous questions regarding impact of an e-learning format on quality of educational experience. Specifically, in this part of the study we will try to find out whether online testing with an open book format is compromising integrity of assessment in e-learning by encouraging cheating among students, and what is the difference in online testing and assessment between students in web based (fully online) class versus web enhanced class.

1. Introduction and literature review

E-learning is becoming increasingly dominating delivery format for training and education. It has been widely adopted by the corporate world as it is extremely cost effective in delivery of internal corporate training [1]. The same can not be said about education – especially higher education, where objectives of instructional activities are broader and more complex than objectives of typical training. Also, universities seem to have more problems with incorporating this new technology into an overall strategy and business processes since – ironically – they are more resistant to change [2]. This may be a reason for apparent differences between number of online credit courses and degree programs offered by lower tier unaccredited institutions and those fully accredited. Accredited degree programs seem to be much more cautious in adopting e-learning format out of concern about quality of education and requirements of accrediting institutions. Big part of this skepticism

is attributed to legitimate questions about reliability of online testing and assessment, especially at the undergraduate level. Specifically, issue of security or lack of it in a web based testing has been preoccupying researchers like Adams and Armstrong (1998) leading to numerous software solutions like their Eval program used for testing at undergraduate level [3].

Hodgins (2002) in his vision paper developed for the American Society for Training and Development (ASTD) emphasized “Assessment and Certification” as one of the main areas where impact of technology on e-learning has to be closely monitored and controlled [4]. Similarly, Dobbs (2002) in his definition of the state of online learning is concentrating on four fundamental obstacles to high quality of e-learning. Number one problem identified by him is a flawed perception that “reading is learning”. He is suggesting that more interaction should be built into the e-learning as well as effective assessment mechanism [5].

Assessment seems to be an important part of study in the area of designing and evaluating online learning environment like the one proposed by Hoffman and Ritchie (2001) [6]. However, its impact on the quality of educational experience is hardly ever measured and assessed in empirical settings.

A serious of quantitative studies based on a solid samples of web based students performance has recently been completed shading more light on the issues of viability of e-learning (see Wei-Fan Chen, 2005) [7].

At the same time some authors warned against Digital Doctrine that greatly overestimates impact of technology on economy and education (see – Albreht and Gunn, 2000) [8]. Some anticipate that dot-com bust could be repeated with disappointments in the field of e-learning, due to irreproducibility of some important components of face to face learning process.

This study attempts to continue a trend of

verifying myths created around e-learning with statistically sound samples of data.

2. Methodology and hypothesis

A sample of 230 students took an upper division undergraduate MIS course, which was delivered fully online using Blackboard 6 – a comprehensive e-learning environment. At the same time another 186 students took the same course with the same instructor and using the same text book but in a web enhanced mode. Web enhanced mode is defined here as a paperless class with all materials, handouts and communication delivered in a digitized form (using Blackboard content), with all tests administered online but with students still participating in a traditional lectures in classroom settings.

Couples of hypothesis were formulated addressing different dimensions of quality of assessment process:

- Online open book delivery format of quizzes and tests is conducive to cheating and abuse, therefore test scores will be impacted by the assessment feedback
- Online open book delivery format of quizzes and tests is conducive to cheating and abuse, therefore test scores will be impacted by the level of questions randomization used in the assessment
- Students in a web enhanced class taking online tests have more chance to organized themselves for a purpose of cheating than those taking fully web based class, thus they will do better on the same set of online tests
- Web based delivery format as more “anonymous” and conducive to cheating will provide for higher average grades than web enhanced format

3. Experiment Design

A sample of 416 students took 12 quizzes and 2 tests during one semester upper division MIS course. This means that total number of graded assignments (quizzes and tests) used in this study is equal to 5824. It has been insured that the level of difficulty was uniform for all students by using the same pools of questions, the same textbook and the same time frame for the assignments. About a half of the sample were web based students (online course), which had almost no face to face contact with the instructor and each other. The other half of the sample included students that participated twice a week in a regular lecture,

knew each other and benefited from instructor’s face to face consultation hours.

3.1. Variables and Treatments

Blackboard environment provides numerous settings for designing of an online test. Every design could be more or less conducive to cheating, depending on such parameters as:

- time allocated to every question,
- enforcement of sequential way of answering questions (one at a time) versus scrolling page,
- type of provided feedback (just the score, identification of questions missed and the score, identification of question missed and correct answer)
- questions randomization from a larger pool versus the same set of questions

Every combination of these parameters could be used as a treatment. For purpose of this study only two of those parameters were used to create treatments in statistical analysis of scores:

- type of provided feedback and
- randomization

The first of the two parameters are described in Table 1 (below) as quiz or test setup controls of e-learning environment called Blackboard:

Table 1. Feedback types in the blackboard assignment setup

Feedback Mode		
SCO	<input type="radio"/>	Score Only Only the final score is presented.
DRNA	<input checked="" type="radio"/>	Detailed Results The user’s answers, whether they are correct, and the final score are presented, but not the correct answers.
SCA	<input type="radio"/>	Show Correct Answers The user’s answers, the correct answers, and the final score are presented, but not any question feedback.

The second parameter may take on one of the two possible values:

- select questions at random from a larger pool – R

- select the same questions for every student – NR

Those treatments represented arrangements under which cheating during an open book online quiz or test could be either very easy or very difficult. A variable that was measured for every treatment was an average score (class mean) on a given test or quiz with specific format. It was assumed that - should students abuse an online format of testing – the mean of scores should consistently drop as we move from “easy to cheat” treatments to “difficult to cheat” treatment. In other words – if there was any abuse of online testing among students, it was expected that difference between the mean scores will be statistically significant as we compare different combined setups shown in Table 2 below.

Table 2. Combined setups for delivery of online assignments

SCA-NR	show answers; the same set of questions
DRNA_NR	show missed questions but no correct answer; the same set of questions
SCO_NR	show only total score; the same set of questions
SCA-R	show answers; randomized questions
DRNA-R	show missed questions but not correct answer; randomized questions
SCO-R	show only total score; randomized questions

It is reasonable to assume that above formats (assessment setups) represent an increasing degree of difficulty in cheating, therefore treatments from the first row to the last may be viewed as a scale of increasing “degree of difficulty in cheating.”

Separate statistical tests were conducted on quiz scores and tests scores due to the difference in settings of the assessment process in both cases: quizzes included only 10 multiple choice questions and were allocated 10 minutes, whereas tests included 50 questions and were allocated 50 minutes. All remaining setup parameters of the assessment process were the same for all collected data: all questions were multiple choice questions, there was always 60 seconds time allocated to every question, and there was always a possibility of answering questions in any order (a scrolling mode enabling student to answer questions in any sequence).

3.2. Statistical Tests

Numerous statistical tests have been conducted to verify some of the hypothesis listed above. Primary focus of this analysis was on the issue of searching for

statistically significant difference in the mean scores on online assignments administered under different settings, which were more or less conducive to cheating and abuse by the students.

The first test was conducted using One-Way ANOVA F-test for verification of significant difference among the mean scores on assignments administered with different level of **feedback** (treatments). Null hypothesis H_0 about equal means on scores obtain in assignments delivered with different level of feedback could not be rejected even at $\alpha = .05$ with value of $F=1.77$ and $p\text{-value} = .1759$. Post hoc Tuckey analysis of p-values for pairwise t-tests confirmed this result.

Lack of impact of type of feedback on the mean score is clearly visible on the Fig. 1 .

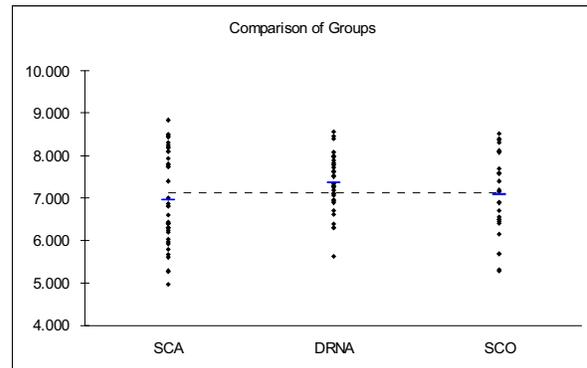


Fig. 1. Distribution of means of scores between three levels of feedback

Similarly, One-Way ANOVA F-test was used for verification of significant difference in the mean scores obtained on online assignments administered with different form of **randomization** (treatments). Surprisingly, mean scores on assignments with and without randomized questions shown even more uniformity. Null hypothesis H_0 about equal means on scores obtain in assignments delivered with and without randomized questions could not be rejected even at $\alpha = .05$ with value of statistics $F=0.60$ and $p\text{-value} = .4406$.

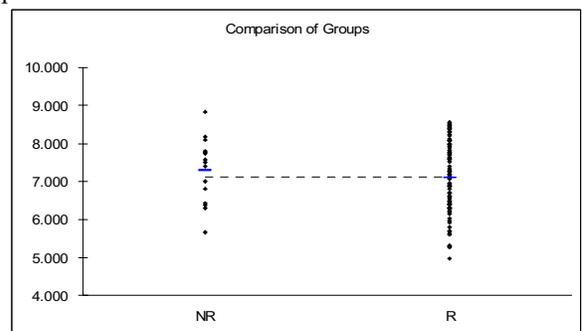


Fig. 2. Distribution of means of scores between two types of questions randomization

Lack of impact of questions randomization on the mean score is clearly visible on the Fig. 2 above. Post hoc Tuckey analysis of p-values for pairwise t-tests confirmed this result.

The next test utilized Randomized Block Design experiment with blocks identified as two different levels of **randomization** (R and NR) and treatments as three levels of **feedback**. Its intention was to remove any variance between investigated means that could be possibly caused by the fact that some assignments used randomized questions and some did not. Again, null hypothesis H_0 about equal means on scores obtain in assignments delivered with combined settings of randomization and feedback could not be rejected even at $\alpha = .05$. Value of $F=0.20$ for treatments (level of feedback) and $F=0.33$ for blocks (randomization) with p-values equal respectively 0.83 and 0.62 would clearly indicate statistically solid uniformity of means.

Combined and clearly inconsistent impact of feedback and randomization on the means of scores is shown below in Fig. 3.

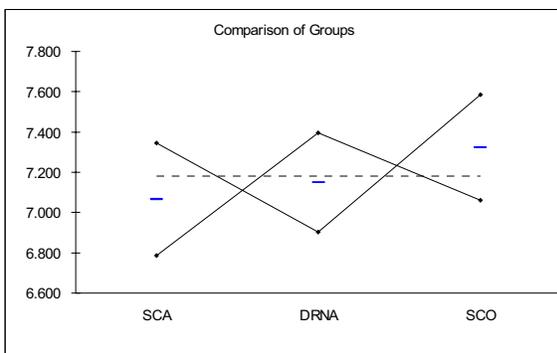


Fig. 3. Comparison of impact of feedback and randomization on mean scores

Finally, the Two-Way Analysis of Variance experiment was conducted with one factor (1) defined as a **form of instructional delivery** (web-based versus web-enhanced) and another factor (2) defined as levels of **feedback**. A purpose of this test was to establish if an additional factor such as specific format of a class (web-based versus web-enhanced) played any role in distribution of means of scores. Since overall objective of this study was to verify degree of integrity and determine scale of cheating and cooperation between students taking tests online, it was reasonable to expect that more cooperation will take place between students who take web-enhanced class and meet physically twice a week in the classroom. More cooperation should translate into higher scores in online testing. However, here also null hypothesis H_0 about equal means on scores obtain in assignments delivered with

different formats – web-based and web-enhanced could not be rejected even at $\alpha = .05$.

Obtained values of statistics F and respective p-values for Factor 1 (format), Factor 2 (feedback), and Interactions were as follows:

	F	p-Value
Factor 1 (format)	0.07	.7883
Factor 2 (feedback)	1.96	.1498
Interaction	0.60	.5521

It is clear that neither of the two Factors nor Interactions between them have any significant impact on difference between the means of scores. This would imply that no interaction experienced by students in a regular (web-enhanced) class is used for a purpose of cooperation/ cheating when taking online quizzes and tests. Some interesting fact has been noticed however when analyzing a chart of means spread in Interaction Plot by Factor 1 – see Fig. 4 below.

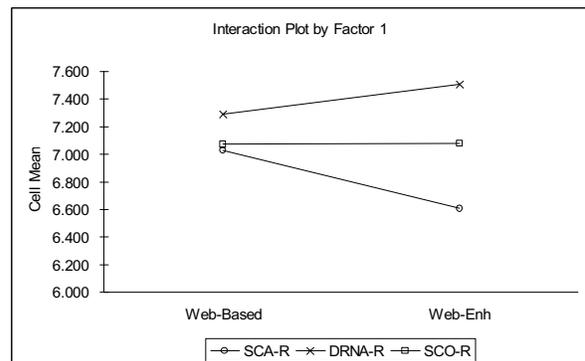


Fig. 4. spread of mean scores for web-based and web-enhanced class format

It seems that mean scores are much more spread out in the web-enhanced classes than in web-based classes. Moreover, the feedback type that seems to deliver lower mean is SCA-R, which is “showing correct answer in a randomized questions test”. One possible explanation could be that there are some attempts of sharing the answers between students, which without adjusting for randomization could have an adverse effect on the scores. This however needs more study and by no means is changing the fact, that there is no statistical base to imply any difference between the means of the scores due to cheating.

All tests were conducted based on statistically sound samples collected through delivery of online tests and quizzes as well as using some quantitative analysis tools available as part of course statistics in Blackboard [9]. We are assuming that Blackboard being one of the most popular LMS platforms is a good representation of e-learning environment.

4. Preliminary Results

Preliminary results seem to contradict couple of myths to which academic community often prescribes:

- in general, delivery of quizzes and tests in an online/ open book format does not seem to be conducive to cheating as it does not lead to variations in scores obtained by students under different assessment setups,
- it appears that making answers to questions available to students right after completion of an assessment (treatments SCA) does not have statistically significant impact on average score regardless whether questions were randomized or not,
- randomization of questions when delivering an online quiz or test does not cause statistically significant difference in the means of scores,
- there is no significant difference between the mean scores of online tests between web based (fully online) students and those using web-enhanced format, which would imply that a better chance for cooperation in the classroom settings does not necessarily translates into cheating when taking online tests. However, bigger spread among means for web-enhanced format requires some more study and may be an indication for some ever so misguided cooperation between students taking online tests.

There are some additional data available to author, which will require some continuing research.

5. Conclusions

An overall conclusion should perhaps be formulated in the following way: an average student taking an online class is less mischievous and interested in cheating as he/she is overworked, disconnected and ill organized to be an effective cheater. Cheating and abusing online testing environment can be easily made very time consuming and difficult for students by a skillful instructor. Randomization of the questions seems to have a minimal effect on mean scores, whereas revealing answers upon completion of the assignment does not increase possibility of cheating.

This study is on going and covers period from Fall of 2003 till present. Therefore, it will be possible to reformulate it in the future into longitudinal study and observe possible changes in the mean scores over time.

It seems that complexity of the material studied has much more to do with the mean scores than with a format of online tests measured in terms of “degree of difficulty in cheating.” This would indicate that online testing format of e-learning does not necessarily diminish quality of educational experience and does not compromise integrity of assessment process. This perhaps can not be said about online content delivery that most likely is depriving recipients of a significant “value added” associated with face to face lecturing and interaction with the instructor [10]. This may hopefully be alleviated overtime with ever more powerful e-learning technology.

6. References

- [1] Zhang, D., Nunamaker, J., “Powering E-Learning In the New Millennium: An Overview of E-Learning and Enabling Technology”, *Information Systems Frontiers*, 2003, 2, (5), pp: 207 - 218
- [2] Jones, N., O’Shea, J., “Challenging hierarchies: The impact of e-learning”, *Higher Education*, 2004, 48, (3), 2004, pp: 379 - 395
- [3] Adams, J.C., Armstrong, A., Web-based testing: A study in insecurity, *World Wide Web*, Volume 1, Issue 4, 1998, Pages 193 – 208
- [4] Hodgins, W., “Learnativity: Into the Future”, in *E-learning Handbook*, McGraw Hill, New York 2002, pp.38-64
- [5] Dobbs, K., “The State of Online Learning – What the Online World Needs Now: Quality”, *Training*, 9, (2), 2000
- [6] Hoffman, B., Ritchie, D., An instructional design-based approach for developing online learning environment”, in Kahn, B. (ed.), *Web-Based Training*, Educational Technology Publications, Englewood Cliffs, NJ, 2001
- [7] Wei-Fan Chen, "Effect of web-browsing interfaces in web-based instruction: a quantitative study," *Education, IEEE Transactions on*, vol.48, no.4 pp. 652- 657, Nov. 2005
- [8] Albrecht, K., Gunn, R., “Digital Backlash”, *Training and Development*, 11, (3), 2000, ASTD Publications
- [9] Chia-I Chang, F., “Quantitative Analysis of Distance Learning Courseware”, *Multimedia Tools and Applications*, 2003, 20, (1), pp: 51 - 65
- [10] Stella, A., Gnanam, A., “Quality assurance in distance education: The challenges to be addressed”, *Higher Education*, 2004, 47, (2), pp: 143 – 160