

Post-Graduate Student Performance in ‘Supervised In-class’ versus ‘Unsupervised Online’

Multiple Choice Tests: Implications for Cheating and Test Security

Richard K. Ladyshevsky,

Curtin Graduate School of Business, Curtin University, 78 Murray Street, Perth, Western Australia, Australia 6000.

Email. rick.ladyshevsky@gsb.curtin.edu.au

Profile: http://business.curtin.edu.au/courses/cgsb/staff_profiles.cfm/Rick.Ladyshevsky

Richard K. Ladyshevsky is a Professor of Managerial Effectiveness and Director, Online Teaching and Learning at the Curtin Graduate School of Business, Curtin University and Fellow of the Higher Education Research Development Society of Australasia. His research interests include scholarship of teaching and learning, leadership and management effectiveness and peer coaching and professional development.

Post-Graduate Student Performance in ‘Supervised In-class’ versus ‘Unsupervised Online’ Multiple Choice Tests: Implications for Cheating and Test Security

Abstract

This research explores differences in multiple choice test (MCT) scores in a cohort of post-graduate students enrolled in a management and leadership course. A total of 250 students completed the MCT in either a supervised in-class paper and pencil test or an un-supervised online test. The only statistically significant difference between the nine test scores was for one test where the students scored significantly lower in the unsupervised online test. There was no increase in mean test scores over time and the mean test scores for the unsupervised online test were not significantly higher than the mean test scores for the supervised in-class test. The study suggests that unsupervised online MCTs can be a viable tool for assessing knowledge in post-graduate students provided they meet best practice standards for online assessment. Concerns about increased cheating in unsupervised online testing is not supported.

Key Words: online assessment, e-testing, academic integrity, proctored versus un-proctored testing

Introduction

The growth in online distance education enrolments is accelerating at a rapid pace. In the United States, during the fall semester of 2002, only 9.6 percent of students in post-secondary institutions were undertaking at least one distance education course which rose to 29.3 percent in the fall semester of 2009 (Fask, Englander, & Wang, 2014). A post-secondary institution is typically a university, college, vocational or technical college for adults or graduates of a secondary school. A secondary school offers education from kindergarten through to grade 12 or 13. A more recent survey places this at 34 percent in the United States of America with online enrolments increasing five times faster than total enrolments (Bailey, Barton, & Mullen, 2014). Online learning has created significant flexibility for learners looking to further their education whilst balancing work and life demands. Online learning has also become critical to the ongoing success of many higher education providers

(Trenholm, 2007). Further, the advent on massive open online courses (MOOCs) is also revolutionizing how we think about educational delivery.

In business schools, particularly at the post-graduate level, online courses have proliferated and many degrees can be obtained in-part or fully online. Consumers of these courses are typically busy individuals with full time jobs, have family commitments and an urgency to fast track their academic accomplishments so they can advance their careers. Many business schools must reconsider their strategy around educational delivery to accommodate these learners (Bailey, Barton, & Mullen, 2014). Once such strategy that has changed considerably is student assessment. For many institutions, the desire to expand online markets to secure funding has meant they have had to develop retention strategies for students who want flexibility in their study plans (Fask et al., 2014). This has led to more online unsupervised testing which the students can complete off campus at a time that suits their schedules (Schultz, Schultz, & Round, 2008).

One question that emerges with the use of unsupervised online testing (e.g. multiple choice tests MCTs) is the possibility that cheating is greater (Fask et al., 2014; Styron & Styron, 2010). While there is some research which suggests this possibility, there does not appear to be any conclusive evidence that these new technologies have led to an increase in cheating (Fask et al., 2014; Underwood, 2006). However, the literature broadly suggests that faculty have concerns regarding the greater potential for cheating in online courses (Fask et al., 2014). In one study, which surveyed over 1900 undergraduate students, evidence suggests that students were more likely to cheat in online courses compared to traditional courses (Grijalva, Nowell, & Kerkvliet, 2006). Yet in another study of students in both public and private college programs the opposite was found (Dille', 2011). In this study a larger percentage admitted to cheating in traditional courses than in online courses. Further, those who admitted to cheating in traditional classes were seven times more likely to cheat in online classes. The only significant demographic variable was related to age. Younger students admitted to cheating more often. The authors conclude that cheaters are going to cheat regardless of the mode of instruction (Dille', 2011). The problem with using self-report surveys to assess the extent to which

cheating is occurring in online and traditional environments may not yield valid results even with assurances of anonymity (Fask et al., 2014).

What appears to be occurring is that student cheating as a whole has seen an increase in higher education (Styron & Styron, 2010). What constitutes 'cheating', however, is also changing.

Traditional views of cheating by Universities may not be in keeping with the digital era of learning which involves greater and greater use of open-source collaboration and the ready sharing of ideas, knowledge and information (Harkins & Kubik, 2010). These authors even go as far as discerning a difference between 'cheating' and 'ethical cheating' in keeping with changes in information accessibility and availability.

Online Testing

Online testing involves several components which include supervisory arrangements if any, the delivery of the instrument, the grading and analysis of the test, and the testing parameters and process (Erturk, Ozden, & Sanli, 2004). This can occur in a computer assessment centre or through stand-alone computers. While there are many types of electronic assessments and tests that can be created for online courses, this research explores the use of multiple choice question (MCT) testing.

Examples of these other assessments include asynchronous discussions, e-presentations, wikis, blogs, e-portfolios and virtual simulations to name a few (Styron & Styron, 2010). In many of these assignments, cheating can be controlled by creating more subjective and personal assignments, which make copying from other sources harder.

Cheating on online multiple choice tests, particularly if it is unsupervised, may increase in incidence (Trenholm, 2007). Hence, to control for cheating during these types of tests, some universities have set up computer testing centres. However, supervision of the test is still required and the cost of setting up this type of system, and associated software, can be expensive. These on-campus computer testing centres are typically located on the campus and require students to travel to the institution so they can appear on a set day and time for their test. This doesn't provide a solution for students who are studying online at the university and live interstate or overseas from the host institution.

With the advent of learning management systems tests such as the MCT have evolved to online formats where the student can complete the test from their own personal computer. Movement towards online testing by academics is increasing because it can reduce the effort required for grading and can allow for more frequent and flexible testing arrangements (Erturk et al., 2004). Several other benefits of online testing have been identified (Maguire, Smith, Brallier, & Palm, 2010). These benefits include reduced instructional time dedicated to assessment and testing, more flexibility in scheduling and administering tests, rapid test item analysis and increased test security, and increased speed of feedback to instructors and students. There is also evidence of studies finding that students prefer online tests to written ones (Maguire et al., 2010).

These learning management systems have also streamlined many approaches to online testing (Prince, Fulton, & Garsombke, 2009). For example, having in-built testing systems within the learning management system overcomes issues of time and geography for students because they can access the test from their own computer provided they have a reliable internet connection. However, these unsupervised assessment strategies have introduced other complicating factors which are described below.

Cheating and Online Testing

Increasing cheating in online courses can be explained in part by the levels of anonymity online courses can provide, changing views on the morality of cheating, increasing pressure to have higher degrees to remain competitive in modern society, and the growth and availability of online courses that can 'fit' into peoples' busy lifestyles (Trenholm, 2007). Students cheat for a range of reasons (Hensley, 2013) and these reasons are not limited to online courses. For example, Hensley (2013) notes that common reasons include low grades and ineffective study strategies, poor time management skills, personal values and views which relate to achievement, fear of punishment, class attendance and peer pressure. Performance based pressure and student competition also leads to cheating in comparison to environments which emphasize learning and collaboration. Other reasons described in the literature based on a survey of over 2300 students in the USA and Israel, includes

extrinsic versus intrinsic motivations to learn, and age, with younger people more likely to cheat than older individuals (Peled, Barczyk, Eshet, & Grinautski, 2012). Further, students in face-to-face courses had a statistically higher propensity to cheat than their counterparts in distance learning courses.

The most significant issue associated with using unsupervised multiple choice testing within a learning management system is student authentication (Maguire et al., 2010; Schultz et al., 2008) and the possible consequences of cheating (Trenholm, 2007). Cheating during an unsupervised online test can involve a myriad of strategies from ranging from using books, searching on the internet, being on the phone with a classmate, having your classmates with you when you complete the test or having someone else do the test on your behalf. Digital technologies in and of themselves have not necessarily generated an increase in academic malpractice. What they have done however, is create more ways in which to cheat (Underwood, 2006).

While supervisors can be nominated or hired to supervise the student taking the test, this can become administratively complex and time consuming if you have students all over the world taking the course. While there is work being undertaken to develop technologies that can authenticate individuals using web cameras and remote supervisors (Kolowich, 2013), or finger print and identification checking mechanisms and retinal scanning, these are costly and often raise questions about personal privacy (Trenholm, 2007). Other products, when integrated within the learning management system, can lock students into the testing environment once they log in to the test. When locked in, they cannot print, use print screen or other capture devices. Copying and pasting is disabled along with screen sharing and right-click options. These functions are only disabled when the student submits the test for grading.

Other approaches have also been suggested to prevent cheating, some of which relate to all forms of assessment and others more specifically to online testing. One important factor which is noted extensively in the literature is the presence of an institutional statement and guidelines describing academic integrity and the consequences of cheating (Hensley, 2013; Maguire et al., 2010; Underwood, 2006). Further, for those who breach these guidelines appropriate action must be taken so

students see that there are real consequences to cheating. Surprisingly, despite academic concerns about cheating in online assessments, there is ample evidence to suggest that faculty often do not take aggressive action to combat student cheating in online courses (Fask et al., 2014).

Other strategies to help reduce cheating in online MCTs relate to the design and value of the test. An online assessment worth half of a course grade may entice students to cheat whereas an online MCT worth a quarter of a course grade may not invoke the same level of energy to cheat. Some of the inbuilt tools for online testing can also reduce the efficacy of cheating. Setting a reasonable and fixed amount of time to answer each question precludes students from having extra time to call a friend or copy the question (Schultz et al., 2008). Permitting only one question to be viewed at a time and preventing backtracking will prevent students from going back to questions if they finish early and copy them. Having the test questions drawn randomly from a larger pool of questions which are stored in the online test folder is a common practice. Students end up with different overall tests and questions which appear in different orders, making collusion (sitting together to do the test) more difficult (Maguire et al., 2010; Prince et al., 2009; Trenholm, 2007; Underwood, 2006), particularly if there is a time limit on each question. Questions which encourage critical thinking rather than knowledge regurgitation are also more effective as they are harder to cheat on (Harper, 2006).

The benefits of using these measures to increase the security of the online MCT means the test can be made available over a set time period, for example, 24 – 48 hours. This ensures all students, regardless of geography or time zone, are able to find a time to complete the online MCT - an important factor in promoting flexible learning and assessment.

Supervised versus Unsupervised Online Testing

Reviews of the literature from the 1980s through to the 1990s during the early stages of online testing indicate that online test scores tended to equate to paper administered test scores when the testing environment was similar (e.g. supervised) (Alexander, Bartlett, Truell, & Ouwenga, 2001; Bugbee Jr., 1996). More recent studies have explored the efficacy of online testing. One study examined a group of students receiving the same in-class instruction in accounting (Maguire et al., 2010). Forty three students completed an online test in a supervised computer testing centre. The other 92 students

completed the equivalent supervised paper-and-pencil test. Students in the computer testing centre did significantly better than cohort sitting the paper-and-pencil test. The study confirmed that computer based testing did not in itself present limitations on performance compared to the traditional testing format.

Another study explored differences in 75 student test scores for a paper based versus computer based test in several undergraduate accounting courses (Anakwe, 2008). All courses were taught by the same instructor and all students had previously taken at least one online test. Each student was given two computer based tests, two paper based tests, and a final in-class assessment. The four tests were multiple choice in nature involving 25 questions. Each test was of 50 minutes duration. The computer based tests and in-class tests were supervised. There were no significant differences across the methods of assessment in the students' test scores.

Other studies have explored supervised versus unsupervised online tests. In one study four classes (n=76) in three different disciplinary subjects within a school of management ran supervised and unsupervised online tests (Prince et al., 2009). All students were online students and the online test questions were drawn from a randomised pool of questions. The means for the supervised tests were significantly lower than for the unsupervised tests. To control for instructor, test scores of two classes of the same course taught by the same instructor were analysed. Similar patterns were seen.

A contrasting study found no statistically significant differences in MCT scores between undergraduate finance students taking an unsupervised online test versus those taking the supervised test in-class (Peng, 2007). The e-test was available over 30 hours, took 80 minutes to complete and questions were allocated randomly from a pool of questions. The value of the test was only worth 10 percent of the overall grade and the authors note that this in itself may have been a deterrent to cheating. Another study examined 850 grades from a variety of undergraduate mathematics distance education courses and found no statistically significant differences between the 406 students who completed an in-class supervised examination and the 444 students who completed their examination entirely in an unsupervised online environment (Yates & Beaudrie, 2009).

In contrast to these two studies, Harmon and Lambrinos ran four online MCTs in an economics course across two different study periods (Harmon & Lambrinos, 2008). The students could complete each MCT over a three day window. All questions were randomised and once the student logged in they had 60 minutes to complete the test (90 minutes for the final fourth test). In the second administration of the course, the only difference was that the final fourth test was conducted as an in-class supervised assessment. Using a mathematical model (regression and GPA as a predictor of assessment scores) they inferred that the variation in scores between the supervised and unsupervised test results being due to cheating in the latter group.

Looking at post-graduate students, one study tracked the performance of 114 masters and doctoral level students whom undertook an online unsupervised test or a traditional test in a supervised setting (Rakes, Fields, & Cox, 2006). Those in the traditional supervised test setting ($n = 42$) scored significantly higher than those taking the test online. Another study involving post-graduate students explored overall course grades of students who had courses with supervised and unsupervised tests (Schultz et al., 2008). In all four courses investigated there was no significant difference between final grades between courses with supervised assessments compared to those with unsupervised assessments, although in 3 of the 4 courses the mean grades were higher in the supervised assessment group.

Another study involved graduate students in a pharmacy program (Pierce, 2012). The face to face cohort comprised of 213 students who experienced didactic instruction via recorded lectures to review at their leisure. The other cohort were practicing pharmacists studying for a Doctor of Pharmacy degree and this course was delivered exclusively online. The content of the curriculum was identical. All students completed the same online test, however the students in the face to face group sat the exam with a supervisor present. The Doctoral students completed the online test unsupervised. They also were only given 1.5 – 1.7 minutes per question and the test locked out when the time expired. For three of the four tests, there were no significant differences in the test scores between the two cohorts. The fourth test (infectious diseases) demonstrated a significant difference between the test scores in

favour of the Doctoral cohort who completed the unsupervised online test suggesting there may be other extraneous variables at play.

In a study involving physiology students (Kibble, 2007), all were provided with unsupervised formative multiple choice questions which they could use to prepare for more formalised summative test at a later date. Students who received no credit for the formative test but used them to study for their summative test did best on the final summative assessment. Those students who received credit for the formative tests, to offset their final mark for the summative assessment, did less well on the final assessment, and were found to be 'cheating' by using other resources to answer the questions so they could get 90 – 100% on the formative test. This study provides some suggestion that the value of a test in terms of its overall final contribution to the course grade can influence cheating.

In another study, students in the same class could complete their online test unsupervised or in a traditional classroom environment at a set time with a supervisor (Granger & McGarry, 2003). Both groups were able to use books and notes during the test. They were allocated the same amount of time to complete the test. The 29 students who took the in-class test had a mean score of 97 percent whereas the 31 students in the online cohort had a mean of 95 percent.

These differences in test scores between traditional and online tests in supervised and unsupervised settings produce mixed results and there are likely several factors influencing these findings.

Research appears to demonstrate that students in unsupervised online testing environments do less well in comparison to students taking in-class supervised tests (Fask et al., 2014; Hollister & Berenson, 2009) although there is significantly more variation in scores for those in unsupervised settings even when the student cohorts are controlled for grade point average (Hollister & Berenson, 2009). The reasons for this lesser performance in unsupervised online tests appears to relate to confounding effects such as not being able to ask the instructor for clarification regarding a question, ambient noise and distractions in the testing environment (home, dormitory office), and/or computer and connectivity issues. These confounding effects have been demonstrated to be statistically significant and may be masking cheating effects in studies that only look at test results (Fask et al., 2014). The reasons why students in supervised online tests may do better relates to the accountability

inherent in being supervised during a test. This may have an indirect effect on student learning and performance particularly if the supervisor is their course instructor (Wellman, 2005).

Research Question

This research explores whether unsupervised online testing leads to increases in possible student cheating in a sample of post-graduate business students by examining MCT scores across time. The educational community has called for further studies investigating the possibility of increased cheating in online courses (Fask et al., 2014; Styron & Styron, 2010). In particular, at the post-graduate level where there is limited research on testing in online environments (Liu, 2008). Most of the available research in this area is either anecdotal in nature and/or focussed at the community college/undergraduate level. There are limited amounts of research on online testing in comparison to traditional testing particularly where supervision may or may not be involved (Maguire et al., 2010). Further research on online testing is called for as there have only been a handful of studies over the past 15 years that have used statistical methods to investigate cheating in online tests (Fask et al., 2014). Much of the research on online cheating looks at either survey methods or copying information from the web rather than exploring how students may cheat in multiple choice knowledge tests (Trenholm, 2007).

Hence, the null hypothesis to be tested in this research is that there will be no significant differences in mean test scores between those students taking supervised in-class MCTs and students taking unsupervised online multiple choice tests. This hypothesis tends to support what is written in the literature but is specific in this case to post-graduate students. The second null hypothesis is that there will be no progressive increase in the mean test scores over time. Increases in test scores could suggest a breach of test security. One reason for this increase might be that a student had copied questions from the test and the questions were now circulating amongst the student body.

Method

Post-graduate students enrolled in a management and leadership business course were the sample of convenience used in this study. The students attended classes in face to face mode. The course was

offered either once a week in the evening for three hours, or in a blended-intensive mode over two to three weeks (e.g. full day classes two or three times per week). The course content and assignments were the same over the course of this study although there were different faculty teaching the course in three of the nine course offerings. The courses ran for 14 weeks with students completing their assignments within this time frame. The students in the course are typically in their mid-thirties and are working full time. The gender ratio in the course is approximately 60 percent male and 40 percent female. About 10 percent of the students in the course would be international students and the other 90 percent domestic students. The international students in this cohort require a higher English language score (e.g. TOEFL) in comparison to other post-graduate business programs in the faculty. In one case, the course was run in a semester format as a service course for another Master's program. These students generally have a lower TOEFL score for admission and enrolments in this class have a higher percentage of international students.

Students were assessed on their applied knowledge using a MCT format. Each multiple choice question had a short case scenario with four options which required students to demonstrate some critical thinking as the answer had to be applied to a business scenario.

James Arlett is classified as an ISTJ (introvert, sensing, thinking, judging) according to the Myers Briggs Personality Indicator. His dominant function is Sensing (in the introverted mode). During highly stressful situations, he goes into a 'grip' response. Given his dominant function of 'Sensing' what behavioural pattern is James most likely to exhibit when he is highly stressed?

1. Crying, sadness, despondency
2. Vindictive, harsh, hyper-critical
3. Negativity, catastrophizing, awfulizing
4. Obsessive, drawn to escapist behaviour

Prior to the administration of the MCT, each question (n = 49) was assessed for its overall ability to discriminate the students' level of knowledge. The 50th question was a question which all students could answer as it was given as a bonus mark. Table 1 provides an excerpt of a matrix where each question was scrutinised by each student's response. Where questions were found to be too difficult or ambiguous because a large number of students got them wrong (e.g. Question 7), they were reviewed to increase clarity and improve discrimination. Questions that were seen to be 'easy' because very few

students got them wrong (e.g. Question 14) were reviewed and adjusted accordingly. A large number of correct responses was not necessarily a problem in and of itself, particularly if it was addressing a core aspect of knowledge which was covered well in the course.

Table 1 – Excerpt from Test Item Analysis Matrix

St.	Qu																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	...	49
1					x		x				x	x			x		
2				x			x										
3						x					x	x		x			
4			x	x			x					x			x		x
14...	x		x				x					x					x
40	x				x	x	x				x						
41						x	x	x		x	x	x	x				
42	x			x	x	x	x	x	x								
43		x												x			
44						x	x										
45							x				x				x		
46	x		x				x					x			x		
47		x	x			x	x	x		x		x	x		x		x
48	x	x					x						x				
49	x			x			x	x		x	x		x				
X	17	8	12	11	14	18	45	17	8	7	21	22	10	4	7		9
%W	35	16	24	22	29	37	92	35	16	14	43	45	20	8	14		18

X = number wrong; %W = percentage of questions wrong

The MCT once reconstructed after this pilot test, was administered over eight trimesters and one semester. The first four MCTs were paper-and-pencil tests which were completed in a secure environment under supervision. Test security was very high. At the end of the MCT students had to submit both the test booklet with the questions and their answer sheet with their names registered on both documents in order to receive a grade. There were no breaches of test security during these four administrations and as far as the researcher is aware, students did not have previous copies of the test. The MCT was worth 25 percent of the overall course grade and was comprised of 50 multiple choice questions. The students were given 130 minutes to complete the test or just over two and half minutes per question.

The next five administrations of the MCT were done online as part of a purpose built online test in Blackboard. Many of the suggested ways of reducing cheating in online tests were incorporated in to the design of the e-test (randomization, time limits, no backtracking). There were 25 multiple choice

questions on the online test. These were randomly drawn from the original pool of 50 test items used to test the students in the traditional paper-and-pencil test. Students were given a four day window to log in and complete the online multiple choice test. When they logged in they had 75 minutes to complete the MCT or three minutes per question. There was no backtracking. The test was unsupervised.

Results

The results of the nine tests, including the number of students in each group, means (averages), standard deviation, standard error, lower and upper bounds for the confidence interval of 95 percent and minimum and maximum scores are illustrated in Table 2. A total 136 students completed the in-class (IC) supervised MCT and another 114 students completed the unsupervised online MCT (eT). Courses taught by a different instructor to the principal instructor are illustrated with a * in Table 2.

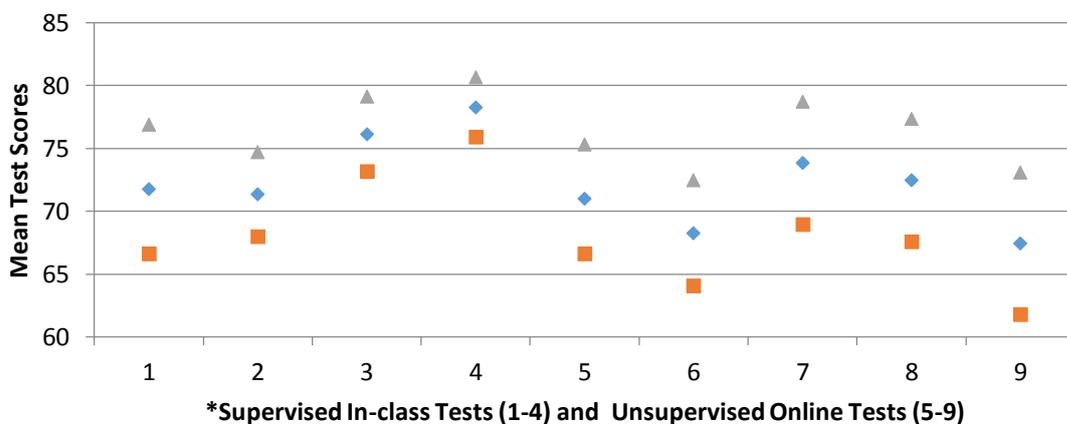
Table 2: All Test Scores

Test	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
IC1	24	71.7500	12.16284	2.48273	66.6141	76.8859	48.00	96.00
IC2*	38	71.3158	10.20097	1.65482	67.9628	74.6688	48.00	88.00
IC3	34	76.1176	8.51297	1.45996	73.1473	79.0880	52.00	90.00
IC4	40	78.2500	7.41706	1.17274	75.8779	80.6221	62.00	94.00
eT5	23	70.9565	10.03393	2.09222	66.6175	75.2955	52.00	88.00
eT6*	32	68.2500	11.58141	2.04732	64.0745	72.4255	40.00	88.00
eT7	22	73.8182	11.02221	2.34994	68.9312	78.7052	44.00	92.00
eT8	18	72.4444	9.78728	2.30688	67.5773	77.3115	52.00	88.00
eT9*	19	67.4211	11.66817	2.67686	61.7972	73.0449	40.00	84.00
Total	250	72.7000	10.56221	.66801	71.3843	74.0157	40.00	96.00

The results of the one way ANOVA for all nine tests indicate significant differences in test scores $F(8, 241) = 3.628, p = .001$. The main reason for these differences is due to the last test (eT9) having a relatively low mean test score in comparison to the rest of the mean test scores.

As seen in Figure 1 below, test scores vary each time the MCT is administered and do not necessarily reveal any distinct patterns, for example, a progressive increase in mean test scores. Supervised in-class mean test scores (data points 1 – 4) show an increase in test scores from test 2 – 4. Test 2 was taken by the students with a different instructor teaching the course rather than the regular instructor. Unsupervised online test scores (data points 5 – 9) do not demonstrate any pattern although data points 7 – 9 are decreasing. Test 6 and 9 were taken by the students with different instructors teaching the course from the regular instructor. The three vertical figures for each data point in Figure 1 represent the mean test scores in the middle (diamond) and the upper (triangle) and lower (square) limits of a 95 percent confidence interval for these test scores.

Figure 1: Mean Test Scores & Confidence Ranges (upper & lower) – All Tests



*Tests 2, 6 and 9 taught by other instructors

The results for the courses taught by the regular instructor are given in Table 3. The results of the six tests, including the number of students in each group, means, standard deviation, standard error, lower and upper bounds for the confidence interval of 95 percent and minimum and maximum scores are illustrated in in Table 3. A total of 98 students completed the in-class (IC) Supervised test whereas 63 completed the unsupervised online e-test (eT).

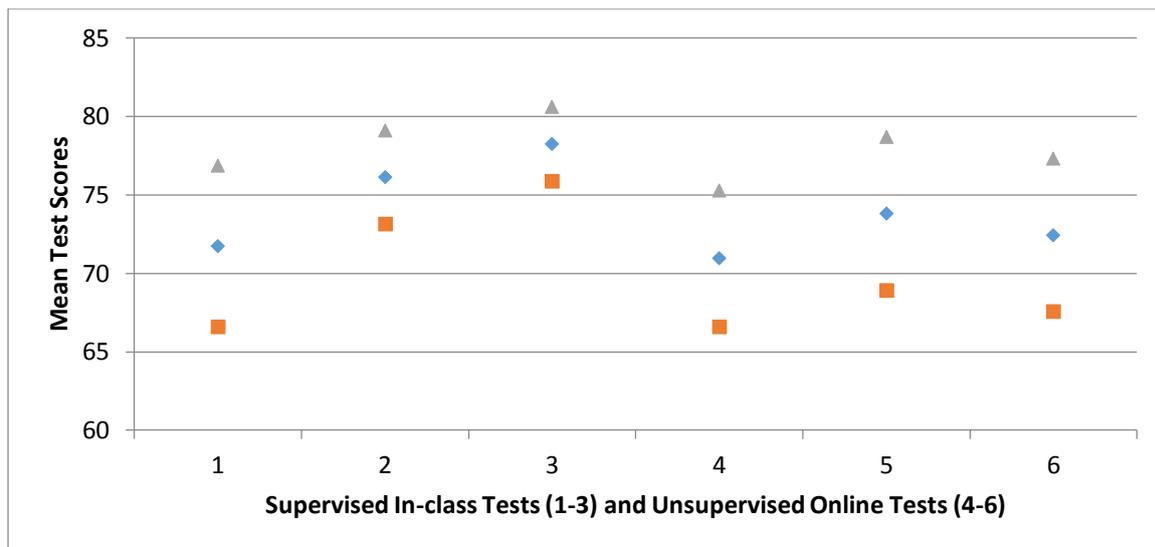
Table 3: Test Scores For Regular Instructor

Tests	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
IC1	24	71.7500	12.16284	2.48273	66.6141	76.8859	48.00	96.00
IC3	34	76.1176	8.51297	1.45996	73.1473	79.0880	52.00	90.00
IC4	40	78.2500	7.41706	1.17274	75.8779	80.6221	62.00	94.00
eT5	23	70.9565	10.03393	2.09222	66.6175	75.2955	52.00	88.00
eT7	22	73.8182	11.02221	2.34994	68.9312	78.7052	44.00	92.00
eT8	18	72.4444	9.78728	2.30688	67.5773	77.3115	52.00	88.00
Total	161	74.5342	9.85586	.77675	73.0002	76.0682	44.00	96.00

The results of the one way ANOVA for all six tests illustrate significant differences in test scores, $F(5, 155) = 2.612, p = .027$, but no particular upward or downward direction over time.

In Figure 2 below, test scores vary each time the MCT is administered and do not necessarily reveal any distinct patterns. In this figure the students taking these tests are receiving instruction from the same instructor who most regularly teaches the course. Supervised in-class mean test scores (data points 1 – 3) show an increase in test scores. Unsupervised online test scores (data points 4 - 6) do not demonstrate any pattern. The three vertical figures for each data point in Figure 2 represent the mean test scores in the middle (diamond) and the upper (triangle) and lower (square) limits of a 95 percent confidence interval for these test scores.

Figure 2: Mean Test Scores & Confidence Ranges (upper & lower) – Tests Controlled by Instructor



Discussion

The results of this research adds support to other studies that claim to have found no significant differences in test scores between supervised in-class tests and comparable unsupervised online tests. This pattern of no significance was demonstrable across 8 administrations of the MCT. The mean test score for the ninth administration of the MCT was particularly low, and produced a significant difference, but in a direction not suggestive of cheating. Hence, the null hypothesis for this study, that there will be no significant differences in student performance between supervised in-class MCTs and unsupervised online MCTs is supported. This results of this research is particularly noted against other studies that have investigated post-graduate students taking unsupervised online tests (Pierce, 2012; Rakes et al., 2006; Schultz et al., 2008) finding similar results.

Other than an increase in mean test scores for the supervised in-class MCTs (tests 2 – 4), there was no pattern of increasing test scores across all 9 administrations of the test. For the in-class tests, test security was excellent and the instructor(s) followed procedures ensuring all test booklets and answer sheets were returned against the student's name. All materials such as phones and bags were not allowed alongside the student during the supervised test. Hence, it is unlikely that old tests were circulating amongst the group or that students were using technology to cheat. Reasons for the progressive higher scores seen between MCTs 2 – 4 may be due to the instructor(s) teaching more towards the test over time once it was developed. The other may be instructor variability as the second

test administration was taken by students who had an instructor who was different from the regular instructor teaching the course. The absence of any increase in test scores in the unsupervised online tests (5 – 9) may be due to confounding elements in the online MCT which are not present in the in-class test which lowered test scores in the online testing group (Fask et al., 2014; Hollister & Berenson, 2009). Two of these five tests were also taken by students who had an instructor teaching the course who was not the regular instructor. Hence, variability in teaching and content coverage would have an impact on test scores. The randomisation of the questions in the unsupervised online MCT could have also resulted in a greater proportion of harder questions being sampled, and thus the lower scores in this group. There is also evidence that students in face to face classes who take supervised tests do better than students in online classes who take unsupervised tests because of the relationship between instructor and student and the greater sense of personal accountability in the testing centre (Wellman, 2005).

Other reasons for the results seen in this study may lie in the nature of the sample, the design of the test – particularly the online MCT and the setting. First of all, the sample were post-graduate business students. The mean age of the student body approximates the mid-30s. The research suggests that older students who are working and are more mature are far less prone to cheat than younger students (Prince et al., 2009; Underwood, 2006). These individuals are busy, most work full time, and have to balance work, study and family life. As they are time poor, they may not invest time into strategizing methods for cheating and just get on with their assessment tasks. The value of the test was also only 25 percent of their overall grade, and may have not been significant enough in value to warrant cheating. The online MCT was designed to incorporate many of the principles of best practice described in the literature. Questions were application based and linked to a business scenario which reduces memorisation of facts and potential sharing of answers (Eplion & Keefe, 2005). As the test was secure, students did not have copies of the right answers. Hence, they could not create a grapevine effect (Williams, Lloyd, & Simonton, 1990) where earlier test takers could share and distribute correct answers to later test takers. If this was occurring there would be an escalation of scores as the test progressed in time across all nine tests. The test items on the online MCT were also

randomly selected from a larger pool of questions and had time limits imposed on each question. Students were also not allowed to back track. All of these security measures appeared to be effective in mitigating cheating. As the in-class cohort could backtrack on their test, it is possible that part of their higher scores were due to being able to review and change their answers. The setting in which this study took place also has an academic integrity and plagiarism policy which is well advertised to students. Students are assigned penalties systematically for cheating and this would have also been a deterrent (Hensley, 2013; Maguire et al., 2010; Underwood, 2006).

The results, however, don't necessarily mean that the students did not cheat in the unsupervised online test. However, in a MCQ test there is usually limited time to access external resources because of the number of minutes allocated to each question. When the questions are scenario based and require more critical thinking, rather than memorization of facts, using external resources becomes even more difficult. Hence, allowing open book access in both the supervised in-class MCT and the unsupervised online MCT would likely not have much of an impact on mean test scores between groups, although it would be good to test this out in future research.

Another reason masking potential cheating in the unsupervised online testing environment may be due to confounding factors that influence test scores (eg. computer and network issues, peripheral noise and distraction, lack of ability to ask instructor for clarification of a question during the test, in built testing measures such as no back-tracking). As this study did not control for test environment, these confounding factors could lower scores, masking any actual cheating that occurred during the test. (Fask et al., 2014; Hollister & Berenson, 2009).

There is evidence to suggest that students in blended/intensive courses at the post-graduate level select these types of courses because of their need to balance work and lifestyle and to accelerate their program completion. Students who selected blended/intensive courses were the least concerned about learning benefits in comparison to those students who took traditional weekly face to face courses and fully online courses (Ladyshevsky & Taplin, 2013). In light of these learner effects, this may also be another reason for the lower scores seen in the unsupervised online test group, as the majority of the student in this group were enrolled in the blended/intensive version of the course.

Limitations and Suggested Further Research

There were limitations to this study. The first relates to the format of the test. The first four administrations of the test had 50 questions and the next five administrations only had 25 questions drawn from the same pool of 50 questions. The random sampling may have resulted in some tests being harder with a lesser range of questions to demonstrate true knowledge, thus driving down the scores in the unsupervised online multiple choice test. Second, there were two different instructors teaching the course in addition to the main instructor, which can introduce variation into final test scores depending on how well content is taught and covered. There was also one course out of the overall nine courses that had a larger international student cohort with lower TOEFL admission scores in comparison to the rest of the sample. However, when the study was controlled by instructor (which did not include the international student cohort), similar patterns of results are seen.

Future research might include test scores from post-graduate students in other courses who are studying fully online and take unsupervised multiple choice tests. Studies with larger sample sizes would also be beneficial across other post-graduate programs. Studies which control for other features such as literacy and language capabilities where there is a large international student cohort and cultural background would also be interesting. As online supervision tools and software programs become more effective and commonplace, the impact of these on test scores would be of great interest.

Conclusion

This study responds to the call for more statistically based investigations exploring differences in unsupervised online tests and supervised in-class tests (Fask et al., 2014; Maguire et al., 2010; Styron & Styron, 2010) particularly at the post-graduate level (Liu, 2008). The results of this research provides support to other studies that have found no statistically significant differences in mean test scores when these tests are administered in supervised in-class settings versus unsupervised online settings. Issues of test design, value of test, age and maturity of the student, and confounding factors may impact on these test scores and are important considerations when designing and implementing tests in unsupervised online settings. The results do suggest, however, that unsupervised online

MCTs can be carried out in the right circumstances and that fears about increased cheating may not be as great as feared. Cheaters seem to cheat regardless of the testing modality.

References

- Alexander, M., Bartlett, J., Truell, A., & Ouwenga, K. (2001). Testing in a Computer Technology Course: An Investigation of Equivalency in Performance between Online and Paper and Pencil Methods. *Journal of Career and Technical Education*, 18(1), 69-80.
- Anakwe, B. (2008). Comparison of Student Performance in Paper-Based Versus Computer-Based Testing. *Journal of Education for Business*, 84(1), 13-17.
- Bugbee Jr., A. (1996). The equivalence of Paper-and Pencil and Computer Based Testing. *The Journal of Research on Computing in Education*, 28, 282-290.
- Dille¹, E. (2011). *A Multi-Institutional Investigation into Cheating on Tests in College Online Courses*. (Doctor of Philosophy), University of South Carolina, South Carolina. (UMI 3488362)
- Eplion, D., & Keefe, T. (2005). *Best Practices for Preventing Cheating on On-Line Exams*. Paper presented at the World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education, Chesapeake, Virginia.
- Erturk, I., Ozden, Y., & Sanli, R. (2004). Students' Perceptions of Online Assessment: A Case Study. *Journal of Distance Education*, 19, 77-92.
- Fask, A., Englander, F., & Wang, Z. (2014). Do Online Exams Facilitate Cheating? An Experiment Designed to Separate Possible Cheating from the Effect of the Online Test Taking Environment. *Journal of Academic Ethics*, 12(2), 101-112.
- Granger, M., & McGarry, N. (2003). *Incorporating On-line Testing into Face-to-Face Traditional Information Systems Courses*. Paper presented at the International Academy for Information Management, Annual Conference: International Conference on Informatics Education & Research.
- Grijalva, T., Nowell, C., & Kerkvliet, J. (2006). Academic Honesty and Online Courses. *College Student Journal*, 40(1), 180-185.
- Harkins, A., & Kubik, G. (2010). "Ethical" Cheating in Formal Education. *On the Horizon*, 18(2), 138-146.
- Harmon, O., & Lambrinos, J. (2008). Are Online Exams an Invitation to Cheat? *Journal of Economic Education*, 39(Spring), 116-125.
- Harper, M. (2006). High Tech Cheating. *Nurse Education in Practice*, 6, 364-371.
- Hensley, L. (2013). To Cheat or Not to Cheat: a Review with Implications for Practice. *The Community College Enterprise*, Fall, 22-34.
- Hollister, K., & Berenson, M. (2009). Proctored versus unproctored online exams: studying the impact of exam environment on student performance. *Decision Sciences Journal of Innovative Education*, 7(1), 271-294.
- Kibble, J. (2007). Use of Unsupervised Online Quizzes as Formative Assessment in a Medical Physiology Course: Effects of Incentives on Student Participation and Performance. *Advances in Physiology Education*, 31, 253-260.
- Kolowich, S. (2013). Behind the webcam's watchful eye, online proctoring takes hold. *The Chronicle of Higher Education*. <https://chronicle.com/article/Behind-the-Webcams-Watchful/138505/>
- Ladyszewsky, R., & Taplin, R. (2013). Factors influencing mode of study preferences in post-graduate business students. *International Journal of Management Education*, 11(1), 34-43. doi: <http://dx.doi.org/10.1016/j.ijme.2012.12.001>
- Liu, S. (2008). *Design and Use of Assessment Tasks in Online Graduate Courses: Instructors' Practices, Reflections, and Perceptions*. (Doctor of Philosophy), Indiana University, Indiana. (UMI 3330791)

- Maguire, K., Smith, D., Brallier, S., & Palm, L. (2010). Computer-Based Testing: A Comparison of Computer-Based and Paper-and-Pencil Assessment. *Academy of Educational Leadership Journal*, 14(4), 117-125.
- Peled, Y., Barczyk, C., Eshet, Y., & Grinautski, K. (2012). *Learning Motivation and Student Academic Dishonesty - A Comparison Between Face-to-Face and Online Courses*. Paper presented at the Chais Conference on Instructional Technologies Research, Raanana, Israel.
- 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.
- Pierce, R. (2012, June). *Proctored Versus Non-Proctored: How Does Assessment Setting Affect Student Achievement on Web-Based Assessments?* Paper presented at the World Conference on Educational Multimedia, Hypermedia and Telecommunications, Denver.
- Prince, D., Fulton, R., & Garsombke, T. (2009). Comparisons of Proctored Versus Non-Proctored Testing Strategies in Graduate Distance Education Curriculum. *Journal of College Teaching & Learning*, 7(6), 51-62.
- Rakes, G., Fields, V., & Cox, K. (2006). The Influence of Teachers' Technology Use on Instructional Practices. *Journal of Research on Technology in Education*, 38(4), 411-426.
- Schultz, M., Schultz, J., & Round, G. (2008). Online Non-Proctored Testing and its Affect on Final Course Grades. *The Business Review, Cambridge*, 9(2), 11-16.
- Styron, J., & Styron, R. (2010). Student Cheating and Alternative Web-Based Assessment. *Journal of College Teaching*, 7(5), 37-42.
- Trenholm, S. (2007). A Review of Cheating in Fully Asynchronous Online Courses: A Math or Fact-Based Course Perspective. *Journal of Educational Technology Systems*, 35(5), 281-300.
- Underwood, J. (2006). Digital Technologies and Dishonesty in Examinations and Tests. Report of the Qualifications and Curriculum Authority, United Kingdom. pp. 1-24. Accessed April 4, 2014. http://dera.ioe.ac.uk/6987/1/qca-06_digital-technologies-dishonesty-exams-tests-report.pdf
- Wellman, G. (2005). Comparing Learning Style to Performance in On-Line Teaching: Impact of Proctored v. Un-Proctored Testing. *Journal of Interactive Online Learning*, 4(1), 20-39.
- Williams, R., Lloyd, J., & Simonton, D. (1990). *Sources of OSCE examination information and perceived helpfulness: a study of the grapevine*. Paper presented at the Current Developments in Assessing Clinical Competence, Ottawa.
- Yates, R., & Beaudrie, B. (2009). The impact of online assessment on grades in community college distance education mathematics courses. *The American Journal of Distance Education*, 23(2), 62-70.