AC 2010-2137: OPEN-BOOK VS. CLOSED-BOOK TESTING: AN EXPERIMENTAL COMPARISON

Leticia Anaya, University of North Texas

Leticia Anaya, M.S. is a Lecturer in the Department of Engineering Technology at the University of North Texas College of Engineering. She is currently working in her PhD in Management Science at the University of North Texas. She received her M.S. in Industrial Engineering from Texas A&M University. Her research and teaching interests include Thermal Sciences, Statistics, Quality Assurance, Machine Design, Simulation and Educational Teaching Methods. She has published previously in ASEE Conferences and has developed three laboratory manuals in the following areas: Thermal Sciences, Fluid Mechanics and Mechanics of Materials.

Nicholas Evangelopoulos, University of North Texas

Nicholas Evangelopoulos, PhD is an Associate Professor of Decision Sciences at the University of North Texas and a Fellow of the Texas Center for Digital Knowledge. He received his Ph.D. in Decision Sciences from Washington State University and his M.S. in Computer Science from the University of Kansas. His research interests include Statistics and Text Mining. His publications include articles appearing in MIS Quarterly, Communications in Statistics, ASCE Journal of Hydraulic Engineering, and Computational Statistics & Data Analysis.

Uyi Lawani, University of North Texas

Uyi Lawani, M.S. is a doctoral student in strategy in the Department of Management and also a Fellow of the Robert B. Toulouse School of Graduate Studies at the University of North Texas. While his doctoral minor work was in Economics, he has a B.S. degree in Microbiology and received his MBA in Finance from East Carolina University. His research interests include Organizational governance structures: mergers; acquisitions; and alliances. His solo authored refereed paper has been published in the proceedings of the Decision Science Institute' Department of Management.

Open-Book vs. Closed-Book Testing: an Experimental Comparison

Abstract

This research adds to the ongoing dispute on what is the better method of assessing college students during examinations: open- book or closed- book. The open book assessment method is considered by many to be a realistic method that resembles the actual professional setting of demonstrating acquired knowledge in the field. On the other hand, the closed book assessment method has been used for centuries in traditional institutions as a rigorous method for knowledge assessment. In this research, engineering and business students from a large university in the southwestern United States participated in an experimental comparison designed to determine whether open-book or closed-book is the better approach to access academic knowledge during examinations. The Latin Square experimental design was used to block the variation due to the order in which students received the open-book and closed book treatments, as well as differences in material content tested on the exam. The research study produced mixed results for engineering students that were tested in three different classes: Statics, Mechanics of Materials and Quality Assurance. After adjusting for material content and treatment order differences, in two out of the three classes, the engineering students attained higher scores, a possible indication of achieving a higher learning level, when they were tested in the closed book approach. For the business side, the results indicate that the students attained higher scores, indicating a possible higher learning level, using the open book approach. The implications for this research can be extended to today's online testing and certification environments, which are typically "open-book". The open-book nature of online testing is viewed by some as a necessary evil that poses a validity threat, and by others as a simulation of the professional environment. As a direction for future research, this study could be followed up with future experiments that will attempt to reproduce the results in an online environment.

Introduction

The traditional invigilated closed book approach for testing has been used for generations in various institutions of higher learning. But with the advent of modern technology, the open book format for testing is becoming more common. Controversy exists as to what is the best method of assessing academic learning and performance between these two approaches of testing. Each method has its critics and its supporters. Although the closed book invigilated style is the traditional format that has existed for generations, it is not necessarily problem-free. The main arguments against the closed book format is that this format is irrelevant to real life professional practice, it encourages recall type learning rather than application focused learning, it encourages cheating and it is more costly to administer[1]. Before presenting our research study in detail, these four arguments are first examined here.

The first argument is that the traditional invigilated closed book format is considered to be unrealistic from actual professional practice. In the engineering field, practicing engineers tend to rely on manuals, technical books, Internet and any other extraneous source to be able to solve real life engineering complex problems. Shine and his associates in their article "In Defense of Open-Book Engineering Degree Examinations" defended the open-book engineering testing

format based on the fact that open-book testing resembles most realistic the real life of a working engineer.[1] Although the authors of that article recognized that some engineering schools assess by the open book format and some by the closed book format, the authors advocated the open book format for engineering testing based on four pedagogical arguments: (a) the open book (OB) format is preferable because the environment for open book format is more akin to the real life practicing scenario; (b) the format can be adjusted to incorporate additional content and additional testing time; (c) the format allows for the content can be revised to be more application oriented; and (d) the format is easily adjusted to incorporate computer applications (e.g., software packages) in testing. In the field of business administration, practicing business professionals similarly rely on many external sources to solve business related problems. These sources include Internet research, business consulting, manuals, books and journal research. Thus, the argument that open book exams present a more realistic scenario is also applicable to business students.

The next argument against closed book format is that it encourages recall type learning instead of application focused learning. The main argument against this type of learning is that this type of recall type learning is easily forgotten by the students. Williams and Wong report that invigilated closed book exams "encourage cramming the night before and data dumping on the day with little knowledge retention thereafter[4]." A study by Theophilides and Koutselini focused on evaluating the behavior of the students prior to and during the exams in each of the open book and closed book formats[2]. Specifically, their survey study found that the typical behavior prior to the closed book examinations is that the students tend to (a) memorize knowledge or information, (b) pay more attention to facts, (c) study only the assigned texts, (d) apply surface study, and (e) postpone study and exam preparation until the end of the semester. In contrast, the typical student behavior prior to open book examinations is that the students (a) apply higher order thinking (analyze, synthesize, evaluate), (b) study the course material in depth, (c) read external sources, and (d) interrelate information acquired. The study concluded that during the examination, the open book students tend to work more creatively, make the best use of the material, apply critical thinking, analyze and synthesize, and probe deeper than the closed book students.

But expanding on the argument of knowledge retention, conflicting reports exist to the results that indicate that closed book testing allows for more knowledge retention than open book testing. Moore and Jensen conducted open-book and closed-book research on an introductory biology course that concluded that even though the students that were tested under the open-book format scored slightly higher than their closed-book format counterparts, at the final cumulative exam which was closed-book for all, the formerly open-book format students scored substantially lower than the students who had been tested during the traditional format[6]. This indicated that at the end of the semester, the students tested under the closed book format were able to recall more information than the students that had been tested under the open book format during the semester.

The third argument against closed-book examination is cheating. The presence of opportunity for cheating in closed-book exams examinations is a debatable issue. Shinge et. al. also discussed some extreme measures that some students take to pass closed book invigilated exams. Incidents have been documented where test takers in closed book formats even go to the extent

of hiding material in washrooms which was caused some universities to provide washroom escorts to these candidates[1]. Invigilated closed-book exams require more closely monitoring to ensure that applicants do not bring extra materials or devices. In-class invigilated open-book examinations reduce the tendency for cheating as the student are allowed to bring any material possible to class and they do not have to go to extreme measure to hide any information. But open-book open-web exams present other opportunities for cheating. Without invigilation, it is difficult to assess whether the person who is taking the exam is the actual applicant. Third party cheating is a common issue that affects the integrity of online courses. Williams and Wong, in their study concluded that neither the traditional invigilated closed-book exam nor the open-book open-web (OBOW) exam is foolproof from cheating, but the OBOW exam offers flexibility of completion to the students as it allows students to fit the exam schedule better with work and family arrangements.[4]

The fourth argument against closed-book exams is that invigilated closed book format exams usually require more resources to administer. In large classes, they require more than one invigilator to administer the exam to prevent incidents of cheating. In addition, for certain subjects, additional addendums attached to the exams create bulky exams that are costly to academic institutions. In-class open-book exams require less resources from the institutions as the students are allowed to bring any material possible and the OBOW exam once established, can easily be administered at a very low cost to as many globally located students as possible. These latter exams do not require the presence of an invigilator and thus are less costly to the institution.

Aside from arguments against closed-book examinations, the main arguments against open-book examinations is that the belief that students will tend to score substantially higher than in closed-book formats. The open-book format almost guarantees that the answers will be readily available to the participants. But conflicting results exist in open-book format. Moore and Jensen's, research of open-book format examinations versus closed-book examinations in an introductory biology course concluded that under open-book format, the students scored higher grades in individual exams, attended fewer lectures, and submitted fewer extra credit assignments[6]. But even if the students scored higher in open-book format, the students did not score exceedingly higher. On the first and second open-book exam, the students scored 76% versus 71% and then 75% versus 70%[6]. The drawback of these exams that they were testing basic recall of factual information and synthesis of information on one specific course. The results may be highly course-specific. In another research study by Shinte et al., three different engineering courses were tested at one university and one course was tested at another university. The first course consisted of three different subjects, each assessed by a different instructor. The results for this course is that the closed-book exam format mean was 61% while the open-book format was 53% while the cumulative exam was 62%[1]. The second course consisted of two different subjects, each assessed by a different instructor. The mean for the closed-book format was 54% and for the open-book format was 48% while the cumulative end-of-semester exam yielded 50%[1]. The third course, taught by two different instructors under an open-book format, yielded a mean exam score of 67%[1]. The last course, taught at a different university, yielded a mean of 50% for open-book and a mean of 46% for closed-book[1]. In short, the notion that open-book exams automatically yield exceedingly high scores was diffused by this research. The drawback to this study is all the exams were created and graded by different

instructors. No consistency of grading and subject existed to make accurate comparison between open and closed-book formats.

Research Problem Background

Academic testing in the engineering fields and business fields differ from each other. In the engineering field, the exams tend to be highly mathematical and analytical and rely on problem solving approaches that tend to be step by step procedures to arrive at an answer. Because of the length of these procedures, it is very easy for the students to make mathematical errors and arrive at a wrong answer. Moreover, some of the mathematical formulas needed for these exams can be highly complex. Because of this, the possibility always exists that engineering students may fail an exam because of their inability to recall mathematical formulas instead of their inability to apply engineering concepts to solve problems. The issue that is often presented in engineering examinations is: What is the exam actually testing? Mathematical principles or engineering principles? Fortunately, most engineering instructors recognize this issue and tend to add formula sheets (in addition to already attached charts, plots) to the exams, a practice that increases the administration costs for these exams. In the business field, recall-type content can be tested in multiple choice, fill-in-the-blanks exams and essay-type questions. These exams tend to be less bulky than engineering exams as no additional attachments are needed. In both fields, unstructured and semi-structured case testing can also be formulated. In the engineering field, a case-structure exam can be created to design an item under specific constraints, while in the business field a case-structure can be created to determine the best strategy in pursuing profit increase for a corporation.

Research Statement

The aim of this research is to study the difference in performance between open-book and closed-book testing in the business field and in the engineering field. These two environments were chosen for comparison as these two environments are generally considered to be complete opposites in the academic field. The representative academic environment for the business environment was a freshman-level business course entitled Free Enterprise System in a Global Environment and the representative academic environment for the engineering field were three engineering classes: Quality Assurance, Statics, and Mechanics of Materials. The business class was taught by one instructor using a conceptual approach for testing that included multiple choice and fill-in-the-blank. In the engineering classes, also taught by one instructor, the analytical approach for testing was used, which includes students having to solve mathematical problems. The hypothesis that was investigated was that there was no significant difference in academic performance between the open-book and the closed-book approaches of testing students under either academic environment. What is unique about this research is that an experimental Latin Square design was used to conduct the research which allowed for analysis of the data to include controlling for group order and content. The data obtained was analyzed in two forms: with and without controlling for these two variables. The results presented here include the results under both scenarios.

Exam Administration Procedure

The engineering classes that were part of this research are: Quality assurance (senior level course), Statics (sophomore level course), and Mechanics of materials (sophomore level). The exams were highly mathematical problem solving exams and they were administered to each respective class. During testing, two equivalent exams were created for each class. Each exam was divided into two parts, I and II creating a total of four exam segments that were printed as four differently colored exams. The questions in both parts in each exam were very similar (e.g., if in one exam a part I question was to determine the torque given angular velocity and radius, in the next exam, the related part I question was to determine the angular velocity given a torque and a radius.) In either case, the students used the same approach and formula to solve the problems. The students were randomly assigned seats, and half of each exam was administered under the closed-book approach and the other half under the open-book approach using the following Latin square design format.

Latin Square Design			
		Part 1	Part 2
Group 1 Students	Group 1	Open	Closed
Group 2 Students	Group 2	Closed	Open

Table 1. Latin Square Design Format

After the exam was taken, students were given a survey to determine key demographic variables about the students. In the first two exams, the survey was not immediately administered but was administered the next class meeting. In the last class, the survey was attached to the exam.

The business course that participated in this study was an introductory business course entitled "Free Enterprise Systems in a Global Environment." The business students were similarly divided in two groups, with each group receiving part I and part 2 segments of two separate exams. The administration of the exams was also conducted using the Latin Square format outlined in Table 1. A survey was administered immediately following the examination.

Survey

A survey was administered to all the classes asked the student's academic status, their major course of study, gender, the study time allocated to the exam, expected grade for this exam, their current GPA, their expected GPA after completion of course, and the number of business (engineering) courses taken prior to the class. The idea behind this survey was to determine whether any of these variables impacted academic performance for both closed-book and open-book formats. The determination of this was achieved through the performance of a linear regression analysis on all the closed-book scores and open-book scores for these classes.

Demographics

The survey indicated demographics for the classes tested in this experiment that are summarized in Table 2.

	Engineering Course: Quality Assurance	Engineering Course: Mechanics of Materials	Engineering Course: Statics	Business Course: Free Enterprise Systems
Sample Size	51	60	19	34
Majors	67% Mechanical Engineering Technology, 33% Electrical Engineering Technology	68% Mechanical Engineering Oriented, 22% Construction Engineering Oriented	37% Mechanical Engineering Oriented, 47% Construction Engineering Oriented	44% Business Majors, 41% Construction Engineering Majors
Gender	94% Male	95% Male	100% Male	65% were females
Academic Status	70% Seniors, 26% Juniors	44% Juniors, 44% Seniors	47% Juniors, 42% Seniors	34% Sophomores, 31% Juniors, 21% Seniors
Study Habits	32% allocated 2-3:59 hours, 30% allocated 6-7:59 hours	45% allocated 2-3:59 hours, 29% allocated 4-5:59 hours	28% allocated 0-1:59 hrs, 28% allocated 2-3:59 hours, 39% allocated 4-5:59 hours	34% allocated 0-1:59 hrs, 40% allocated 2-3:59 hours, 15% allocated 4-5:59 hours
Expected Grade on Exam	52% expected "B", 19% expected "A", 19% expected "C"	61% expected "B", 18% expected "A", 21% expected "C"	37% expected "B", 37% expected "A", 21% expected "C"	60% expected "B", 38% expected "A"
CURRENT GPA	51% -- 3:00-3:49 GPA, 38% -- 2:50-2.99 GPA	40% -- 3:00-3:49 GPA, 29% -- 2:50-2.99 GPA, 21% -- above 3.5 GPA.	32% -- 3:00-3:49 GPA, 26% -- 2:50-2.99 GPA, 21% -- above 3.5 GPA.	38% -- 3:00-3:49 GPA, 34% -- 2:50-2.99 GPA, 12% -- above 3.5 GPA.
Expected GPA at end of Course	59% --- 3:00-3:49 GPA, 24% --- 2:50-2.99 GPA	40% -- 3:00-3:49 GPA, 35% -- 2:50-2.99 GPA, 24% -- above 3.5 GPA.	42% - 3:00-3:49 GPA, 26% - 2:50-2.99 GPA, 21% - above 3.5 GPA.	41% - 3:00-3:49 GPA, 34% - 2:50-2.99 GPA, 15% - above 3.5 GPA.
Engineering (Business) Courses Taken Prior Class	62% taken more than 10 engineering courses, 28% had taken between 7-9 classes.	28% taken more than 10 engineering courses, 30% had taken between 7-9 classes, 30% had taken between 4-6 classes.	79% had taken between 4-6 engineering classes, 32% had taken between 1-3 classes.	34% had never taken a business course, 50% had taken between 1-3 classes.

Table 2. Survey Participant Demographics

Results for the Basic Open-Book/Closed-Book Comparison.

The results of this study have been divided into two parts. The first part consisted on determining whether there was any significant difference between open-book and closed-book exam scores without controlling for any effects of group order and content. This was assessed by performing a basic t-test comparison of the means test. This test essentially compares the mean of one group with the mean of another group. The t-test comparison test assumptions are that the populations being compared are normally distributed, have equal variances and are independent from each other. The second part consisted of determining whether there was any significant difference between open and closed-book exam scores by controlling the effects of group order and content. This was assessed by performing a Latin Square design. The Latin Square design is an experimental design where the number of rows (representing one blocked factor), the number of columns (representing another blocked factor) and the number of treatments are equal. Restrictions on randomization of treatment exist as each treatment is present in exactly one row and one column of the design. In our particular design, the rows represented group order, the columns represented content, and the treatments were closed-book and open-book testing. The analysis of the results is done by statistically accounting for ("blocking") the effects of group order and content order. Table 3 below illustrates the results that indicate that there is no significant difference between open-book and closed-book scores under these conditions.

Paired Samples Test Results for the Engineering Data									
Subject	Pair T test Comparison	Paired Differences							
	Variables	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		T	Df	Sig. (2-tailed)
					Lower	Upper			
Quality Assurance	OverallOpen – OverallClosed	-0.56	9.199	1.301	-3.174	2.054	-0.43	49	0.669
Mechanics of Materials	OverallOpen – OverallClosed	-0.5	4.926	0.641	-1.784	0.784	-0.78	58	0.439
Statics	Open Book – Closed Book	1.542	5.055	1.16	-0.894	3.979	1.33	18	0.2

Table 3. Results of Engineering Open Book Closed Book Data Under No Blocking Effects

When analyzing the business data through a simple paired t test, the results indicate that there is a significant difference between average closed book scores and average open book scores for the business students. Table 4 illustrates the results from analyzing the business data under conditions of not controlling for the effects of group order and content. The linear regression results indicate that only the statics open book data was the closest that could fit a linear

regression equation with an R^2 of .689. However, even under these conditions, none of the variables surveyed were deemed significant at the 0.05 significance level. A linear regression model was not the appropriate model for any other open book or closed book scores for either the business class or the engineering classes.

t-Test: Paired Two Sample for Means		
	Average Closed	Average Open
Mean	16.63235294	15.23529412
Variance	1.262254902	3.200534759
Observations	34	34
Pearson Correlation	0.153647527	
Hypothesized Mean Difference	0	
Df	33	
t Stat	4.15430067	
P(T<=t) one-tail	0.000108339	
t Critical one-tail	1.692360258	
P(T<=t) two-tail	0.000216678	
t Critical two-tail	2.034515287	

Table 4. Business Course Analysis Assuming No Blocking Effects.

Part II of the analysis involves the analysis of the Latin Square design illustrated by Table 1. Under the Latin Square format, all groups were exposed to both open and closed book conditions of testing. The Latin Square analysis indicates that when controlling for variation due to group order and content, mixed results were obtained for the engineering classes. The results are summarized in Table 5. Under these conditions, the Quality Assurance class and Mechanics of Materials scored highest under closed book conditions and the Statics course scored higher under open book conditions. (One of the possible reasons that the Statics course performed differently from the other engineering courses is that the Statics course is one of the introductory engineering courses that engineering students take. Perhaps the fact that these students are inexperienced in the academic rigors of the engineering field may have caused them to score differently from the other two courses.) The business course scored higher under the open book conditions when controlling for content and group order. Although it is difficult to quantify and assess higher level of learning and an ongoing dispute exists as to what is considered an effective measurement to assess student learning, higher examination scores are usually positively correlated with higher levels of learning. Table 5 illustrates that engineering students scored higher in closed book tests in two out of three cases and thus, possibly attained a higher level of learning under closed book

conditions. The business class scored higher under open book conditions and thus possibly attained higher level of learning under the open book conditions.

LATIN SQUARE RESULTS			Significant Difference	Mean Open	Men Closed	Higher Scores
Quality Assurance	Part 1	Part 2	Yes	17.346	18.525	Closed Book
Group 1,25 Students	15.08	22.32				
Group 2,26 Students	14.73	19.612				
Mechanics of Materials	Part 1	Part 2	Yes	14.030	14.573	Closed Book
Group 1, 26 Students	14.288	15.192				
Group 2,33 Students	13.954	13.772				
Statics	Part 1	Part 2	Yes	25.574	24.147	Open Book
Group 1, 10 Students	27.48	23.85				
Group 2, 9 Students	24.444	23.667				
Business Course	Part 1	Part 2	Yes	15.584	15.532	Open Book
Group 1, 18 Students	15.167	16				
Group 2,16 Students	15.063	16				

Table 5: Results of Latin Square Analysis.

Discussion and Conclusion

The results indicate that content and group order can make a significance difference in the results of the open-book and closed-book scores. After controlling for effects of content and group order, significant differences in scores were detected in the open-book and closed-book data for the engineering classes. Without controlling for these effects, there was no significance difference between the scores. This finding underscores the importance of a well-designed study in obtaining correct results. Similarly for the business data, the preference for open-book and closed-book shifted toward an open-book format after accounting for the effects of content and group order. These findings imply that a researcher should ensure that all important effects have been accounted for in a similar study before a valid comparison of an open-book and closed-book study can be made. Content and group order are a couple of the effects that were tested in this research. But other effects that can affect the exam scoring of open-book and closed-book exams may exist. Future research should extend to determine what other factors could affect these scores. Theophilides and Koutselini's past research has identified two problems for open-book examinations: 1) students wasting time looking for information and having less time to formulate responses properly and 2) the false sense of confidence created by open-book exams that cause them to be less prepared for the exams[2]. Based on these observations, one possible explanation for closed-book exams yielding higher scores is that the students may be forced to study more before the exam, and focus more during the exam. Since in our study the students

were asked to take both open-book and closed-book exams and did not specifically anticipate one or the other type, the latter (i.e., focusing more during the exam) is more plausible than the former (studying more before the exam.) While at present this is only speculation, one direction for future research would be to determine how these two factors can be incorporated into a similar future study. For example, the student survey, on top of demographics, could incorporate questions related to the student's preparation effort, as well as questions related to test-taking anxiety.

Lastly, the future of the traditional invigilated closed book is subject to speculation. With the ongoing trends in education moving more and more toward an online oriented education, the future of this type of exam is questionable. Strictly online universities are moving toward the open-book open-web based (OBOW) exams because of the logistics associated with online exam participation. Williams and Wong have basically demonstrated that with higher education evolving in terms of technological changes that include uses of multimedia, online pedagogies, and flexible delivery of courses, the traditional final examination has become a somewhat inflexible proposition for such emerging online global universities[4]. Shine et al. also mentioned that closed-book examinations are not highly adaptable for online learning as they present an inconvenience for online participants.[1]

Bibliography

1. Shine, S., Kiravu, C., and Astley, J., "In Defense of Open-Book Engineering Degree Examinations." International Journal of Mechanical Engineering Education, Vol. 32, No. 3, July 2005, pp. 197-211.
2. Theophilides, C. and Koutselini, M., "Study Behavior in the Closed Book and the Open Book Examination: A Comparative Analysis," Educational Research and Evaluation, Vol. 6, No. 4, December 2000, pp. 379-393.
3. Theophilides, C., and Dionysiou, O., The major functions of the open-book examination at the university level; A factor analytic study, "Studies in Educational Evaluation, Vol. 22, No. 2, 1996, pp. 157-170.
4. Williams, J.R., and Wong, A. (2009), "The Efficacy of Final Examinations: A comparative study of closed book, invigilated exams and open-book, open-web exams." British Journal of Educational Technology, Vol. 14, (2), 2009, pp. 227-236.
5. Jennings, M, Fox, J., and Graves, B., (1999), "The test-takers' choice: an investigation on the effect of topic on language-test performance" Language Testing, Vol 16(4), pp 426-456
6. Moore, R., and Jensen P., "Do Open-Book Exams Impede Long-Term Learning in Introductory Biology Courses," Journal of College Science Teaching, July/August 2007, pp. 46-49.
7. www.ndsu.nodak.edu/ndsu/horsley/LS.pdf Retrieved December 1, 2009.