A Comparison of Learning Outcomes by 'In-Course' Evaluation Techniques for an On-Line Course in a Controlled Environment

Otis L. Stanley, *University of Southern Indiana*

Abstract

The purpose of this study is to determine if the type of weekly evaluation method used in an on-line course contributes to a difference in the learning outcomes for students. Two methods for the ongoing evaluation of student learning were analyzed for differences in learning outcomes as demonstrated by mid-term and final exam test scores. Using an experimental design, students that enrolled in either a Disease Control course or an Epidemiology course were randomized into one of two sections in each course. Holding the course parameters the same except for the weekly evaluation type (homework or quiz), bivariate analysis using independent t tests supported that sections were similar in both courses with respect to test scores. A statistically significant difference did occur between final exam scores in the Disease Control course with the higher scores occurring in the quiz section. End of course student satisfaction surveys were similar for both types of evaluation methods and for both courses. Of the students that responded to the surveys, the majority felt their overall learning experience was either good or very good, regardless of whether they completed weekly interactive homework assignments or automated quizzes. This study supports the idea that learning outcomes and student satisfaction scores with a weekly automated quiz are equivalent or improved over the more interactive weekly homework assignments in select courses.

Introduction

There is strong evidence that properly structured online education is as effective, if not more effective, than traditional face to face instruction given equal initial student maturity and motivation (Johnson, et al., 2000). Online instruction is well suited for what has been called "learner centered" education if the learner is provided with appropriate venues of information sources and discipline appropriate tools (i.e. methods, process guidance) (Brown, 2005). In the online environment where the teacher is physically absent from view, the student is challenged to be self-directed in exploring the course materials and being an active participant in the learning experience. Rather than being a passive recipient of information, the student becomes a partner with the course faculty in the pursuit of knowledge (Weimer, 2002). Extending the learner centered model, the online environment is also an appropriate, if not optimal setting for learning facilitation "where the instructor assists students to access information, to synthesize and interpret it and to place it in a context- in short to transform information into knowledge" (Kershaw & Safford, 1998, p287).

But meaningful, appropriate, and quality online instruction faces many challenges. Choosing the best online techniques to use for various course types and students, such as video capture, asynchronous discussion, virtual classroom, interactive assignments, and automated quizzes, is an example of one such challenge. Draves (2000) and Collison et.al (2000) suggest that various techniques work better (or are preferred) for different types of courses or teaching styles. But an equally important consideration is that some assessment methods are more time intensive per student. The amount of time spent weekly per student can be considered as a measure of course interactivity and this is generally thought to improve learning outcomes. But as class size increases in a course, compromises are often required to balance available faculty time with the perceived optimal instructional or assessment method for the subject matter at hand.

Interactive homework assignments, online discussions, and student papers and projects, are all excellent in-course evaluation (and instructional) strategies when properly applied and when the technique matches the course objectives. In addition to capturing a grade, these strategies also serve to re-enforce the importance of key terms and concepts, focus the student's learning activities, provide a medium for reflection, allow the application and synthesis of course

materials, and lastly, serve effectively as a means to establish whether a student is still in the course (i.e. attendance).

Each of these methods may involve considerable faculty time per student, especially when they are used frequently through out a course (for example when used weekly). Replacing such faculty-student intensive interactions with quizzes (or other automated assessment methods) can reduce the amount of time per student required by the faculty, but the concern is that the quality of learning outcomes may suffer. It must be noted that the amount of faculty time required to develop valid and reliable quiz and exam questions is quite extensive. However, once developed, a good test bank of questions can be used over a period of time if properly secured and updated regularly.

Methodology

Purpose

The purpose of this study is to determine if the type of weekly evaluation method used in an on-line course contributes to a difference in the learning outcomes for students. This study examines this issue in the controlled environment of standard courses where only the type of in-course evaluation technique varied. The two questions the findings will answer are;

1. Is there a difference on mid-term and final exam scores between those students who completed weekly interactive homework assignments and those who completed automated weekly quizzes?
2. Is there a difference on end of course satisfaction scores between those students who completed weekly interactive homework assignments and those who completed automated weekly quizzes?

Study Design

An experimental design was used to explore the dependence of scores on mid-term and final exams with the type of method used for weekly in-course evaluation. Simple randomization was used to assign each student to either a “homework” group or a “quiz” group. In addition, a cross-sectional survey was given at the end of the course to evaluate the satisfaction of the student with

the method used and to obtain information on other variables that might be pertinent to the study outcomes.

Study Population

The available population for the study were college students at a small mid-western university who enrolled for one of the courses to be used in the study. The two courses were Disease Control, and Epidemiology. Although the students who chose to enroll in one of the courses may not represent the "average" student, the randomized allocation process of students into study groups did allow inferences to be made about the major study variable (in-course evaluation type) with respect to outcome measures (major exam scores and end of course satisfaction).

Sample Selection and Size

Students were randomly placed into one of two in-course evaluation groups in each of the two courses. Both evaluation techniques had been previously used in similar courses, so no student was exposed to a novel technique. After the completion of the course, students were asked to fill out an end of course questionnaire that addressed perceived satisfaction and information on covariates that might have correlated with scores on exams.

The expected enrollment was approximately 50-60 students for the Disease Control class and 30-40 for the Epidemiology Course. These numbers would only be expected to find a difference of 30% and 40% respectively at commonly used levels of significance and power (.05 and .80 respectively, two-tailed). Because of the sample size limitations per course, it would have been necessary to run 2-3 iterations before sufficient statistical power would be achieved to find a 10% -20% differences in test scores at the standard levels of significance and power. The limitation would remain that even a 10% difference (usually more than the difference between an 'A' and a 'B') is a large enough difference to detect.

Procedure

The same instructor had taught the courses used in this study as well as online instruction for at least 3 years. The courses are standardized by format and include identical instructional techniques (assigned book readings, modular online notes on the subject matter, video capture

supplemental materials, and internet links to related materials or study aides). An asynchronous discussion area was available for student comments and discussion in each course, but discussion was not required as part of the course grade.

Each study group had identical learning inputs and only differed by the use of graded homework assignments or quizzes as the method of in-course evaluation. Mid-term and final exams, which consisted of objective questions in each course, served as the outcome measure to test the hypothesis that in-course evaluation methods are associated with test score results. The two courses, Disease Control (*Course A*) and Epidemiology (*Course B*), differed in that Disease Control is non-analytic and non-quantitative in nature, while Epidemiology has analytic and quantitative content. Because of this difference, each course was analyzed separately.

At the onset of the course, students were randomized into either a "homework" section or a "quiz" section. Equal numbers were used in each. All other aspects of the course were identical (learning materials available to the student, an asynchronous discussion area for student questions, additional learning materials, and the major exams the student would take).

Homework assignments were given weekly and the questions were related to the key terms and concepts for that week's material. Homework assignments were instructor graded and returned to the student with comments about the quality of the work. The weekly homework assignments were worth 20 points each (200 total points for all homework). Quizzes were also completed weekly. The questions were also related to the key terms and concepts of the week's material. The quiz questions were all objective in nature with automated grading (20 points per quiz, 200 points total). No instructor comments about the items missed were provided. Students did have access to correct answers after the quiz.

Students understood that their course grade depended on homework (200 points), a midterm exam (100 points) and a final exam (100 points) if they were in the 'homework' section. If they were in the 'quiz' section, their grade depended on quiz scores (200 points), a midterm exam (100 points) and a final exam (100 points). Both groups had equal access to the instructor via the discussion board, email, or phone.

At the end of the course, students were asked to complete a questionnaire that addressed perceived satisfaction with the course and questions about covariates that might have been strongly correlated with their exam scores. Although student names and identifiers (SSN) were not used in the analysis, individual identity was retained to place test scores with covariates and satisfaction. Student names and identifiers were not to be used outside the content of the courses themselves and the analysis only contained a line number to relate all variables to a single student. In effect, the subjective evaluation of each student was blinded to the instructor.

Findings

The analysis proceeded in a traditional way. Univariate description was completed for each variable. Any unusual values were considered individually. Bivariate analysis was done on the primary outcome measure (exam scores) by group assignment ('homework' or 'quiz'). Perceived satisfaction was also analyzed bivariately by group membership, but low response rate numbers and the potential for differential inputs between those that did respond and those that did not made such analysis not as meaningful. Descriptive summaries of the subjective responses were considered.

Cofactor information was available and was considered if it was thought to have an influence on the outcome measures (i.e. academic major, junior vs senior class, gender). Random assignment of students into groups should have also randomized covariates, thus primary bivariate analysis was expected to be sufficient to obtain an unbiased estimate of group differences. In the case that a covariate was strongly correlated to the outcome measure, adjustment on the covariate could have been done to improve precision. However, no major variation was observed with respect to student type (major, class, gender) so no adjustments were done in the analysis.

Results

Course A: Disease Control

In the Disease Control course, 51 subjects remained at midterm with 26 in the Quiz section and 25 in the Homework section. Attrition was minimal and comparable in both sections from the

beginning of the course. The first point of analysis was after the midterm exam. The average midterm score was 73.69 for the Quiz section with a standard deviation of 11.76. The average midterm score was 70.68 for the Homework section with a standard deviation of 9.88 (See table 1). Total points up to that point were 152.46 for the Quiz section with a standard deviation of 36.97. Total points at that time for the Homework section was 145.48 with a standard deviation of 41.32. No extreme values (greater than 2 SD's from the mean) occurred in either group and the assumption for equal variance was not violated, so an independent t test analysis was done on the midterm scores. The calculated t statistic of .99 (p=.328, two tailed) was non-significant, indicating that the groups did not significantly differ from each other for midterm test results. Total points at midterm were also compared with a t statistic = .63 (p=.53, two tailed) which was also not statistically significant (See table 2).

Table 1: Disease Control Midterm Scores

	Homework Group	**Quiz Group**
Mean	70.68	73.69
Standard Deviation	9.88	11.76
Standard Error	1.97	2.3
N	25	26
t statistic	0.98	
p value	.328	

Table 2: Disease Control Total Points at Midterm

	Homework Group	**Quiz Group**
Mean	145.48	152.46
Standard Deviation	41.32	36.97
Standard Error	8.26	7.25
N	25	26
t statistic	0.63	
p value	.052	

At the end of course, the average final exam score for the Quiz section was 86.83, with a standard deviation of 7.91 (n=24). The average final exam score for the Homework section was 79.88 with a standard deviation of 9.4 (n=24). The calculated t statistic was 2.77 (p= .008). This suggests a significant difference between final exam scores (See Table 3). The total course point average for the Quiz group was 324.77 with a standard deviation of 78.27. While the total average for the Homework group was 333.24 with a standard deviation of 70.80. The calculated t statistics was .39 (p=.69, two tailed) which was non-significant overall between the two groups (See Table 4).

Table 3: Disease Control Final Scores

	Homework Group	Quiz Group
Mean	79.88	86.83
Standard Deviation	9.4	7.91
Standard Error	1.91	1.61
N	24	24
t statistic	2.77	
p value	.008*	

Table 4: Disease Control Total Points

	Homework Group	Quiz Group
Mean	333.24	324.77
Standard Deviation	70.80	78.27
Standard Error	14.45	15.97
N	24	24
t statistic	0.39	
p value	.69	

Course B: Epidemiology Course

Due to the quantitative nature of the epidemiology course the enrollment numbers were lower. At midterm, 12 homework section students had an average midterm score of 65.78 with a standard deviation of 17.8, while 12 quiz section students had an average midterm exam score of

63.67 with a standard deviation of 10.38. When the groups were compared by the t-statistic, t= .35 (p=.73) results were statistically non-significant (See Table 5). At end of course, average final exam scores for the homework group was 72.1 with a standard deviation of 11.9 (n=10). While the quiz group had an average score of 78.22 with a standard deviation of 9.2 (n=10). When compared, the t-statistics = 1.2 (p=.21, two tailed) was non-significant. Similar statistics are seen for total course points (See Tables 6 and 7). Attritions rates were similar for both groups.

Table 5: Epidemiology Midterm Scores

	Homework Group	**Quiz Group**
Mean	65.78	63.67
Standard Deviation	17.8	10.38
Standard Error	5.14	2.99
N	12	12
t-statistic	0.35	
p value	.72	

Table 6: Epidemiology Final Scores

	Homework Group	**Quiz Group**
Mean	72.1	78.22
Standard Deviation	11.92	9.19
Standard Error	3.76	2.90
N	10	10
t-statistic	1.28	
p value	.21	

Table 7: Epidemiology Total Points

	Homework Group	Quiz Group
Mean	289.5	267.99
Standard Deviation	98.0	90.98
Standard Error	30.99	28.77
N	10	10
t-statistic	0.50	
p value	.61	

Student Satisfaction Surveys

A short survey was sent to all student participants in the study at the end of the semester. They were asked to rate their overall learning experience in the course and to rate their weekly assessments (homework assignments or quizzes) as tools for preparing for the major exams as 'Very Good', 'Good', 'Average', 'Poor', or 'Very Poor'. They were also asked to answer 'Yes' or 'No' as to whether the assessments focused their learning activities and emphasized core concepts. Another Y/N question was asked about their use of the discussion feature (i.e. asynchronous classroom) of Blackboard. Tables 8 and 9 summarize the survey results for each course.

The Homework (Assignment) and Quiz sections respondents in the Disease Control course were generally positive in their appraisal of both evaluation tools with quiz section responses being slightly more positive (see Table 8). Only in question 4 were there significant differences in the responses. This indicates that the Quiz section relied more heavily on the asynchronous classroom (i.e. discussion board) to have their questions answered than did the Homework section.

A question that must be asked whenever voluntary surveys are performed is, "Do the opinions presented represent the overall population from which they are drawn?" This question is partially addressed by the response rate. In Table 8, 19 of 26 students responded to the survey in the homework section (73% response rate), and 13 of 25 students responded in the quiz section

(52% response rate). A better indicator is to look at mean scores across the sections of the respondents and determine if the mean respondent scores were similar to the overall mean scores of all students. From the Table 8, it can be seen that respondents had similar average scores across sections, but that respondents in each section score higher (on average) than did the section students overall. These results give us both some confidence in the survey results, but also a cause for pause since the better students were more likely to fill out the survey. The results in the Epidemiology course are very similar to the Disease Control course, although overall numbers are smaller. See Table 9.

Table 8: Disease Control Course End of Course Student Satisfaction Survey

Question 1 and 2 answers are given as the proportion of students in that section that answered 'Very Good' or 'Good' to the question.			
	Assign Group	Quiz Group	p=*
Q1: How would you rate your overall learning experience in this course?	16/19 (.84)	12/13 (.92)	.63
Q2: How would you rate the weekly assessments (assignments or quizzes as tools for preparing you for the major exam?	16/19 (.84)	11/13 (.84)	1.0
Questions 3 and 4 answers are given as the proportion of students in the section that answered 'Yes' to the question.			
Q3: Did the weekly assessments enhance your learning experience (i.e. focus your learning, emphasize core concepts)?	18/19 (.84)	12/13 (.92)	1.0
Q4: In addition to the weekly assessments, did the discussion in the Blackboard discussion area enhance your learning experience?	8/19 (.42)	12/13 (.92)	.008**
Comparison of mean total points			
Mean total points of respondents	324 (n=19)	343 (n=13)	.91
Mean total points of entire class section	289 (n=24)	323 (n=24)	.69
Comparing average respondent total points and entire section average.	p=.54	p=.43	

* Probability as calculated by Fischer Exact Test for nominal data and t test for interval level (mean) data.
**Statistically significant results.

Table 9: Epidemiology Course End of Course Student Satisfaction Survey

Question 1 and 2 answers are given as the proportion of students in that section that answered 'Very Good' or 'Good' to the question.			
	Assign Group	Quiz Group	p=*
Q1: How would you rate your overall learning experience in this course?	5/7 (.71)	5/6 (.83)	1.0
Q2: How would you rate the weekly assessments (assignments or quizzes as tools for preparing you for the major exam?	5/7 (.71)	4/6 (.66)	1.0
Questions 3 and 4 answers are given as the proportion of students in the section that answered 'Yes' to the question.			
Q3: Did the weekly assessments enhance your learning experience (i.e. focus your learning, emphasize core concepts)?	6/7 (.71)	5/6 (.83)	1.0
Q4: In addition to the weekly assessments, did the discussion in the Blackboard discussion area enhance your learning experience?	1/7 (.14)	4/6 (.66)	.10
Comparison of mean total points			
Comparing mean total points of respondents	324 (n=7)	306 (n=6)	.65
Comparing mean total points of entire class section	289 (n=10)	267 (n=10)	.65
Comparison between mean respondent total points and entire section mean.	p= .3	p= .29	

* Probability as calculated by Fischer Exact Test for nominal data and t test for interval level (mean) data.

Study Validity

There is no attempt to assume external validity in this study. Specific student types will enroll in these courses. There is also little reason to suspect that differences for in-course evaluation methods would differ by major or institution. Additional variables were available from the class rosters addressing the overall representativeness of the subjects that could have been adjusted as covariates to improve precision. These factors included:

1. Gender
2. Class (sophomore, junior, senior)
3. Academic Major

These variables were distributed fairly proportionally between the groups (as part of the random assignment process), but females predominated in both courses and most students were health profession majors. Almost all students were upper classmen. Other variables considered which were not captured that might have been related to the outcome measures of interest included:

1. College GPA at beginning of course
2. Fulltime student during course
3. Age
4. Prior Public Health course
5. Prior Statistics course
6. Prior Microbiology course
7. Prior clinical course covering disease etiology

Within the courses proper, after the random allocation into groups, no blinding was done since the group membership of the student was obvious to the instructor. The potential bias introduced by differential treatment of students by group was considered However, it is noted that the 'quiz' group was not generally available to the instructor (except in the discussion area), and the 'homework' group (considered as the treatment variable) had the instructor's comments as an integral part of the treatment variable.

Discussion

The courses used in this study are well suited to the learner centered model. These courses are heavily content based and allow students to explore well developed online resources, such as those from the Centers for Disease Control. Through the use of links within the course modules and weekly exercises involving those internet sites, all students interact with a wide variety of learning resources. As a result of this exploration, students are guided through the process of taking disparate bits of information and forming a knowledge base for the discipline.

Much of the literature regarding on-line learning strategies advocate various tools based on anecdotal evidence of the author or as the result of surveys done on groups of students or

instructors (Phillips, 2005; Suen, 2005). Such evaluations may be subject to personal preferences, or worse, the tools are compared when the learning inputs vary from course to course. By using an experimental design, and by using a standard learning environment to maximize study precision, the risk of results being biased by differential student, instructor, or learning environment characteristics were minimized in this study. These controls lend strength to the confidence one may place in the findings.

This study addressed an important concern in on-line learning;

> Do the online methods used to evaluate ongoing student learning during a course significantly affect the learning outcome?

This study compared instructor graded homework assignments (a labor intensive interactive method) to automated quizzes (a labor reduced technique) as they relate to the outcome variable of exam scores (a surrogate measure for student learning). From the data it can be seen, that when the activities used to engage the learner in the learning environment are sufficiently rich, student learning is not decremented by the use of an objective format. In addition, student acceptance of this format is equally positive to the competitive interactive format.

Overall, it is of interest to instructors and administrators alike to know which of the available online learning methods (Discussion Boards, Assignments, Quizzes, Virtual Classrooms, etc.) provide the best student outcomes and the most efficient use of faculty time.

An interesting observation beyond the main study results is that the less interactive "Quiz" group did engage in interaction with the instructor by being more interactive in the asynchronous classroom (i.e. discussion area). This interaction was not student to student, but was directed toward the instructor and was usually related to a quiz question. By being student generated, these interactions became highly targeted and relevant to the learner needs representing a 'teachable moment'. The "Homework" group, it is speculated, did not have the need for explanations about lost points since the rationale for their errors was provided with the instructor comments.

Finally, if the quiz sections did do better overall as the data suggests, why did this happen? Although not testable in the current context of this paper, it is presumed that the missed points on a quiz might have provided stronger and more focused 'clues to action' to "Quiz" students that did poorly on a quiz, than the more general instructor comments provided to the "Homework" students. In other words, missed quiz questions were more likely to effect a positive behavioral response (look up the answer, ask a question in the discussion area) than a homework comment would. This would explain not only the better outcome scores, but also the increased traffic in the asynchronous classroom.

A final disclaimer is the potential limitation in this study as it applies to the external validity of the findings. Both of the courses used here emphasize content that is traditionally presented as factual in nature in disciplines that have established orthodoxy. Hence, these courses did not emphasize value laden material that is more robustly considered in group or community settings where interactivity and group culture are important parts of the learning experience (Our students, AACU, 2004).

References

Brown, T. H. (2005). Beyond constructivism: exploring future learning paradigms. *Education Today*, 2, 1-11.

Collison, G., Elbaum, B., Haavind, S., and Tinker, R. (2000). *Facilitating Online Leaning: Effective Strategies for Moderators.* Atwood Publishers: Madison, Wi.

Draves, W.A. (2000). *Teaching Online.* Learning Resource Network(LERN): River Falls, Wi

Johnson, S., Argon, S., Shaik, N., Palma-rivas, N. (2000). Comparative analysis of learner satisfaction and learning outcomes in online and face-to-face learning environments. *Journal of Interactive Learning Research,* 11, 29-49.

Kershaw, A. and Safford, S. (1998). From order to chaos: the impact of educational telecommunications on post-secondary education. *Higher Education*. 35, 285-298.

Our Student's Best Work: A Framework for Accountability Worthy of Our Mission. (2004). *Association of American Colleges and Universities*. Washington, DC.

Phillips, J.M. (2005). Strategies for active learning on online continuing education. *The Journal of Continuing Education in Nursing*, 36(2), 77-83.

Suen, L. (2005). Teaching epidemiology using WebCT: Application of the seven principles of good practice. *Journal of Nursing Education.* 44(3), 143-146.

Weimer, M. (2002). *Learner-Centered Teaching: Five Key Changes to Practice.* Jossey-Bass: A Wiley Co: San Francisco, Ca.