A Better Way To Teach Principles Of Microeconomics? Evidence From The Classroom

Carol Diane Tallarico, Carroll University
Mary Ann Wisniewski, Carroll University

ABSTRACT

This paper compares the success rates of students in two different sections of Principles of Microeconomics taught in 2013-14. In one, the course was taught using weekly quizzes which culminated in a comprehensive final exam. In the other, two large exams replaced the weekly quizzes. The method of assessment had a clear impact on the DWF rate, increasing from 10% when using the quiz method to 56.67% when using the exam method. The method of assessment showed no impact on other measures of student success such as performance on the comprehensive final exam or improvement from pre-test to post-test.

INTRODUCTION

As instructors of Principles of Microeconomics, or other business courses, we are often looking for ways to increase the success of students taking our courses. In particular, Principles of Microeconomics has long been an area of poor performance [Berrett, 2012]. At many institutions, this course is primarily taken by students because it is required for their various majors rather than due to a particular interest in economics. This potentially makes the teaching of this course even more difficult. How then as an instructor is it possible to increase the effectiveness of learning and the success of students in this course? Is there indeed a better way to teach Principles of Microeconomics?

First we must ask, how do educators teach most effectively? How do learners learn most efficiently? There are probably as many different answers to these questions as there are models of instruction that are derived from distinct learning theories or paradigms. Different models of instruction imply different teaching/learning strategies and different methods of assessing student learning. One vehicle for assessing student learning is the use of quizzes and tests.

Although testing is undoubtedly viewed unenthusiastically by most students who, if given a choice, would choose to take as few tests as possible, research has shown that frequent testing in classrooms can produce significant academic improvement. The purpose of this study, then, was to examine how to structure assessments to maximize student learning and student success. Specifically, this study seeks to determine whether quiz- based learning which assessed students on their learning of smaller chunks of materials resulted in better learning outcomes when compared to exam-based learning which assessed students on larger sections of course material. For the purpose of this study, student learning outcomes were measured by the DWF rate defined as the percentage of students enrolled in the course either withdrawing before completion of the course or receiving a final course grade of D or F, chapter assessment performance, final exam

performance, retention as measured by average chapter assessment performance compared to final exam performance, and improvement as indicated by pre-and post-test measurements.

LITERATURE REVIEW

A review of relevant research has shown that frequency of testing positively impacts student achievement and learning in the following ways:

- Retrieval as demanded by testing aids later retention;
- Testing helps students identify gaps in their knowledge;
- Testing encourages students to study;
- Testing especially enhances performance in students with lower GPAs;
- Student performance on tests provides timely feedback to the instructor; and,
- Finally, students indicate higher course satisfaction for courses in which there was frequent use of testing.

Retrieval As Demanded By Testing Aids Later Retention

Frequent testing enhances student performance; the more a student studies specific material, the better material is remembered. A number of studies describe the positive impact of frequent testing on retention. Nearly fifty years ago, Izawa [1966] described the process of test-potentiated learning, showing that students learned more from taking a test and re-studying the materials than they would if they re-studied without taking an initial test. In short, taking a test enhanced the amount of learning in future study sessions.

According to Karpicke and Roediger [2007], the process of repeatedly retrieving information through taking tests enhances the cognitive processes that maximize learning and long-term retention. "…Tests not only assess learning but also enhance it" [2007, p 152]. A series of experiments on the impact of testing compared to studying showed that, of the two, testing had more impact on long-term retention [Karpicke and Roediger, 2007].

A number of other studies demonstrate that frequent testing aids students in remembering and retrieving information more effectively. Kika, McLaughlin, and Dixon [1992] studied the effects of frequent testing on the performance of students who were administered either a weekly or biweekly test. Results showed that performance was greater when students were given short quizzes on smaller units of instruction than on longer quizzes on bigger units of instruction. Positive effects from frequent testing were demonstrated in measurements of post-test achievement. Dunlosky, Rawson, Marsh, Nathan & Willingham [2013] similarly demonstrated that student performance was enhanced when given one exam per week compared to longer and less frequent assessments, such as 2-3 exams per semester.

In a study by Leeming [2002], students were taught the same content but were assessed at different intervals. Students in some of the classes were administered a short exam at the start of every class, where students in other sections took only four exams throughout the duration of the course. Results demonstrated that grades were significantly better in the "exam a day" class, there were fewer withdrawals from the class, and the students also performed better on a retention test than students in classes in which fewer exams were administered.

Testing Helps Students Identify Gaps In Their Knowledge

Frequent testing enables students to identify information they either do, or alternately, do not know. Having this knowledge, students can focus on learning the material they do not know. Research by Son and Kornell [2008] provides evidence that if students study material after a test, they spend more time re-studying material that they initially got wrong than on material they initially retrieved correctly. Dunlosky and Hertzog [1998] described the discrepancy reduction model which explains how students attempt to reduce the discrepancy between their current knowledge and the knowledge they hope to achieve. According to this framework, students will devote most of their study time to topics they do not know well, while spending less study on topics they know.

Testing Encourages Students To Study

Another benefit of frequent testing is that students are motivated to study on a more regular basis throughout the duration of the course. In courses that have only a midterm and a final exam, it is not unusual to find that students delay studying until the night before the exam. In a study in which tests were given either every day, every week, or every three weeks, Mawhinney, Bostow, Laws, Blumenfeld, and Hopkins [1971] found that more frequent testing throughout the semester encouraged more evenly spaced study, whereas in classes with less frequent testing, students tended to study only before the test. A number of studies found that students who were quizzed weekly on assigned reading material performed 4-24% better on midterm and final exams as compared to students who were not quizzed as often. An explanation, according to Tuckman, [as cited in Frost, 1999], is that students are less likely to procrastinate on studying if they know when quizzes will be given and what content will be included on the quizzes. In student surveys, students readily admitted that they would have been less likely to complete the assigned reading if quizzes would not have been given.

Roediger [as cited in Lahey, 2014] agrees and states that educators should be using quizzes and tests frequently to strengthen learning. He believes that periodic exams motivate students to "cram" material and regurgitate the information shortly thereafter, which is an ineffective strategy for long-term storage of knowledge.

Tuckman [as cited in Frost, 1999] compared weekly testing to the more common approach of administering a midterm and a final by stating, "By the time a midterm rolls around, students are already either successful or in big trouble. If teachers want to increase students' drive and get them to keep up with their schoolwork, we have to evaluate students' performance over shorter intervals of time."

Testing Especially Enhances Performance In Students With Lower GPAs

Many students enter college without many of the essential learning skills to be successful. They may lack study skills, note-taking skills, critical thinking skills, and the ability to evaluate their progress throughout the duration of the course. Higher education institutions have attempted to help lower-achieving students by providing a series of "Introduction to College 101" courses and non-credit remedial courses. However, a more effective approach to increasing performance of students with low GPAs may be to encourage these students to participate in classes in which frequent testing is part of the curricular design. Tuckman [as cited in Frost, 1999] found that students with low GPAs demonstrated improved performance and grades when encouraged to study for a weekly quiz, and earned better grades than their counterparts with average GPAs.

Tuckman suggests that forcing low GPA students to study regularly will help them perform to their maximum level of ability.

Pennebaker, Gosling, and Ferrell [2013] suggested that structuring courses to include frequent testing with immediate feedback enhances the academic performance of students. In their study, courses were designed to include daily online quizzes that provided immediate feedback. They found that performance on exams was about half a letter grade better than the previous semester, and that students who participated in the quiz classes also performed better in other classes, both in the same semester and in subsequent semesters. The daily quiz classes resulted in a 50% reduction in the achievement gap as measured by grades of students of varying socio-economic status groups. The authors attribute the positive changes to the curricular design that required students to develop more effective study skills to effectively prepare for the quizzes. In addition, the frequency of testing chunked the content into smaller segments that encouraged students to focus attention on relevant material. The immediate feedback after quizzes not only provided student with a vehicle for self-evaluation but also strengthened their rate of class attendance. Pennebaker [et. al., 2013] suggest that both short- and long-term performance gains can be attributed to participation in classes with frequent testing.

Student Performance On Tests Provides Timely Feedback To The Instructor

While testing is most often viewed as a means of formally assessing student learning, it is also a valuable mechanism to inform one's teaching. Upon review of each student's response to specific questions, instructors are likely to discover that many students lack understanding of a particular concept. This may serve as a signal that more time needs to be spent re-teaching that concept, rather than moving on to new material. A lack of student understanding may also indicate that another approach to teaching a concept may be necessary. In addition, an awareness of a student's performance on tests enables the educator to assess individual strengths and areas to be further developed. This information can be used in designing future individualized teaching/learning strategies.

Research by Kelly [1999] demonstrated that educators often assume that their students know more than they actually do. Testing often eliminates those lofty assumptions, and pinpoints students' knowledge more realistically. This knowledge helps educators target their teaching/learning strategies at the most appropriate level to match the students' level of understanding.

Students Indicate Higher Course Satisfaction For Courses In Which There Was Frequent Use Of Testing

Although initially students may voice complaints regarding the frequency of testing, a number of studies showed that by the end of the semester, students not only reported a preference for weekly testing [Kika, McLaughlin, and Dixon, 1992; Dunlosky, Rawson, Marsh, Nathan and Willingham, 2013], but the studies also showed that students' perceptions of course satisfaction was higher in courses with more frequent testing. Dunlosky, et.al, [2013] reported that students had more positive opinions of courses in which testing occurred more frequently. In a survey conducted by Lyle and Crawford [2011], students felt that frequent quizzes gave them a chance to practice for exam questions, communicated important course content, encouraged students not only to come to class more regularly but also to pay attention more closely, and finally, to more

thoroughly understand what they learned in each class session. In short, educators may be able to improve the affective outcomes of instruction by testing more frequently [Bangert-Drowns, Kulik, and Kulik, 1991].

Summary

Although students may complain about frequent testing, and the process of grading weekly quizzes may be daunting for educators, Roedinger [as cited in Lahey, 2014] has shown that "taking a test on material can have a greater positive effect on future retention of that material" than other study methods including "re-reading, highlighting, reviewing and writing notes, outlining material and attending study groups." He believes that when students are exposed to frequent quizzing, they are continuously engaged in the material, and have more ownership over their education. Testing, in addition to other teaching/learning strategies, may be a way of creating a more positive climate in the classroom.

HISTORICAL DATA

Twenty sections of Principles of Microeconomics taught by the same full time faculty instructor at the same institution over a period of nine academic years provides interesting historical data on teaching methods used along with the associated DWF rate. At this small private university, Principles of Microeconomics is often filled with students majoring in Business Administration, Accounting, International Business, Nutrition, and Fashion Merchandising. Although both a major and minor in economics were offered, it was a very small program graduating 2-6 students per year. As a result, the Principles of Microeconomics course was often filled with students who were not there because of their interest in the subject or their desire to learn about economics. When asked at the beginning of each semester why students were taking the course, the overwhelming response was "because I have to."

At this particular university, courses required for the major must be completed with a grade of C- or better. If the course is completed with a grade of D, it does not fulfill the major requirement and must be repeated. Given that this course is primarily taken as a major requirement, for students to successfully complete the course they must complete it with a grade of C- or better. Students who do not successfully complete the course would receive grades of D, F, or W. A grade of W or withdrawal occurs if the student chooses to withdraw from the course during the withdrawal period. The withdrawal period begins after the add/drop period which occurs at the beginning of the semester and ends about 2/3 of the way through the semester. Most students who withdraw do so in order to avoid receiving a grade of D or F in the course. As a result the DWF rate provides an accurate measure of the percentage of students who do not successfully complete the course.

The first things to note in Table 1 are the associated averages for the twenty sections taught. The average enrollment is 28.15 students per section with a standard deviation of 7.3504. Next, the average DWF rate is 28.43% with a standard deviation of 0.1105. Quizzes and exams make up 82.35% of the course grade on average, with a standard deviation of 0.0820. The typical cut off for a grade of C- is 64.73% with a standard deviation of 0.0473. Here, section 15 at 48.54% lies 3.42 standard deviations below the mean, implying that this data point can be considered an outlier.

Table 1: Historical Data

Section	Academic Year	Enrollment	DWF Rate	Lowest Grade Earning C-	Textbook	Aplia	Quizzes	Exams	Quizzes and Exams % of course grade	Extra Credit Offered
1	2004-05	23	26.09%	No Accurate Data	Mankiw	NO	2 ch quizzes	2 midterms + Comprehensive Final	80%	yes
2	2004-05	45	31.11%	63.45%	Mankiw	NO	4 ch quizzes	2 midterms + Comprehensive Final	90%	yes
3	2004-05	21	38.10%	63.45%	Mankiw	NO	4 ch quizzes	2 midterms + Comprehensive Final	90%	yes
4	2005-06	39	33.33%	70.00%	Mankiw	NO	3 quizzes (2-3 ch each)	3 exams	75%	yes
5	2005-06	15	33.33%	70.00%	Mankiw	NO	3 quizzes (2-3 ch each)	3 exams	75%	yes
6	2006-07	29	10.34%	61.95%	Mankiw	NO	12 ch quizzes	2 exams	75%	yes
7	2007-08	10	10.00%	62.59%	McConnell	NO	6 quizzes (1-2 ch each) (drop one)	2 exams	65%	no
8	2007-08	32	31.25%	65.75%	McConnell	NO	7 quizzes (1-2 ch each) (drop one)	2 exams	80%	yes
9	2007-08	30	30.00%	65.75%	McConnell	NO	7 quizzes (1-2 ch each) (drop one)	2 exams	80%	yes
10	2008-09	30	33.33%	65.22%	McConnell	NO	12 online ch quizzes (drop two)	2 exams	90%	yes
11	2008-09	30	13.33%	65.22%	McConnell	NO	12 online ch quizzes (drop two)	2 exams	90%	yes
12	2008-09	27	40.74%	65.01%	McConnell	NO	4 quizzes (2 ch each)	2 exams	80%	no
13	2009-10	29	41.38%	67.16%	McConnell	NO	10 online ch quizzes + 5 quizzes (1-2 ch each) (drop one quiz score)	2 exams	95%	no
14	2009-10	28	42.86%	67.16%	McConnell	NO	10 online ch quizzes + 5 quizzes (1-2 ch each) (drop one quiz score)	2 exams	95%	no
15	2009-10	29	13.79%	48.54%	McConnell	YES	5 ch quizzes	2 exams	75%	no
16	2010-11	31	25.81%	61.54%	McConnell	YES	10 online ch quizzes	2 exams	80%	built in 3%
17	2010-11	28	39.29%	62.64%	McConnell	YES	8 online ch quizzes (drop one) + pop quizzes	3 exams	90%	yes
18	2011-12	30	36.67%	70.00%	Mankiw e-text	YES	4 ch quizzes	2 exams	72%	no
19	2012-13	30	26.67%	67.63%	Mankiw e-text	YES	11 ch quizzes (drop one)	Comprehensive Final	85%	no
20	2012-13	27	11.11%	66.76%	Mankiw e-text	YES	11 ch quizzes (drop two)	Comprehensive Final	85%	no
MEAN		28.1500	28.43%	64.73%					82.35%	
STDEV		7.3504	0.1105	0.0473					0.0820	
N		20	20	19					20	

There are two sections in the table which should be removed from further analysis. Sections 7 and 15 were both atypical. Section 7 was the only section taught on a different schedule than the Fall or Spring semester. It was taught over 10 weeks on site for an employer for employees seeking their undergraduate degrees. As a result, the students in this section were mostly adult students whose tuition was paid for by the employer provided that the students received a grade of B- or better. The lower than average DWF rate for this section is most likely due to the lower than average enrollment, the adult population, and the tuition reimbursement model that was particular to this section. As mentioned earlier, section 15 had an abnormal grading scale. In this section, the cut off for a C- was 48.54%, well below the average of 64.73%. This shift in the grading scale caused a reduction in the DWF rate as well.

When these two sections are removed from the analysis, there are three remaining sections (6, 11, and 20) with DWF rates well below average. Section 6 (DWF rate = 10.34%) and 20 (DWF rate = 11.11%) both have in common the fact that they required in class quizzes for each chapter covered in the course. Interestingly, section 11 had a DWF rate of 13.33% while section 10 which was taught using the same methodology had a DWF rate of 33.33%. These two courses had the same enrollment, were taught on the same days, in the same semester, in the same room with a gap of 1.5 hours in between them. It is unclear why one had a 20% lower DWF rate than the other. It seems as though the only explanation lies in the abilities of the students enrolled in these courses.

After an initial examination of this data it appears as though the in class quiz for every chapter method may lead to a lower DWF rate that other methods used. Sections 6, 19, and 20 all used the method of having an in class quiz for each chapter and resulted in an average DWF rate of 16.04% with a standard deviation of 0.0921 well below the average DWF rate of 28.43%. By comparison, sections 10, 11, 13, 14, 16, and 17 all used an online quiz for each chapter and they resulted in an average DWF rate of 32.67% and standard deviation of 0.1136. Even with these very small samples, the in class quiz method leads to a significantly lower average DWF rate than the online quiz method (one tailed p-value = 0.0328). Other factors such as the text used (Mankiw or McConnell), the use of Aplia, the semester (Fall or Spring), if extra credit was offered, or if quiz grades were dropped did not have any significant effect. In all cases, the difference between the mean DWF rates for the two groups examined was not significantly different from zero.

This preliminary analysis shows that the use of in class quizzes for each chapter seems to result in a lower DWF rate than other assessment methods. Using quizzes on small groups of chapters did not lead to lower DWF rates, nor did using online quizzes for every chapter.
In order to further the analysis to ensure that it is indeed the method of assessment and not other aspects which are leading to these results it is necessary to run a controlled experiment.

CONTROLLED EXPERIMENT

In the 2013-14 academic year, a controlled experiment was run using two sections of the Principles of Microeconomics; one offered in the Fall semester and the other offered in the Spring semester. In both sections, the enrollment was 30 students, the Mankiw e-text with Aplia was used, quizzes and exams made up 85% of the final course grade, and no extra credit was offered. In the Fall, the lowest grade earning a C- was 66.19%, while in the Spring it was 64.43%, both consistent with the historical average of 64.73%.
At the beginning of each section students were given a pre-test of 85 multiple choice questions which was the multiple choice portion of the final exam. This pre-test was used to control for existing knowledge of economics which might potentially skew the results. In the Fall semester,

for the 28 students who completed the pre-test the average was 41.13% with a standard deviation of 0.0794 and a high score of 57.65%. In the Spring semester, for the 29 students who completed the pre-test the average was 37.89% with a standard deviation of 0.0846 and high score of 64.71%. Statistically there is no difference between the two means (two tailed p-value = 0.1420). We can be confident that students entering the course had approximately the same level of economic knowledge in both semesters.

The only difference between the two sections lies in how the knowledge of the individual chapters was assessed. In the Fall semester the course was taught using 12 quizzes, one for each chapter, along with a comprehensive final exam. In the Spring semester, the course was taught using two large exams, composed of 6 chapters each, along with the same comprehensive final exam.

Given the results of the historical data, it was presumed that the quiz method would lead to a lower DWF rate than the exam method. Given that many of the students in the course tend to be first year students, the Fall should experience more students "adjusting to college" than the Spring. As a result, it was decided to use the quiz method in Fall and the exam method in Spring in order to potentially bias the results towards the Spring. All quizzes and exams (except the comprehensive final exam and pre-test) were returned to students during the semester to aid in their preparation for the comprehensive final exam. Given that the Spring exams were created from the Fall quizzes this should also create bias in favor of higher grades in the Spring if students were to access the quizzes used in the Fall semester from students who took the course then.

Table 2: Course Grading Weights

Item	**Fall 2013 Quiz Method**	**Spring 2014 Exam Method**
Pre-Test	1%	1%
Paper	1%	1%
Aplia HW	13%	13%
Quizzes	60%	X
Exams	X	60%
Comprehensive Final Exam	25%	25%

In the Fall semester 12 quizzes, one per chapter, were given with the lowest two quiz grades dropped meaning that the remaining 10 quizzes counted for 6% each. In order to make the grading symmetric in the Spring semester, two exams were given worth a total of 60%. Each exam was made up of 6 chapter sections, comprised of the chapter quizzes used in the Fall semester. In the Spring, Exam I covered Chapters 1, 2, 4, 5, 6, 7 while Exam II covered Chapters 21, 10, 11, 13, 14, 15. The same chapters were covered in the same order in the Fall semester. At the end of the Spring semester, the grades on the two lowest chapters were dropped so the overall exam grade was comprised of 10 chapters each worth 6%.

A general description of the 12 chapter assessments is included in Table 3. The final exam was composed of 85 multiple choice questions, the same questions used on the pre-test, and 15 points of short answer questions. Table 4 shows how these questions were dispersed amongst the various chapters.

Table 3: Chapter Assessments

Chapter	Topic	Points	Description of Assessment
1	Ten Principles of Economics	15 points	15 multiple choice questions covering Principles 1-7
2	Thinking Like an Economist	15 points	15 multiple choice questions covering the circular flow diagram, PPF, micro vs. macro, positive vs. normative, and factors of production
4	Supply and Demand	10 points	2 multiple choice questions, 4 supply and demand graph questions
5	Elasticity	10 points	2 multiple choice questions, 4 elasticity problems incorporating calculations and interpretation
6	Price Controls and Taxes	27 points	2 multiple choice questions, 4 price control analyses, and a tax implementation problem
7	Consumer, Producer, and Total Surplus	15 points	15 multiple choice questions
21	Theory of Consumer Choice	18 points	10 multiple choice questions, 3 graph interpretation questions, and an optimization problem
10	Externalities	18 points	18 multiple choice questions
11	Public Goods and Common Resources	18 points	18 multiple choice questions
13	Costs	16 points	2 multiple choice questions, calculation and interpretation of accounting versus economic profit, and cost and productivity calculations
14	Perfect Competition	34 points	Short answer questions relating to the characteristics of perfect competition and an explanation of adjustment to LR equilibrium, calculation of SR and LR profit maximizing quantities, SR profits, LR equilibrium price, and shutdown price.
15	Monopoly	30 points	Short answer questions relating to characteristics of monopoly, the dilemma of regulation, two-way price discrimination, and barriers to entry, calculations of efficiency under a profit maximizing monopoly with no price discrimination vs. a perfectly price discriminating monopolist.

Table 4: Final Exam Coverage

Chapter	Multiple Choice (85 Points)	Short Answer (15 points)
1	6 questions	
2	5 questions	
4	10 questions	
5	7 questions	2 points (calculations)
6	8 questions	
7	8 questions	
10	8 questions	3 points (graph interpretation)
11	6 questions	4 points (categorization)
13	7 questions	
14	8 questions	1 point (calculation)
15	7 questions	
21	5 questions	5 points (graph interpretation)

RESULTS

The results of the experiment offer some differing findings. In terms of the DWF rate, the model of frequent testing led to significantly better results. On the final exam, however, performance was essentially the same regardless of the assessment method used during the course. The average improvement from pre-test to post-test was also no different. It appears that the students in the section using the quiz method initially learned more of the material, but that those gains were then lost by the time of the final exam.

DWF Rate

The results of the controlled experiment are quite significant when looking at the DWF rate. In the Fall semester using the quiz method the DWF rate was 10%. In the Spring semester using the two large exams the DWF rate was 56.67%. The exam method caused a tremendous increase in the DWF rate.

It seems as though the organization of the course and the frequency of assessment have a significant impact on the DWF rate. Much of the increase in the DWF rate was caused by withdrawals from the Spring section following the first exam. In the Spring, 46.67% of the students withdrew from the course compared to only 3% in the Fall, confirming the results by Leeming [2002] that there were fewer withdrawals from class when more regular assessments were used.

This increase in withdrawals resulted from the poorer performance on the first six chapter assessments in the Spring. Scores on Exam I ranged from 34% to 98% with 43% of students scoring 70% or above and 37% of students scoring in the F range (below 60%). By comparison in the Fall, a similar analysis of the first six chapters, shows grades ranging from 50.31% to100% with 77% of students scoring 70% or above and only 6% of students in the F range. In the Spring, by having only one exam instead of multiple quizzes, the first assessment in the course, Exam I, completely determined the outcome of the course. Given their low performance on the first exam

and that the second would not occur until after the final date to withdraw from the course, many students in the Spring chose to withdraw rather than risk a potential final course grade of D or F.

Also, it seems that in the Spring, students did not have the opportunity to learn what to expect in terms of assessment or how to correct their study habits. If we look at the performance on Ch1 versus Chapter 2 in both semesters we see some significant differences. Looking at Table 5, we see that in both semesters, the students who scored below 70% on Chapter 1, averaged 57.33% or an F between them. In the Fall those 10 low performing students on Chapter 1, then improved to an average of 80.67% on Chapter 2. In the Spring, the 15 low performing students on Chapter 1 increased their average to 63.56% on Chapter 2. By having feedback and time to act on it, the Fall students were able to improve their scores by almost 4 times as much as the Spring students.

Table 5: Improvement from Ch.1 to Ch. 2

	Fall	**Spring**
Number of students scoring below 70% on Ch 1	10	15
Their Average on Ch 1	57.33% F	57.33% F
Their Average on Ch 2	80.67% B	63.56% D
Improvement	23.34%	6.23%

These results seem to confirm the findings of Pennebacker [et al, 2013] in that the provision of more immediate feedback improved student performance in the Fall. They also seem to confirm the findings of Tuckman [as cited in Frost, 1999] in that by taking only one exam in the Spring, the students who were prepared were well prepared and performed well over all six chapters while those that were in "big trouble" stayed that way for multiple chapters. Success on the first exam was essentially determined by the studying that occurred well before the exam itself.

Chapter Assessment Performance

In each section of the course, a few students did not consent to participation in this study. As a result further analysis and comparisons between the two sections will not include all of the enrolled students in the sections. In the Fall semester 27 students consented to participate in the study and in the Spring semester 26 students consented to participate in the study. The results of the participating students generally coincide with the results for the full set of students enrolled in the sections.

One of the most important results from Table 6 is the overall chapter average by course. In the Fall, the average score on all 313 chapter assessments taken was 78.31%. In the Spring when the quizzes are combined into exams the average drops to 61.98% over 240 chapter assessments. This drop of almost 16.33% is highly statistically significant with a p-value of 0.000. By combining the quizzes into two large exams we see that there is a significant decline in performance. Again, these results seem to confirm those of Pennebaker [et al, 2013] that frequent testing with more immediate feedback enhances the academic performance of students.

Table 6: Performance on Chapter Assessments

	Fall			Spring			T-test for difference of two means		
Assessment	MEAN	STDEV	N	MEAN	STDEV	N	Test Statistic	One-tailed p-value	Two-tailed p-value
Ch 1	77.18%	0.1585	26	72.31%	0.1651	26	1.0850	0.1416	0.2831
Ch 2	85.68%	0.1222	27	73.08%	0.1229	26	3.7421	0.0002	0.0005
Ch 4	77.21%	0.2337	26	45.96%	0.2871	26	4.3044	0.0000	0.0001
Ch 5	79.17%	0.1458	26	51.92%	0.2569	26	4.7039	0.0000	0.0000
Ch 6	75.93%	0.1997	27	57.34%	0.1909	26	3.4618	0.0005	0.0011
Ch 7	74.32%	0.2065	27	65.90%	0.2350	26	1.3870	0.0858	0.1715
Ch 21	77.62%	0.1300	26	60.71%	0.2087	14	3.1624	0.0015	0.0031
Ch 10	83.44%	0.1193	26	71.03%	0.2011	14	2.4579	0.0093	0.0187
Ch 11	82.00%	0.0979	25	76.98%	0.1459	14	1.2850	0.1034	0.2068
Ch 13	69.64%	0.2168	26	57.47%	0.2031	14	1.7300	0.0459	0.0918
Ch 14	89.93%	0.1351	26	71.48%	0.2097	14	3.3839	0.0008	0.0017
Ch 15	67.33%	0.1948	25	44.17%	0.2463	14	3.2374	0.0013	0.0025
Overall Chapter Average	78.31%	0.1770	313	61.98%	0.2331	240	9.3645	0.0000	0.0000
Pre-Test	41.65%	0.0816	25	38.19%	0.0874	26	1.4599	0.0754	0.1507
Final Exam	63.48%	0.1457	26	63.04%	0.1316	14	0.0941	0.4628	0.9255
Final Exam Multiple Choice (MC)	64.12%	0.1406	26	65.88%	0.1303	14	-0.3871	0.3505	0.7009
Final Exam Short Answer (SA)	59.87%	0.2732	26	46.90%	0.1656	14	1.6178	0.0570	0.1140

The overall decline shown by the overall chapter average results from the fact that students in the Spring scored significantly lower than students in the Fall on most of the individual chapters. Only for chapters 1, 7, and 11 were the average scores in the Spring not significantly lower than those in the Fall. These chapters were all assessed using exclusively multiple choice questions pointing to the possibility that students who have to "wait" to take an exam may perform relatively better on multiple choice than short answer questions.

Final Exam Performance

Although the chapter assessment performance was quite different between the two sections, the final exam performance was almost exactly the same. The average final exam score was 63.48% for the 26 students participating in the Fall versus 63.04% for the 14 students participating in the Spring. These results on the final exam mirror results from an article in the Chronicle of Higher Education regarding principles of microeconomics where at George Washington University, "On the final, with its 100 possible points, the average was 63." [Berrett, 2012)].

In terms of learning it appears that the students in the Fall learned more initially as shown by their higher chapter assessment scores, but any gains were removed by the time of the final exam.

Retention: Chapter Assessment Performance Compared To Final Exam Performance

When looking at Table 7 and the scores for the 40 individual students who completed the course and participated in the study, we see that the average on the chapter sections ranges from 49.59% to 97.77%. By comparison the final exam scores range from 35% to 90%. Only two

students (both in the Spring) scored higher on the final exam, than their average chapter score. For the others, something was lost between the chapter assessment and the final exam. The most dramatic decline, occurring in the Fall, was 33.59%. For the students who declined, the average decline in the Fall was 15.30% with a standard deviation of 0.0869 and N=26. In the Spring, the average decline was 6.69% with a standard deviation of 0.0390 and N=12. In the Spring, the average decline is significantly less than the Fall (one tailed p-value = 0.0012).

At some level, these numbers appear to be informing us about the retention of information from the time of the chapter assessments to the time of the final exam. Generally, retention appears to be worse in the Fall because more students have large declines, but it may be that the spring students never really learned as much so they did not have the opportunity to forget as much. Here is appears as though frequent testing did not have any impact on overall retention given that students performed at the same level on the final exam regardless of the assessment method used during the course. Learning material in small chunks may not lead to any better retention than learning through large exams. Such results contradict the findings of Leeming [2002] which showed that students who had more frequent testing performed better on a retention test than students in classes where fewer exams were administered.

Improvement: Pre-Test Performance Compared To Post Test Performance

When looking at the pre-test compared to the multiple choice portion of the final exam, we see that students did learn something in the course. One student in the Fall scored exactly the same on both the pre-test and post-test, but for the remainder, the course added some knowledge and students improved their scores with the maximum increase at almost 50%. In the Fall, the average improvement was 22.25% with a standard deviation of 0.1025 and N=24. In the Spring, the average improvement was 23.28% with a standard deviation of 0.1183 and N=14. The differences between these means are not significantly different, implying that for most students, improvement was not impacted by the course structure.

Table 7: Individual Student Performance

Student	Semester	ChAVG	Final Exam	Final Exam minus ChAVG	PreTest	Final MC	Final MC minus PreTest
1	Fall	79.59%	46.00%	-33.59%	44.71%	47.06%	2.35%
2	Fall	68.30%	40.00%	-28.30%	38.82%	38.82%	0.00%
3	Fall	93.05%	67.00%	-26.05%	43.53%	71.76%	28.24%
4	Fall	60.96%	35.00%	-25.96%		36.47%	
5	Fall	71.65%	46.00%	-25.65%	23.53%	52.94%	29.41%
6	Fall	76.53%	54.00%	-22.53%	34.12%	54.12%	20.00%
7	Fall	84.38%	62.00%	-22.38%	48.24%	68.24%	20.00%
8	Fall	69.38%	47.00%	-22.38%	38.82%	47.06%	8.24%
9	Fall	81.90%	61.50%	-20.40%	40.00%	63.53%	23.53%
10	Fall	84.21%	65.00%	-19.21%	43.53%	67.06%	23.53%
11	Fall	73.14%	57.50%	-15.64%	40.00%	62.35%	22.35%
12	Fall	84.51%	69.00%	-15.51%	43.53%	74.12%	30.59%
13	Fall	88.79%	74.00%	-14.79%	56.47%	77.65%	21.18%
14	Fall	78.31%	64.00%	-14.31%	48.24%	64.71%	16.47%
15	Fall	64.91%	51.00%	-13.91%	36.47%	54.12%	17.65%
16	Fall	95.50%	85.00%	-10.50%	44.71%	83.53%	38.82%
17	Fall	79.83%	70.00%	-9.83%	54.12%	71.76%	17.65%
18	Fall	77.50%	68.00%	-9.50%	42.35%	63.53%	21.18%
19	Fall	69.12%	60.50%	-8.62%	43.53%	55.29%	11.76%
20	Fall	94.53%	86.00%	-8.53%		84.71%	
21	Fall	71.38%	63.00%	-8.38%	29.41%	62.35%	32.94%
22	Fall	97.77%	90.00%	-7.77%	57.65%	88.24%	30.59%
23	Fall	81.56%	76.00%	-5.56%	38.82%	72.94%	34.12%
24	Fall	94.16%	90.00%	-4.16%	47.06%	88.24%	41.18%
25	Fall	63.14%	60.00%	-3.14%	41.18%	60.00%	18.82%
26	Fall	64.11%	63.00%	-1.11%	32.94%	56.47%	23.53%
27	Spring	57.86%	45.00%	-12.86%	35.29%	48.24%	12.94%
28	Spring	77.09%	65.50%	-11.59%	35.29%	68.24%	32.94%
29	Spring	65.92%	55.00%	-10.92%	34.12%	58.82%	24.71%
30	Spring	88.71%	80.00%	-8.71%	35.29%	83.53%	48.24%
31	Spring	61.27%	54.00%	-7.27%	42.35%	58.82%	16.47%
32	Spring	78.08%	72.00%	-6.08%	49.41%	74.12%	24.71%
33	Spring	87.95%	82.00%	-5.95%	64.71%	84.71%	20.00%
34	Spring	57.62%	52.00%	-5.62%	40.00%	55.29%	15.29%
35	Spring	56.31%	51.00%	-5.31%	37.65%	56.47%	18.82%
36	Spring	54.79%	50.00%	-4.79%	40.00%	51.76%	11.76%
37	Spring	82.75%	82.00%	-0.75%	44.71%	84.71%	40.00%
38	Spring	71.37%	71.00%	-0.37%	56.47%	74.12%	17.65%
39	Spring	49.59%	52.00%	2.41%	44.71%	51.76%	7.06%
40	Spring	55.75%	71.00%	15.25%	36.47%	71.76%	35.29%

CONCLUSION

Course structure has a significant impact on the DWF rate as shown by the increase in DWF rate from 10% to 56.67%. Course structure also impacted performance on the chapter assessments as shown by the decline in overall chapter performance from 78.31% to 61.98%. Within the course, students seem to perform much better when they have in class quizzes on every chapter versus large exams. However, when it comes to final exam performance, the results are essentially exactly the same. In fact, it appears as though any additional learning gained from the quiz method initially is lost by the time students take the comprehensive final exam. The students in the Fall clearly have a more significant decline in their performance on the final exam compared to their chapter assessments than the students in the Spring. It appears as though the students in the Spring struggled to learn the material initially, but maintained whatever the learned through the final exam while the students in the Fall initially learned significantly more material, but then seemed to forget significantly more of the material by the time the final exam occurred. All in all, the multiple in class quiz method has clear impacts on the DWF rate, but did not lead to improvement on the comprehensive final exam. This appears to be a better way to teach microeconomics in that it kept more students engaged in the course for a longer period of time, but if we are looking at overall microeconomic knowledge, it appears that any short term gains that result from frequent quizzing are lost by the time the semester ends.

DIRECTIONS FOR FUTURE RESEARCH

There are two sets of preliminary results in this study that could provide interesting directions for future research. First, the results of the historical data show that in class quizzing leads to significantly lower DWF rates than online quizzes. Faculty may be concerned about using class time to conduct quizzing but it appears as though online quizzing may not be as effective. More thorough research on this topic should be conducted. Second, students who were subjected to the exam method, performed relatively better on multiple choice portions of those exams than short answer portions. Exam design may also have an impact on student performance, especially as the time from initial learning to assessment increases. Further research on this topic could be useful as well.

REFERENCES

Bangert-Drowns, R.L., J. A. Kulik, and Chen-Lin C. Kulik. 1991. "Effects of Frequent Classroom Testing," *The Journal of Educational Research,* 85 (2), 89-99.

Berrett, D. "3 Colleges' Different Approaches Shape Learning in Econ 101," *Chronicle of Higher Education.* (June 18, 2012).

Dunlosky, J., K. A. Rawson, E. J. Marsh, M. J. Nathan, and D. T. Willingham. 2013. "Improving Students' Learning with Effective Learning Techniques: Promising Directions from Cognitive and Educational Psychology." *Psychological Science in the Public Interest.* 14 (1) 4-59.

Dunlosky, J., and C. Hertzog. 1998. "Training Programs to Improve Learning in Later Adulthood: Helping Older Adults Educate Themselves. In *Metacognition and Educational Theory and Practice*. Edited by D. J. Hacker, J. Dunlosky, and A. C., Greaser. Mahwah, NJ: Erlbaum.

Frost, P. "Students Get Better Grades When Tested Frequently in Class." *The Ohio State University Research News*. Retrieved from http://researchnews.osu.edu/archive/motivat.htm, (1999).

Izawa, C. 1966. "Reinforcement-test sequences in paired-associate learning." *Psychological Reports*, 18, 879-919.

Kelly, C. M. 1999. "Subjective Experience as a Basis for 'Objective' Judgments: Effects of Past Experience on Judgments of Difficulty." In *Attention and Performance XVII.* Edited by D. Gopher & A. Koriat. MIT Press.

Kika, F.M., T. F. McLaughlin, and J. Dixon. 1992. "Effects of Frequent Testing of Secondary Algebra Students." *Journal of Educational Research*, 85, 159-162.

Lahey, J. "Students Should Be Tested More, Not Less." 2014. *The Atlantic*. Retrieved from http://www.theatlantic.com/education/archive/2014/01/students-should-be-tested-more-not-less.

Leeming, F. C. 2002. "The Exam-a-Day Procedure Improves Performance in Psychology Classes." *Teaching of Psychology*, 29, 210-212.

Karpicke, J. D. and H. L. Roediger III. 2007. "Repeated Retrieval During Learning is the Key to Long-Term Retention." *Journal of Memory and Language.* 57, 151-162.

Kyle, K. B., and N. A. Crawford. 2011. "Retrieving Essential Material at the End of Lectures Improves Performance on Statistics Exams." *Teaching of Psychology*, 38, 94-97.

Mawhinney, V.T., D. E. Bostow, D. R. Laws, G. J. Blumenfeld, and B. L. Hopkins. 1971. "A Comparison of Students Studying-Behavior Produced by Daily, Weekly, and Three-Week Testing Schedules. *Journal of Applied Behavior Analysis*, 4, 257-264.

Pennebaker, J. W., S. D. Gosling, and J. D. Ferrell. "Daily Online Testing in Large Classes: Boosting College Performance while Reducing Achievement Gaps." *PLoS ONE* (November, 2013) doi:10.1371/journal.pone.007977.

Roedinger, H. L., III, A. L. Putnam, and M. Smith. 2011. "Ten Benefits of Testing and Their Applications to Educational Practice," *Psychology of Learning and Motivation.* 55, 1-36.

Son, L. K., and N. Kornell. 2008. "Research on the Allocation of Study Time: Key Studies from 1890 to the Present (And Beyond). In *The handbook of memory and meta-memory.* Edited by J. Dunlosky, & R. A. Bjork. Hillsdale, NJ: Psychology Press.