

THE EFFECT OF TEST RETAKES ON LONG-TERM RETENTION

by

BRADLEY A. PAFF

A Master's Paper

Submitted to the Graduate Faculty

in partial fulfillment of the requirements for the Degree of

Master of Science in Education - Physics

---

Advisor's Signature

---

Date

University of Wisconsin – River Falls  
November 26, 2012

## ABSTRACT

This study examines the effect of an optional retesting strategy on long-term retention of course content. It does so by comparing the number of retests a student chooses to take during the semester to the student's performance on the semester exam. The rationale for this instructional practice is that when a student takes the time to restudy material for a retest, the mastery level and long-term retention of that content should increase, which would lead to better performance on cumulative assessments such as the semester exam. This study does find a small trend toward higher performance on semester exams for students who choose to take more retests throughout the semester, though the effect size is not statistically significant.

## TABLE OF CONTENTS

Literature Review	1
Purpose of the Project	5
Methods used	7
Results	10
Table 1. Number of Retests Taken and Semester Exam Average within Ability Groups	11
Graph 1. Number of Retests vs. Semester Exam Score of Each Student	12
Table 2. Number of Retests Taken and the Difference between Sem. Exam Scores and Students' Average on Original Unit Tests	13
Graph 2. Retests Taken vs. Difference Between Semester Exam and Original Test Average	14
Analysis	14
Summary	17
Works Cited	18

## Literature Review

Two of the most important questions for every educator to ask are, “How much are my students learning?” and “How do I help those students who are struggling to learn?” The first of those questions forms the rationale for the many forms of assessment to which students are subjected. The second question leads educators to constantly rethink their approach to teaching, including the forms and frequency of assessments that are used to measure student learning.

Before the twentieth century, assessment commonly took the form of written essays or one-on-one oral examinations. With the advent of written objective tests in the early 1900s, the administration and grading of written exams became a much faster process, and for the first time it became convenient to give frequent written assessments to entire classes of students. A number of studies conducted during the 1930s examined the effects of frequent testing, exploring the idea that higher testing frequency would encourage more frequent review of course content and result in better recall and improved learning (Bangert-Drowns, *et al.*, 1991). These studies reported mixed results, with some finding that more frequent testing improved student learning, and some finding the opposite. The usefulness of frequent testing has been debated and studied ever since. Bangert-Drowns, *et al.* (1991) performed a meta-analysis of 35 previous studies conducted between 1929 and 1989, and found that 29 of the studies reported positive effects from frequent testing, while 6 reported negative effects. Of these, four studies showed students in the frequent-testing treatment groups increase their exam scores by 0.70 standard deviations or more, while one study showed a decrease of -0.80 standard deviations in exam scores as a result of frequent testing. The complete analysis indicated that 13 of those studies reporting a positive effect determined the effect to be

statistically significant, while only one of the negative reports was found to be statistically significant. Overall, their analysis found that increased frequency of testing improved student achievement by an average of 0.23 standard deviations, but with a diminishing rate of return as the frequency of testing increased. The biggest positive outcome of frequent testing may be in the effect it has on student attitudes, where students who were tested more often reported a higher level of satisfaction with their instruction, a finding which has been repeated by several studies (Bangert-Drowns, *et al.*, 1991; Friedman, 1987; Catanzano & Wilson, 1977; Davidson, 1984).

Many of these studies focused purely on the frequency of regular assessments such as quizzes, unit tests, or midterm exams. However, beginning in the late 1960s, the growing popularity of teaching practices such as mastery learning and personalized systems of instruction led to a number of studies on the effects of students *retaking* versions of previous tests in order to demonstrate an increased mastery of the content (Juhler, *et al.* 1998; Catanzano & Wilson, 1977). As recently as last year, Diegelman-Parente (2011) studied the effects of the mastery learning approach, where students are given multiple opportunities to learn material and demonstrate their knowledge by retesting. The study found positive student attitudes toward the mastery learning approach, which was attributed to students feeling more incentive to relearn the material they had not yet mastered. In further support of the mastery approach, Diegelman-Parente cites earlier studies by Benjamin Bloom (1978), who found that “about 80% of students in a mastery learning class reach the same final achievement level as the top ~20% of the students under conventional instruction” (as cited in Diegelman-Parente, 2011).

While the mastery learning approach forces struggling students to retake assessments in order to complete course requirements, other instructional methods have developed the use of optional retesting to promote higher student achievement. Catanzano and Wilson (1977) compared seventh grade science students in three groups, including a control group with no retesting opportunity, a group with a mastery-learning format, and another group with optional retesting. They found that while the mastery and optional retesting formats were equally effective at improving student achievement compared to the control group, the optional retesting format was viewed more positively than the mastery or no-retest methods on measures of student attitude toward their instruction.

Optional retesting has since been explored in several studies in a variety of disciplines. Davidson (1984) studied college students in introductory psychology courses, and found that 48% of students who scored an “F” on an original midterm exam were able to raise their exam score by a letter grade on a retest, along with increases by 52% of students originally earning a “D” and 57% of students earning a “C” on their original exams. In addition, 80% of students reported less test anxiety because of the option to retake the exam, even among those who did not use the retake option. Geving, *et al.* (2005) studied potential real estate agents during their attempts to pass a licensing exam, exploring the effects of repeated content on retake tests, as well as the effect of retaking parallel forms of the same test multiple times. Findings from this study included the observation that repeated use of the same questions does not necessarily contribute to higher test scores, as scores increase over the first few retakes, but then begin to plateau, similar to the Bangert-Drowns, *et al.* (1991) study showing the diminishing returns of frequent testing. This study also reported that examinees who retook the tests at longer

intervals after the previous test experienced a greater gain in test scores, presumably due to taking more time to study the material. Abraham (2000) used college algebra and calculus courses to explore the relative effect of two grading policies: dropping the lowest test score versus allowing students to retake exams. The study found that students who were allowed to drop their lowest test score averaged 64% on the final exam, while students who were allowed to retake tests averaged 73%. This effect was presumed to be due to students in the retake sections taking more time to study and relearn the material between the original and retake tests, leading to greater long-term retention. In contrast, a study of college algebra students by Juhler, *et al.* (1998) found that although 90% of students using the retake option improved their grade from the test to the retest, there was no significant correlation between the number of tests retaken and the final exam score. A study of college statistics students by Friedman (1987) found that students who retook two or three exams during the course scored significantly higher on the final exam than students who retook one or none of the exams during the course. One limitation to this study is that it only included students who had an average grade of 'B' on the initial exams.

There is still uncertainty as to the effects of retesting. While there is clear evidence that students perform better on the retake exams themselves, some researchers attribute this improvement mostly to practice effects and test familiarity (Friedman, 1987; Catanzano & Wilson, 1977), while others attribute the higher retest grades to the feedback that the initial exam provides about students' weaknesses and the time students spend restudying the material (Juhler, *et al.*, 1998; Davidson, 1984; Geving, *et al.*, 2005). Some of the studies, including those by Friedman (1987) and Abraham (2000), have found improved student achievement at the end of the course, while others, especially

Juhler, *et al.* (1998), find no increase in performance. One outcome the studies seem to agree on is that the ability to retake exams lowers students' test anxiety on the initial exams, and leads to higher student satisfaction with the course and instruction overall (Davidson, 1984; Catanzano & Wilson, 1977; Friedman, 1987; Juhler, *et al.*, 1998).

As pointed out by Juhler, *et al.* (1998), students in certain fields of study such as mathematics should benefit from an optional retesting approach, because the first test provides a learning experience and feedback that is "particularly important in a sequential curriculum because each concept and procedure serves as a stepping stone to more complex concepts and procedures." Certainly, many instructors have found that once students get lost in a sequential curriculum, those students' chances for success decrease greatly. Many of these students will be likely to drop out of the class after failing an exam if the opportunity exists, or languish and fail the course if dropping is not an option.

### **Purpose of the project**

The high school science department connected to the current study noticed the same effect, where students who fail a test in a sequential curriculum will rarely make the effort to relearn the missed concepts, leading to knowledge gaps that continue to hamper their understanding of future units. This cycle often continues until the student is so far behind that he or she drops the course, or simply gives up any serious attempt to learn. It is exactly this idea that prompted the current study of optional test retakes.

This study has been conducted at Medford Area Senior High, located in Medford, Wisconsin. The community of Medford has a population above 4,300 residents, many of them employed in manufacturing, with significant agriculture in the surrounding rural area. The school district covers an area of over 425 square miles, with a total student

population of about 2,150. Approximately 650 of these students are enrolled in the high school, where the study was conducted. District-wide, about one in three students qualifies for free or reduced lunch based on family income. The district's graduation rate is typically 96%, with nearly half of graduates planning to attend college. Students in the study were enrolled in the researcher's science courses, which had an average class size of twenty students in the grade nine physical science courses, and sixteen students in the chemistry courses.

Since 2005 the district has employed a professional learning community philosophy, with the staff members organized into collaborative teams. The six teachers of the high school science and agriculture departments, including this researcher, make up one such team. The main goal for our team has been to examine our assessment practices to evaluate how much our students are learning, and then seek ways to help those students who are struggling to learn the essential course content. At the beginning of the 2009-10 school year, our science/agriculture team implemented an optional test-retake policy, which is described in the 'Methods' section below. We expected a retesting option would give struggling students extra motivation to relearn material that they may need in future units. To determine the effectiveness of this policy, our team began compiling data on students' test and retest scores, as well as the overall ability level of the students using the retest option. The current study is a more detailed analysis of the data representing the students from the researcher's courses only, as the other members of the team collected and recorded their data in ways that were not compatible with the format of this study.

As with the previously mentioned studies on mastery or optional retesting strategies, our initial data supported the observation that most students receive higher

scores on the retests than on the original tests. As a result, one of the main questions our team discussed was whether this practice is simply a form of grade inflation, or if the option to retake tests actually leads to increased long-term retention of the course content.

The purpose of the current study was to further explore the effect of test retakes on long-term retention. This study was intended to address gaps in the body of study by examining performance of students from all ability levels, and focused specifically on an optional retesting strategy. It was hypothesized that students who use the retesting option would spend more time restudying the course material, and therefore gain greater understanding of the content. This would be measurable in the form of higher scores on the semester exam than students who use the retake option less often or not at all.

### **Methods used**

Data for this study were collected during the 2010-11 academic year, from students enrolled in the researcher's physical science and chemistry classes. Only students who were enrolled in the researcher's class for an entire semester and completed the final exam at the end of the semester were included in the study. During the first semester, this included a total of 61 grade nine students in three sections of physical science and 32 students from grades ten through twelve in two sections of chemistry. During the second semester, there were 61 grade nine students in three sections of physical science, and 30 students from grades ten and eleven in two sections of chemistry. In total, there were 184 student-semesters represented in the data.

All students in the study had the option to take one retest of each chapter or unit test. Students were not required to complete specific additional work to prepare for the retest, but were required to wait at least two school days between receiving feedback on

the original test and taking the retest. Each retest consisted of a parallel test with size, structure, and difficulty similar to the original test. If a student chose to take a retest, the retest score replaced the original test score, even if the score on the retest was lower.

The following data was collected from the course gradebooks for each student in the study: original scores on each chapter or unit test, retest scores for any chapter or unit tests that were retaken, and score for each semester final exam. During the first semester, there were six chapter/unit exams given in both physical science and chemistry classes. During the second semester, there were three chapter/unit exams given in physical science, and seven given in the chemistry classes.

Chapter/unit tests and retests in the physical science course consisted of a mix of question types, mostly short-answer and problem-solving, along with fill-in-the-blank, multiple-choice, and matching. Semester exams in physical science were approximately half multiple-choice and half problem-solving. In the chemistry course, chapter/unit tests and retests were approximately equal portions of multiple-choice and problem-solving, while the semester exam consisted entirely of multiple-choice questions.

The main focus of the study was to determine if students who used the retest option more times during the semester scored higher on their semester exams than students who retested fewer times or not at all. To control for the wide range of ability between students, the data were sorted into groups based on the average of each student's original test scores. These groups were delineated by the letter grades within the standard grading scale at the researcher's high school:

A: 90% and above

B: 80-89.99%

C: 70-79.99%

D: 60-69.99%

F: below 60%

Once the student data were sorted based on the average of the original test scores, they were further sorted by the number of retest opportunities each student had used during the given semester. These data were then graphed using Vernier Software's Logger Pro 3.7, comparing the number of retests taken to the student's final exam score. Analysis was done using the software's linear regression analysis feature for the data within each letter-grade-based ability group. This analysis was performed to examine the effects of the number of retests on the final exam score.

In a further analysis, the data were resorted first by the number of retests each student had taken during the semester, then by the difference between the semester exam score and the average score on the original tests. These data were graphed to compare the number of retests taken and the difference between the student's final exam and original test averages. In this comparison, students were not sorted by ability group. A linear regression analysis was also performed using the Logger Pro software.

## **Results**

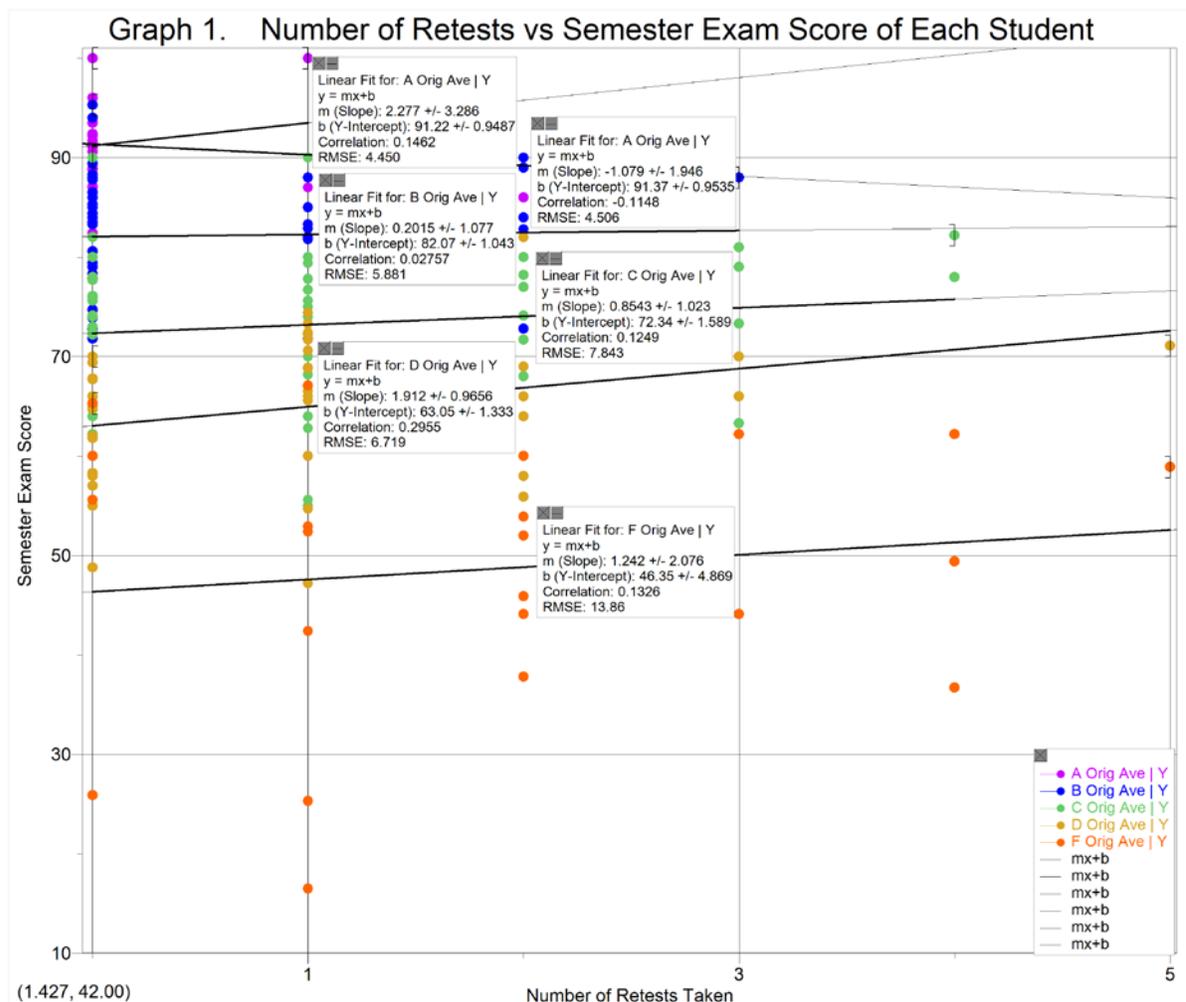
A total of 184 student-semesters are represented in the data collected. Of these, 92 students did not use the retest option during the semester, 50 students used it one time during the semester, and 42 students used the retest option two to five times during the semester.

Table 1 indicates the number of retests taken and the average semester exam score for students within each grade-based ability group. For example, within the 'C' ability group (students who had averaged a 70-79.9% on their original tests), there were two students who used 4 retests during the semester, and they averaged 80.1% on the semester exam.

Table 1. Number of Retests Taken and Semester Exam Average within Ability Groups

Original Test Average Grade	Sample size (# of students)	Retests taken per semester	Semester exam average
A	1	2	86.0
	2	1	93.5
	22	0	91.2
B	1	3	88.0
	6	2	82.6
	12	1	80.8
	29	0	82.5
C	2	4	80.1
	4	3	74.2
	7	2	73.6
	15	1	71.9
	18	0	73.3
D	1	5	71.1
	2	3	68.0
	6	2	65.8
	15	1	66.6
	19	0	62.2
F	1	5	58.9
	3	4	49.4
	2	3	53.2
	6	2	48.9
	6	1	42.7
	4	0	51.7

Graph 1 includes the data for each individual student, and shows the linear fit of the data within each grade-based ability group. With the omission of one outlier in the 'A' ability group, each group shows an increase in semester exam scores as the number of retests increases, although other than the 'D' ability group, the uncertainty of the slope for each fit line is greater than the effect size. As a result, the data sets have correlation coefficients ranging from 0.0276 for the 'B' ability group to 0.296 for the 'D' ability group. For the 'A' ability group, there are two fit lines, one including the outlier of a single student who took two retests during the semester and performed poorly on the semester exam, and one omitting this same data point.

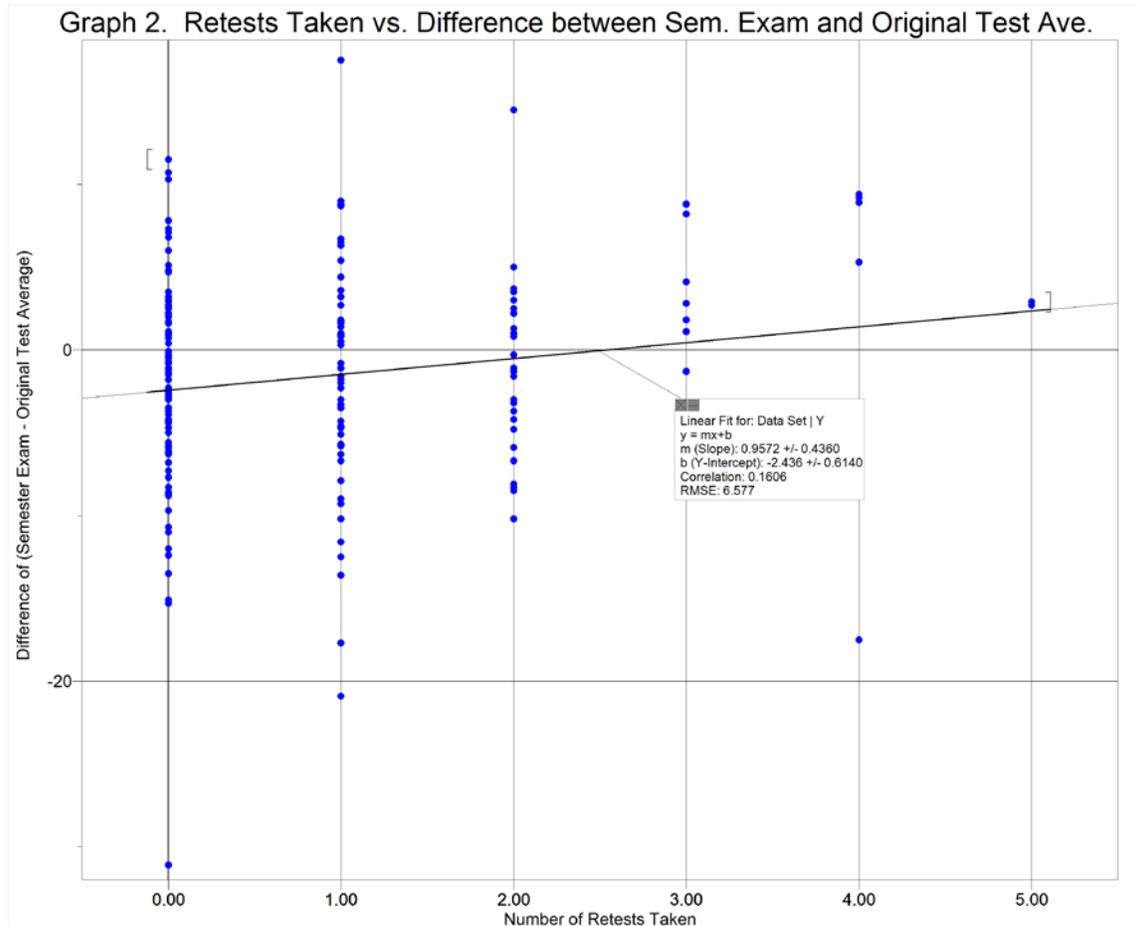


A comparison between the number of retests taken and the difference between the student's semester exam score and average score on the original tests is given in Table 2 and Graph 2. Table 2 groups the data by the number of retests each student took, and indicates that on average, the 16 students who used the retest option 3 or more times during a semester increased their semester exam grade above the average score they had received on their original unit tests. In contrast, on average, the students who had used the retest option 2 or fewer times during the semester scored lower on the semester exam than their original test average.

Table 2. Number of Retests Taken and the Difference between Sem. Exam Scores and Students' Average on Original Unit Tests

# of retests taken during a semester	average difference of (semester exam score – average of original tests)	Sample size (# of students)
5	2.8	2
4	3.1	5
3	1.3	9
2	-1.6	26
1	-1.6	50
0	-2.3	92

Graph 2 provides the data points for all 184 students, and examines the relationship between these two variables on the basis of individual students instead of group averages. The linear fit for the Number of Retests Taken vs. Difference between Semester Exam and Original Test Average shows an increase of 0.96% on semester exam scores for each retest taken during a semester, with an uncertainty of  $\pm 0.44\%$  and a correlation coefficient of 0.16.



## Analysis

This study shows a small increase in students' long-term retention, as evidenced by semester exam scores, as the number of retests taken during the semester increases. The analysis of Graph 1 shows this increase for all ability groups (with the omission of one data point in the 'A' ability group), with effect size ranging from a 0.2% increase in exam scores per retest for the 'B' ability group, to a 1.9% increase in the 'D' ability group's exam scores for each retest taken. Although the uncertainty is greater than the effect size for most ability groups, the positive trend across all ability groups indicates that retesting seems to provide some performance gains to most students. In general, the higher ability groups experienced fewer gains than lower ability groups.

This difference in effect size between ability groups may be due to several factors. One factor is that the higher ability groups simply do not have as many concepts to relearn and improve on, while lower ability groups have more room for improvement. Given that the greatest gain was in the 'D' ability range, it may also be that this group of students felt the greatest motivation to relearn the material to ensure that they received a passing grade for the course. It is also possible that the improvement seen in lower ability groups is partly caused by improvement in test-taking strategies, which the higher ability groups may have already possessed.

Along this line, it is interesting to note that the 'F' ability group did not make greater gains than the 'D' ability group. It is likely that part of this effect stems from a difference in student attitudes, where some students in the 'F' group adopted the approach of, "I already failed the first test, I have nothing to lose by retaking it, but why bother studying for it?" Lower motivation to learn the material is certainly part of the reason some students are in this group, which may help explain the smaller gains by the group during this study.

In the second comparison, relating the number of retests taken and the difference between the students' semester exam and original unit test average, there is also a noticeable trend supporting the hypothesis that optional retesting aids long-term retention of the material. As indicated in Table 2, students who did not use the retake option during the semester scored 2.3% lower on their semester exams compared to their original test average. Students taking one or two retests also scored lower on the semester exam than their original tests. However, students who used three or more of the retests scored higher on their semester exams than on their original tests. Students taking

four or five retests during the semester scored approximately 3% higher on their semester exam than they had on their original unit tests.

Although this study did not collect specific data pertaining to student attitudes toward the retesting option, anecdotal feedback from students throughout the course was quite positive. Similar to the studies by Davidson (1984), Friedman (1987), Catanzano & Wilson (1977), and Juhler, *et al.* (1998), the students in this study offered feedback reporting less test anxiety because of the retesting option. This lower anxiety was reported on the original test because of the ability to ‘try again’ if the first test was unsatisfactory, and also reported on the retests because the students felt they knew more of what to expect on the retest, having already taken a similar test. While this supports the finding by Friedman (1987) and Catanzano & Wilson (1977) that scores on the retests themselves are higher due to practice effects, it does not seem sufficient to account for the increased performance on semester exams as found in this study.

The findings in this study support the practice of optional retesting as a method for helping students learn and retain material. Although the effect is small, the potential gains for lower ability students make the practice worthwhile. There is some additional work for the instructor to develop the retests at the onset of the retesting program, but the opportunity provided to the students definitely warrants the time invested.

This study would benefit from a larger sample size, especially in the number of students taking more than two retests during a semester. If the correlation between retesting and semester exam scores is upheld by more extensive studies, it would be interesting to study whether the effect is also present in subject areas with less sequential curricula.

Another suggestion for an extension of this study would be to determine if students should be required to complete additional work as remediation or reinforcement before taking a retest. There are many aspects of this question that could be explored, including the effect on long-term retention if students are required to complete additional work on the entire unit, only the concepts that they struggled with the most, or no additional work. It would also be interesting to find how many students decline to take optional retests if additional work is required.

### **Summary**

This study examines the effect of an optional retesting strategy on long-term retention of course content. It does so by comparing the number of retests a student chooses to take during the semester to the student's performance on the semester exam. The rationale for this instructional practice is that when a student takes the time to restudy material for a retest, the mastery level and long-term retention of that content should increase, which would lead to better performance on cumulative assessments such as the semester exam. This study does find a small trend toward higher performance on semester exams for students who choose to take more retests throughout the semester, though the effect size is not statistically significant.

## Works Cited

- Abraham, Paul. "A Microscopic Look at Assessment: Dropping a Lowest Test Score versus Allowing a Retake Test." *AURCO Mathematics Journal* 1 (2000): 44-51. Print.
- Bangert-Drowns, Robert L., James A. Kulik, and Chen-Lin C. Kulik. "Effects of Frequent Classroom Testing." *Journal of Educational Research* 85.2 (1991): 89-99. *MasterFILE Premier*. Web. 23 Aug. 2011.  
<<http://web.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=3&hid=110&sid=6e09de59-a56a-4d23-b9fb-92409e9d5072%40sessionmgr113>>.
- Bloom, Benjamin S. "New Views of the Learner: Implications for Instruction and Curriculum." *Educational Leadership* 35.7 (1978): 563-568, 570-576. *ASCD*. Web. 23 Aug. 2012.  
[http://www.ascd.org/ASCD/pdf/journals/ed\\_lead/el\\_197804\\_bloom.pdf](http://www.ascd.org/ASCD/pdf/journals/ed_lead/el_197804_bloom.pdf)
- Catanzano, Robert, and Melanie S. Wilson. "The Effect of Retesting Contingencies on Achievement, Anxiety, and Attitude in Seventh Grade Science." *Science Education* 61.2 (1977): 173-80. Print.
- Davidson, William B., William J. House, and Thomas L. Boyd. "A Test-Retest Policy for Introductory Psychology Courses." *Teaching of Psychology* 11.3 (1984): 182-84. *Academic Search Premier*. Web. 23 Aug. 2011.  
<<http://search.ebscohost.com/login.aspx?direct=true&AuthType=cookie,ip,cpid&custid=ns217454&db=aph&AN=18596751&site=ehost-live&scope=site>>.

- Diegelman-Parente, Amy. "The Use of Mastery Learning With Competency-Based Grading in an Organic Chemistry Course." *Journal of College Science Teaching* 40.5 (2011): 50-58. *Education Research Complete*. Web. 24 Aug. 2011. <<http://search.ebscohost.com/login>>.
- Friedman, Herbert. "Repeat Examinations in Introductory Statistics Courses." *Teaching of Psychology* 14.1 (1987): 20-23. *Academic Search Premier*. Web. 23 Aug. 2011. <<http://search.ebscohost.com/login.aspx?direct=true&AuthType=cookie,ip,cpid&custid=ns217454&db=aph&AN=6381456&site=ehost-live&scope=site>>.
- Geving, Allison M., Shannon Webb, and Bruce Davis. "Opportunities for Repeat Testing: Practice Doesn't Always Make Perfect." *Applied H.R.M. Research* 10.2 (2005): 47-56. Print.
- Juhler, Sandra M., Janice F. Rech, Steven G. From, and Monica M. Brogan. "The Effect of Optional Retesting on College Students' Achievement in an Individualized Algebra Course." *The Journal of Experimental Education* 66.2 (1998): 125-37. *MasterFILE Premier*. Web. 21 Sept. 2010. <<http://web.ebscohost.com/ehost/detail?vid=4&hid=110&sid=6e09de59-a56a-4d23-b9fb-92409e9d5072%40sessionmgr113&bdata=JkF1dGhUeXBIPWNvb2tpZSxpcCxjcGlkImN1c3RpZD1zNzMyNDk2NCZzaXRIPWVob3N0LWxpdmUm c2NvcGU9c2l0ZQ%3d%3d#db=f5h&AN=496441>>.