Quizzing and retention in the high school science class

Donell DeBacker Evans
Louisiana State University and Agricultural and Mechanical College, donellevans@gmail.com

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_theses

Recommended Citation
Evans, Donell DeBacker, "Quizzing and retention in the high school science class" (2013). LSU Master's Theses. 3780.
https://digitalcommons.lsu.edu/gradschool_theses/3780

This Thesis is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Master's Theses by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
A Thesis

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Master of Natural Sciences

in

The Interdepartmental Program in Natural Science

by

Donell DeBacker Evans
B.S., Northwestern State University, 1999
M.S., University of Texas, 2013
August 2013
I dedicate this paper to my mother and father who made me believe there is nothing that I cannot achieve and who supported my education. To my high school teachers Mrs. Hale, Mrs. Servello, and my high school Principal Mr. Schwartz for their support of my education and future. To Mr. Avant an amazing mentor teacher. To God who has kept me strong and unwavering in the winds of change. To my family who stood by me through all the endless hours of absence.
ACKNOWLEDGEMENTS

Thanks to Mrs. Whitlock and the Vernon Parish school board for their support of teachers and the advancement of science. Thanks to Dr. Joseph Siebenaller for his guidance, time, humor, support of teachers, and support in completing this thesis. Thanks to Dr. Gregg, Dr. Wischusen, and Dr. Watkins for their advice and time in support of the advancement of teachers. Thanks to Dr. Madden, Dr. Neubrander, Dr. Hopkins, Dr. Browne, and Mrs. Blanchard for their support of education and teacher advancement. A big thanks to Christine Mayeux for her incredible kindness.
# TABLE OF CONTENTS

**ACKNOWLEDGEMENTS** .................................................................................................................. iii

**LIST OF TABLES** .............................................................................................................................. v

**LIST OF FIGURES** ............................................................................................................................. vi

**ABSTRACT** .......................................................................................................................................... vii

**CHAPTER**

**INTRODUCTION** ............................................................................................................................... 1
  Rationale for the Study ......................................................................................................................... 1
  Utilizing Testing as a Tool ..................................................................................................................... 1
  Memory Mechanisms ............................................................................................................................ 3

**MATERIALS AND METHODS** ......................................................................................................... 5
  Designing the Pre- and Post-test ........................................................................................................... 5
  Administering the Tests ......................................................................................................................... 10
  Definition of the Study Population ...................................................................................................... 11
  Calculating Mean scores and Learning gains ...................................................................................... 12

**RESULTS** .......................................................................................................................................... 13
  Analysis of Pre- and Post-Tests .......................................................................................................... 13

**DISCUSSION** ..................................................................................................................................... 23

**CONCLUSION** ................................................................................................................................... 26

**REFERENCES** .................................................................................................................................... 31

**APPENDIX**

  **A  TEST QUESTION EXAMPLES** .................................................................................................. 33

  **B  PARENTAL PERMISSION FORMS** ............................................................................................. 34

  **C  STUDENT CONSENT FORM** ..................................................................................................... 36

  **D  IRB FORMS** ................................................................................................................................. 37

**VITA** ................................................................................................................................................... 38
LIST OF TABLES

1. Experimental Design........................................................................................................................................5
2. Intervening Quizzes ........................................................................................................................................10
3. Study Population Ethnicity ..............................................................................................................................11
4. Chemistry and Biology Chapter 1, 2, and 3 Sample Size Pre-/ Post Test.............................................. 11
**LIST OF FIGURES**

1. Chemistry Chapter 1 percentage of question type ................................................. 6
2. Chemistry Chapter 2 percentage of question type .................................................. 7
3. Chemistry Chapter 3 percentage of question type .................................................. 7
4. Biology Chapter 1 percentage of question type ...................................................... 8
5. Biology Chapter 2 percentage of question type ...................................................... 9
6. Biology Chapter 9 percentage of question type .................................................... 9
7. Chemistry chapter 1 pre-/post-test data analysis .................................................... 13
8. Chemistry chapter 1 post-test learning gains ......................................................... 14
9. Chemistry chapter 2 pre-/post-test data analysis .................................................... 15
10. Chemistry chapter 2 post-test learning gains ....................................................... 16
11. Chemistry chapter 3 pre-/post-test data analysis .................................................. 17
12. Chemistry chapter 3 post-test learning gains ....................................................... 18
13. Chemistry chapter 1, 2, and 3 pre-/post-test data analysis ..................................... 19
14. Chemistry chapter 1, 2, and 3 pooled learning gains .......................................... 20
15. Biology chapter 1, 2, and 9 pre-/post-test data analysis ....................................... 21
16. Biology chapter 1 2, and 9 learning gains............................................................ 22
ABSTRACT

There is a need to identify educational tools and methods that are easily assimilated into a secondary science education classroom. The use of testing as an educational tool, rather than as a summative assessment only, has emerged as a possible solution. Test-enhanced learning is attributed to the testing effect (Roediger & Karpicke, 2006). The testing effect refers to the higher probability of recalling an item resulting from the act of retrieving the item from memory (testing) versus additional study trials of the item (Crane et al., 2008). A comparison of frequent testing/quizzing versus no quizzing outcomes were studied to determine whether the testing effect has a positive influence on learning gains in a high school science setting. Eighty-eight juniors and seniors enrolled in chemistry (4 sections) and advance placement biology (1 section) classes during the 2012-2013 school year were studied. A within student experimental design was used. Students were pre-tested prior to content coverage. Upon completion of each topic section the students were given quizzes with feedback. The quizzes targeted 50% of the pre-test/post-test material but were not identical in wording. Learning gains for the quizzed material were equivalent to the learning gains for the non-quizzed material in 66% of the instances tested. In 44% of the individual classes the learning gains were greater for the quizzed condition. For these pooled data, quizzing resulted in greater learning gains in all chapters. For two of the three biology chapters quizzing resulted in greater learning gains. These results indicate that using tests and quizzes as an educational tool can have a positive impact on student learning gains.
INTRODUCTION

Rationale for the Study

Improving science education both in high schools and colleges is a national priority that if left unaddressed could have dire consequences for the future of the United States (Members of the 2005 “Rising Above the Gathering Storm” Committee, 2011). Because of the lack of progress in the secondary science classroom the United States is losing its ability to fill STEM careers with United States citizens (Stine & Matthews, 2009). The number of native-born Americans who are entering or considering the STEM careers is decreasing, and the programs for fostering innovation in the United States now depend on foreign–born talent (STEM Workforce Challenge, 2007).

Calls have been made to improve teacher preparedness and to introduce pedagogical tools into the class room that improve student engagement and learning (Stine & Matthews, 2009). There are numerous methods being introduced to increase student learning, for example, grouping (collaboration), peer to peer review, and concept mapping, also called graphic organizer. However, many of the methods do not consider how the assimilation of these methods would fit into a teacher’s already existing repertoire.

Utilizing Testing as a Tool

The use of quizzing to increase retention of targeted material has drawn more attention over the last decade. McDaniel et al. (2007) refer to the increase in memory gain by the use of intervening test as the “testing effect”. The common practice in most classrooms is however, to use testing simply as an evaluative tool to measure acquisition of targeted material. Studies have shown that frequent testing can serve as more than a summative assessment (Rohrer, Taylor, & Sholar, 2010). Integrating testing as a pedagogical tool to assist students in retention of targeted material can be an effective and easily implemented tool in an educator’s arsenal. In this respect quizzing serves two purposes, as a measurement tool and as a teaching enhancement tool.
The testing effect is the idea that active processing and retrieval of knowledge through testing enhances the learning process. Testing becomes a method to reconnect with the targeted material. As far back as 1927, memory experiments show the positive impact that reconnecting with targeted material has on retention of that material. Starch (1927) wrote “Since the rate of forgetting is very rapid at first . . . it probably would be highly advantageous to have relearning . . . come very frequently at first and more rarely later on” (as cited in Karpicke et al., 2007, p. 705). More recent studies directed at the idea of testing as an educational tool have shown the positive effects of testing on learning targeted material. Roediger and Karpicke (2006) find that basic memory experiments have shown that on a final criteria test, students better remember information on which they had been tested sometime prior to the test than information on which they had not been tested previously. A review of eight experimental reports by McDaniel et al. (2007) consistently demonstrated the positive effect that tests have on targeted material.

When a teacher uses frequent testing (quizzing) as an educational tool rather than simply as a summative assessment it can benefit students through feedback loops, student metacognition, and learning gains. The testing effect can conceivably benefit multiple subjects. Testing with feedback can be used as a form of Socratic teaching that when used properly provides a formative assessment making tests and quizzes feedback tools between the teacher and students. Because, educational settings have increased the number of students in classes, this does not allow for frequent feedback or conversation between the student and teacher which is a foundational component of Socratic teaching method (Caldwell, 2007). However, quizzes and tests in a larger classroom still give the opportunity for feedback. Quizzing prior to delivery of new content material allows the students to focus on the key ideas expected during the knowledge acquisition. Kornell and Son (2009) and Thomas and McDaniel, 2007 suggest that quizzing could also have indirect effects, such as to improve students’ metacognitive judgments about what they know and do not know, thereby increasing study effectiveness (McDaniel et al., 2011, p. 400). Wood (2004) finds that quizzing can be a tool to reveal student misunderstandings. Many consider it best for students to discuss their errors or wrong answers and the act of recognizing errors and processing correction for the errors strengthens
Retention. Roediger and Karpicke (2006a) state that, even when feedback is not given the testing effect still may increase retention of the target material. The testing effect can also indirectly benefit students by reducing anxiety about the target material (McDaniel et al., 2011).

Memory Mechanisms

The mechanism behind the testing effect is attributed to the act of retrieval. By experiencing (retrieving) the material after an encoding episode a student increases the retention time. Chan and McDermott (2007) found that if a test is taken soon after the encoding process, in most cases, it improves the recall on a subsequent delayed test. In terms of psychology and pedagogical practices, testing initiates the retrieval process. So, the act of retrieving the target material for the test or quiz improves the learning process. The act of recalling or retrieving improves long term memory through a process of consolidation. Memory re-consolidation is when previously consolidated memories are recalled and then actively consolidated all over again, in order to maintain, strengthen and modify memories that are already stored in the long-term memory (Mastin n.d.). Studies find that taking a test over material can improve long-term retention relative to repeatedly studying the material (Karpicke & Roediger, 2007). The effects of testing are consistent among studies with different materials and different designs. McDaniel et al. (2007), discuss an amplification to the testing effect using recall instead of recognition. They note that recall promotes retrieval processing that is more potent to retention than mere recognition. While current data seem to confirm a positive correlation between testing and knowledge acquisition, its use as a measurement device seems to hide testing’s other valuable role. Bjork (1975) states that the use of testing as an evaluative tool loses sight of the valuable role tests play as a powerful memory enhancer.

The testing effect has largely been demonstrated in laboratory and college settings. The question addressed in the present study is whether active processing and retrieval of knowledge through testing (quizzing) increases student ability to learn in a high school science class setting. Testing can become an even more valuable teaching tool if, when properly implemented, the results amplify retention. McDaniel et al. (2007) note that there is presumably much more variability across
students in an actual classroom setting than in a controlled laboratory study. Thus, examining the testing effect in a classroom may provide a more realistic evaluation of the efficacy of a “quizzing” approach to teaching the material. To address the magnitude of the testing effect in the high school science classroom, several biology and chemistry units were taught with some content quizzed and other content not quizzed, but all content taught in a modified conventional lecture fashion. The testing effect was evaluated by comparing the student’s performance on pre- and post-tests on quizzed and non-quizzed material.
MATERIALS AND METHODS

Designing the Pre- and Post-test

The effect of testing was studied on chemistry and biology 11th and 12th grade science classes. This study used a within-student design modeled after McDaniel et al. (2011). For each chapter all students received a multiple choice pre- and post-test. Fifty percent of the pre- and post-test content was targeted with intervening quizzes between pre-test and post-test. The students received feedback in the form of a question and answer session directly after quizzing. During the feedback sessions misconceptions, errors, and problems with calculations were addressed. The other fifty percent of the questions on the pre-/post-test were not quizzed following the pre-test.

Table 1. Experimental Design

<table>
<thead>
<tr>
<th>Activities</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>Experimental</td>
</tr>
<tr>
<td>Pre-test</td>
<td>Post-test</td>
</tr>
<tr>
<td>Variable Lectures</td>
<td>Variable Lectures</td>
</tr>
<tr>
<td>Homework</td>
<td>Homework</td>
</tr>
<tr>
<td></td>
<td>Quizzed (50 % pre-/post-test) content</td>
</tr>
<tr>
<td>Post-test</td>
<td>Post-test</td>
</tr>
</tbody>
</table>

The quizzed questions and the correlating questions on the pre/post-test were grouped by their similarity. However, the questions were not identical to eliminate the possibility of recognition rather than recall. For an example of the quizzed questions see Appendix A. When choosing the non-quizzed questions which relate to the control material, careful consideration was taken to limit the similarity between the non-quizzed questions and quizzed questions. This was to try to reduce the associated memory effect. The associated memory effect is the idea that activation of one concept in memory should produce facilitative effects for related concepts (Chan et al., 2006). For example, if
two questions were being considered as a tested and quizzed question set and both questions discussed covalent bond characteristics but in different aspects they could be used as a tested and quizzed pair for the experimental questions. If a quizzed question addressed covalent bond characteristics a question for the control material could not address covalent bond characteristics. Educational ancillary material generally has similarity to questions found within the workbooks, PowerPoint presentations, and other study materials for a specific textbook. To prevent students from being exposed to questions similar to those on the pre-test and post-test homework, etc., questions were taken from multiple sources. To calculate base knowledge of the students at the beginning of the chapter a pre-test was administered prior to delivery of the targeted content. To determine the amount of knowledge retention an identical post-test was administered upon the completion of the chapter. Intervening quizzes were administered at the completion of chapter sections.

The test and quiz questions for each chemistry chapter were created using Exam View Pro Test generator version 6.2.1 for Pearson Chemistry Louisiana. The classification for the test questions for chemistry differed from biology in that they were classified by difficulty level (I, II, and III). Level III was the highest difficulty level (Figure 1). To determine the difficulty level of test items, a measure called the Difficulty Index is used. It measures the proportions of students who answer the test item accurately. A lower proportion results in a more difficult question. The chapter 1 topic was a basic overview of the scientific method, branches of chemistry, and chemical versus physical changes. The pre- and post-test for this chapter consisted of 30 multiple choice questions and varied in question types (Figure 1).

![Chemistry Chapter 1](image)

**Figure 1:** Chemistry chapter 1 percentage of question type for pre- and post-test control and experimental and the intervening quizzes. Question categorized by difficulty index with level III being the most difficult.
The chapter 2 topic was basic chemistry math that included significant figures, rounding rules, scientific notation, and simple conversion. The pre- and post-test for this chapter consisted of 30 multiple choice questions and varied in question types (Figure 2).

![Chemistry Chapter 2](image1.png)

Figure 2: Chemistry chapter 2 percentage of question type for pre- and post-test control and experimental and the intervening quizzes. Question categorized by difficulty index with level III being the most difficult.

The chapter three topic was chemical reactions and it consisted of balancing equations, predicting products, and reaction types. The chapter 3 pre-test and post-design was altered from the other test by an additional answer choice for each multiple choice question. The additional answer choice was “I do not know”. The pre- and post-test for this chapter consisted of 30 multiple choice questions with different question types (Figure 3).

![Chemistry Chapter 3](image2.png)

Figure 3: Chemistry chapter 3 percentage of question type for pre- and post-test control and experimental and the intervening quizzes. Question categorized by difficulty index with level III being the most difficult.
All questions used in the pre/post-test and quizzes for each Biology chapter were created using test bank questions from Mader Biology 10th Edition, Campbell Biology 6th Edition, and Campbell Biology 8th Edition. The pre and post-test for biology chapter 1 (View of Biology) consisted of 30 multiple choice questions and varied in question types (Figure 4). The topic for Chapter 1 was the study of biology. It included a general overview of biology themes, processes of science, and biology in everyday life. The questions were classified base on Bloom’s domain. The categories used by the book publisher were knowledge/ comprehension, factual recall, application/analysis, and synthesis. Application/analysis and synthesis are considered higher level categories of Bloom’s domain and would be considered more difficult questions.

Figure 4: Biology chapter 1 percentage of question type for pre- and post-test control and experimental and the intervening quizzes. Question categorized by Bloom’s domain.

The pre and post-test for biology chapter 2 consisted of 34 multiple-choice questions and varied in question type (Figure 5). The topic for this chapter was basic chemistry that included atoms, molecules, elements connections, molecular arrangements, and molecular interactions.
The pre- and post-test for the biology unit 9 (Genetics) consisted of 30 multiple choice questions and varied in question types (Figure 6). The topic for this chapter was basic genetics and included Mendel’s laws, Punnettes square calculations, chromosomal basis and behavior, and linkage.

Figure 5: Biology chapter 2 percentage of question type for pre- and post-test control and experimental and the intervening quizzes. Question categorized by Blooms domain.

Figure 6: Biology chapter 9 percentage of question type for pre- and post-test control and experimental and the intervening quizzes. Question categorized by Blooms domain.
Administering the Tests

All students and parents agreed to participate in the study by signed consent forms (Appendix B and C). The consent forms and the project were approved by the Institutional Review Board at Louisiana State University (IRB #E5995).

Before the start of each chapter a pre-test was administered to the students. The test was not recorded as a class grade. The tests were analyzed and filed with no feedback. The content material was covered by sections and the students were given a quiz at the completion of the section. The quizzes were counted as graded class material. After the entire class completed the quiz the material was graded and feedback was given about any errors and misconceptions. The quizzes were returned to the teacher for filing. Upon completion of the chapter (normally a Thursday) a post–test was administered the following Monday after a review session on Friday. The amount and day of the review session varied between chapters. The review sessions covered all content. The chapter post-tests were counted a class grade.

Table 2. Intervening quizzes for biology and chemistry. Number of intervening quizzes, average time between each quiz, and the time between pre- and post-test.

<table>
<thead>
<tr>
<th>Intervening Quizzes</th>
<th>Chemistry</th>
<th>Biology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Number of Quizzes</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Average Time Between Quizzes</td>
<td>1-2 weeks</td>
<td>Several days</td>
</tr>
<tr>
<td>Time Between Pre-/Post-test</td>
<td>4 weeks</td>
<td>3 weeks</td>
</tr>
</tbody>
</table>

The teaching strategies use included: PowerPoint, Whiteboarding, group projects, group and individual problem solving, class discussion, homework, labs, and frequent questioning with feedback. The percentages of each varied depending on the topic however, questioning and feedback was consistent and extensive through all chapters.
Definition of the Study Population

Students from 4 chemistry classes labeled A, C, D, and E and students from one AP biology section participated in the study. The number of participants who were involved in the pre-test and post-test varied from chapter to chapter for various reasons including absenteeism, student moving, and class changes (Table 4). Because of the in-student experimental design, students who did not take all of the pre-test and post-test were excluded from the analysis. The grade level of the students participating ranged from 10th grade to 12th grade. The students who participated were representative of the population of the school as a whole (Table 3). The school population consists of 800 students in 2012. Of those students, 44.5% receive free and reduced lunch, 54.0% percent minority, students with disabilities 6.4%, and 72% percent at achievement level.

Table 3. Number of students in the study group compared to the population of the school (eSchoolPlus+) Ethnicity of students in Leesville High School and the study population.

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>School Population</th>
<th>Study Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>323</td>
<td>39</td>
</tr>
<tr>
<td>White</td>
<td>373</td>
<td>40</td>
</tr>
<tr>
<td>Asian</td>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>Hispanic</td>
<td>64</td>
<td>7</td>
</tr>
<tr>
<td>Other</td>
<td>20</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4. Number of students who participated in the pre- and post-chapter test for chemistry (section A, C, D, and E) chapter 1, 2, and 3 and biology (section B) chapter 1, 2, and 9. Students were excluded from the data if they missed the pre- or post-test.

<table>
<thead>
<tr>
<th>Chemistry Class</th>
<th>Ch. 1 Pre-Test N=</th>
<th>Ch. 1 Post-Test N=</th>
<th>Ch. 2 Pre-Test N=</th>
<th>Ch. 2 Post-Test N=</th>
<th>Ch. 3 Pre-Test N=</th>
<th>Ch. 3 Post-Test N=</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>18, 16</td>
<td>18, 15</td>
<td>18, 16</td>
<td>18, 12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>17, 10</td>
<td>18, 16</td>
<td>18, 16</td>
<td>18, 16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>16, 11</td>
<td>21, 21</td>
<td>14, 12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>12, 9</td>
<td>12, 11</td>
<td>11, 7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biology Class</td>
<td>Ch. 1</td>
<td>Ch. 2</td>
<td>Ch. 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>18, 16</td>
<td>13, 13</td>
<td>14, 11</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Calculating Mean scores and Learning gains

Because of the within student design of this study only students who completed all pre- and post-tests were included in the analysis. Pre- and post-tests were scored to determine the difference in the number of questions correct between the control questions and the experimental questions. Means, standard errors, and statistical test were calculated using Graph Pad Instat version 3.10 software program. Data analysis using the Kolmogorov-Smirnov test determined that the data was not normally distributed. For this reason the comparison between pre-/post-test control versus experimental a non-parametric analog of the repeated measures ANOVA (Friedman test) was used to calculate p-values. Graphs were created using Graph Pad Prism version 6.02 using values that had been, calculated with Instat. Learning gains were calculated for experimental and control questions from the pre-test and post-test using Hake’s (1998) formula. Learning gain = (post-test scores % - pre-test score %) ÷ (100 – pre-test score %). Any negative learning gain values were recorded as zeros before calculation of mean and standard error. Learning gain means were non-Gaussian and required paring so the Wilcoxon matched-pairs signed-ranks test were used to calculate differences in experimental versus control learning gains.
RESULTS

Analysis of Pre- and Post-Tests

The pre-test scores for the control and experimental material did not differ statistically in any of the classes (P > 0.05) (Figure 7). Classes A and C showed significant differences in post-test scores between the controls and experimental teaching methods (class A P < 0.01 and class C P < 0.05) with the experimental group having higher gains. Class D and E did not differ in the post-test scores between the control and experimental teaching methods (P > 0.05).

Chemistry Ch. 1 Science Processes

Figure 7: Pre and Post-test data for 4 classes on the chemistry chapter 1. Class A N = 16, class C N =10, class D N =11, and class E N = 9. Mean and standard errors are shown. Some error values are smaller than the symbol. The asterisks indicate values that differed significantly. * significance at P < 0.05; ** significance at P < 0.01
Learning gains were calculated from pre-test to post-test for the chapter 1 control questions and the experimental questions (Figure 8). There were significant differences in the learning gains between the control and experimental teaching methods for class A ($P = 0.0015$) and class C ($P = 0.0020$) with the experimental group had significantly higher learning gains. No differences in learning gains were found between class D and E control and experimental teaching methods ($P > 0.05$).

Figure 8: Leaning Gains for Control and Experimental data for 4 classes on the chemistry chapter 1. Class A $N = 16$, class C $N = 10$, class D $N = 11$, and class E $N = 9$. Mean and standard errors are shown. The asterisks indicate values that differed. Learning gains were calculated using the Wilcoxon matched-pairs signed-ranks. ** significance at $P < 0.01$
Pre- and post-test scores were compared for all classes for chemistry chapter 2 (Figure 9). The pre-test scores for the control and experimental materials did not differ statistically in the classes. No differences between control and experimental teaching methods outcomes were detected in any of the individual classes (P > 0.05).

Chemistry Ch. 2 Quantitative Measurements

Figure 9: Pre and post-test data for 4 classes on the chemistry chapter 2. Class A N = 15, class C N = 16, class D N = 21, and class E N = 11. Mean and standard error are shown. Some error values are smaller than the symbol. The asterisks indicate values that differed.
Learning gains were calculated from pre- to post-test scores for the chapter 2 control and experimental questions (Figure 10). Learning gains were significantly different between control and quizzed topics for class D (P = 0.0256). No differences in learning gains were found between control and experimental teaching methods for classes A, C, and E (P > 0.05).

Figure 10: Learning gains for control and experimental data for 4 classes on the chemistry chapter 2. Class A N = 15, class C N = 16, class D N = 21, and class E N = 11. Mean and standard errors are shown. The asterisks indicate values that differed. Learning gains were calculated using the Wilcoxon matched-pairs signed-ranks. * significance at P < 0.05
Pre- and post-test scores were compared for all classes for chemistry chapter 3 (Figure 11). The pre-test scores for the control and quizzed materials did not differ in the classes. No differences between control and experimental teaching methods outcomes were detected in any of the individual classes.

**Chemistry Ch. 3 Chemical Reactions**

![Chemistry Ch. 3 Chemical Reactions Graphs](image)

Figure 11: Pre and Post-test data for 4 classes on the chemistry chapter 3. Class A N = 12, class C N = 16, class D N = 12, and class E N = 7. Mean and standard error are shown. Some error values are smaller than the symbol.
Learning gains were calculated from pre-test to post-test for the chapter 3 control questions and the experimental questions (Figure 12). Learning gains were not significantly different between the control and experimental teaching methods for class A (P = 0.0640) but differences were found between the control and experimental teaching methods for class C (P = 0.0134) and class E (P = 0.0313). No differences in learning gains were found between control and experimental teaching methods in class D (P > 0.05).

Figure 12: Learning Gains for control and experimental data for 4 classes on the chemistry chapter 3. Class A N = 12, class C N = 16, class D N = 12, and class E N = 7. Mean and standard errors are shown. The asterisks indicate values that differed. Learning gains were calculated using the Wilcoxon matched-pairs signed-ranks. * significance at P < 0.05
The number of students per class hour ranged from 12 students to 21 students. Because the classes were taught by the same teacher covering the same contents the data were pooled by chapter. The pooled pre-test scores for chapter 1, 2, and 3 did not differ between control and experimental $P > 0.05$ (Figure 13). The pooled post-test chapter data differed between the control and experimental teaching methods (chapter 1 $P < 0.001$, chapter 2 $P < 0.001$, and chapter 3 $P < 0.05$).

**Chemistry Ch. 1, 2, and 3**

![Graphs](image)

Figure 13: Pre and Post-test pooled scores for chemistry chapters 1, 2, and 3 for 4 classes. Chapter 1 $N = 46$, Chapter 2 $N = 63$, Chapter 3 $N = 48$. Mean and standard errors are shown. Some error values are smaller than the symbol. The asterisks indicate values that differed. * significance at $P < 0.05$; *** significance at $P < 0.001$. 
The learning gains for control and experimental questions for all classes were pooled by chapter (Figure 14). In all three chapters learning gains were statistically different between the control and experimental teaching methods for chapter 1 $P = 0.0005$ and chapter 2 $P = 0.0019$, and chapter 3 $P = 0.0020$.

Figure 14: Learning Gains for control and experimental compiled chemistry class data for chapter 1, 2, and 3 for 4 classes. Chapter 1 $N = 46$, chapter 2 $N = 63$, chapter 3 $N = 48$. Mean and standard errors are shown. The asterisks indicate values that differed. Learning gains were calculated using the Wilcoxon matched-pairs signed-ranks. ** significance at $P < 0.01$, *** significance at $P < 0.001$. 
The biology course consisted of only one section. No differences were found between the pre-test control and experimental number of questions answered correctly for any individual chapters (Figure 15).

**Biology Chapters 1, 2, 9**

![Graph of Biology Chapters 1, 2, 9](image)

Figure 15: Pre and Post-test data biology chapters 1, 2, and 9 for one class. Chapter 1 N = 16, chapter 2 N = 13, and chapter 9 D N = 11. Mean and standard errors are shown.
Learning gains for the control and experimental teaching methods were calculated for each individual biology chapter (Figure 16). There were no differences in learning gains between the control and experimental teaching methods for chapter 1. There were differences in learning gains for chapters 2 and 9 between the control and experimental teaching methods (chapter 2 $P = 0.0034$ and chapter 9 $P = 0.0010$).

**Biology Learning Gains Chapter 1, 2, 9**

Figure 16: Learning Gains for control and experimental biology class data for chapter 1, 2, and, 9 for 1 class. Chapter 1 $N = 16$, chapter 2 $N = 13$ and, chapter 3 $N = 11$. Mean and standard errors are shown. The asterisks indicate values that differed. Learning gains were calculated using the Wilcoxon matched-pairs signed-ranks. ** significance at $P < 0.01$


DISCUSSION

Two of the 12 individual classes differed in the number of correct answer choices between the control and quizzed condition (Figures 7, 9, and 11). All of the pooled chapter data differed in the number of correct answer choices between the control and quizzed condition (Figure 13). Learning gains differed between the control and quizzed condition in 42% of the individual classes (Figure 8, 10, and 12). The pooled chapter learning gains differed between the control and experimental teaching method for all three chapters (Figure 14). Four chemistry classes labeled A, C, D, and E were studied.

Chemistry class sections A, C, D, and E learning outcomes were compared for chapter 1. For this unit two of four classes had differences in learning gains between the control and experimental teaching methods (Figure 8). The pooled results showed a difference in the control method and the experimental method in learning gains (Figure 14). The topic for this unit was basic chemistry which is taught in previous courses. Thus, the students had some prior knowledge of the concepts. It is possible that transfer appropriate processing had occurred that increased the memory performance for this chapter. Transfer appropriate processing is when memory performance increases because it matches processing required for subsequent learning (McDaniel et al., 2007). In class discussions for this topic, it was evident more so than other chapters that the students has some familiarity with the topic. Chapter 1 had the highest pre-test scores for both control and quizzed condition (Figure 7).

The only individual class that showed a difference between the control and experimental teaching method’s learning gains for unit 2 was class D (Figure 10). In many aspects this class of students varied considerably from the other groups of chemistry students. Based on my personal observation there was a larger amount of absenteeism, off task behavior, lack of concern for class content, and negative feelings that chemistry was merely a graduation requirement hurdle in comparison with classes A, C, and E. The focus for this unit was basic math of chemistry. This, in itself, poses challenges for students’ mastery of the topic, but class D, even with numerous distractors, displayed a difference in the outcomes from the control and experimental teaching method. Thus, frequent
quizzing may have been more of an impact for this class. Important differences remain between the classroom and laboratory facilities that could possibly reduce or eliminate the testing effect in a classroom setting (McDaniel et al., 2011). A number of factors could have affected the learning outcomes for this unit. Considering the small sample size for the individual classes the data was pooled for unit 2 and again the data indicated that increased quizzing can improve learning gains.

The choice of answers for chapter 3 varied from other chapters by including the choice “I do not know” to address the possible effects of guessing on learning gains. The number of correct answers chosen decreased for all pre-tests in chapter 3 compared to subsequent chapters (Figure 11). A larger number of the students chose “I do not know” on the pre-test versus the post-test. The decrease in the students correct answer choices could indicate that students took less risk when answering questions on the pre-test. McDaniel et al. (2011) discusses the impact of outside influences on the overall calculated quizzing effect when he states that “the present quizzing effect might slightly underestimate the true effect because quizzing augments student performance . . . on related content.” Because of the multifaceted dimensions of knowledge acquisition it is likely that other factors related to prior knowledge and content difficulty affected the results for chapter 3. Another study would need to be conducted to determine if there is a correlation between guessing and the overall learning gains. Even though variability was evident with class D and no differences the number of questions correct between the control and the experimental methods there were differences in learning gains between the control and experimental methods for class C, E, and the pooled chapter 3 data.

One AP biology / biology II class section was used for this study. Three chapters of biology were used for this research. The biology chapter 1 topic was an overview of biology that included scientific method, themes of biology, and biology in everyday life. The topics for biology chapter 2 were basic chemistry. The chapter 9 biology topic was basic genetics. The biology class was a mixture of Advance Placement biology and Biology II students. Both groups of students were taught in the same class at the same time because of scheduling issues. It is important to note that content coverage for both classes was the same however, the ability levels varied greatly among the students.
The class varied greatly in individual student academic grades, class interaction, and amount of homework completed.

Unlike the chemistry classes this class had only one section (Figure 15). The learning gains were greater for the experimental teaching method in chapters 2, and 9 (Figure 16). Learning gains are evident for chapter 1 but not statistically different between the control and experimental teaching method. This may be attributed to some of the students having been exposed to chapter 1 content during a summer Advance Placement prep program. The students who attended the preparatory program were a mixture of both AP students and Biology II students.
CONCLUSION

Improving science education has become a national priority because of the lack of progress in the secondary science classrooms. Lack of academic performance and knowledge gain in high school science classes has discouraged students from entering the STEM fields. This affects the United States’ position as a world leader but more harmful is its effect on the United States’ ability to stay an innovation leader. It is science innovation that saves lives, protects lives, and gives Americans a higher standard of living (Members of the 2005 "Rising Above the Gathering Storm" Committee, 2011). Identifying educational techniques that increase knowledge gain in the secondary science classroom and that are easily implemented will give educators tools that can improve learning gains in the secondary classroom. This can motivate students to enter STEM fields and help create the innovations that save lives and create a better future for all nations.

This necessity has inspired teachers to find effective strategies to increase student progress in science classes. Because testing and quizzing are common components of most educational facilities this study was performed to determine if testing/frequent quizzing can be used as an effective pedagogical instrument in a high school science setting. Evidence suggests that testing can increase student knowledge acquisition in a controlled laboratory setting and in the college classroom through retrieval and processing (McDaniel et al., 2007). These studies however, are not reflective of a high school science setting. This research was performed to determine if the testing effect would have the same potentiating effect in a high school classroom where there is variability in study habits and variability in the spacing of time between tests and quizzes.

The purpose for this study was to determine if the testing effect would be evident in a high school science classroom where there is a greater amount of variability than that of a laboratory setting. Many studies have shown the testing effect to improve student learning (McDaniel et al., 2007). The positive effects of testing on student learning could become a valuable tool for educators. A benefit to using testing as a teaching instrument would be its easy assimilation because of the minimal amount of change needed to incorporate this method.
Questions have been asked about the effect of high school classroom variation, increased length of time between pre-test and post-test, question types, and target content on the testing effect. Some possible insight to those questions may be evident with this study. Laboratory and college settings lend them themselves to a greater degree of homogeneity. College setting would have a greater degree of differences than those of a laboratory setting but more similarity than that of a high school setting. Several of the factors that create variability in the classroom should reduce the potentiating effect of intervening quizzes. The first variation was the difference in wording for the intervening quizzes to reduce recognition of answers. In many studies similarity between intervening quiz wording and other study material to the post-test raised concerns. Is recognition the underlying factor in the apparent increase in post-test score gains? McDaniel et al. (2007) writes that “on this principle, multiple choice quizzes would presumably promote better transfer to the final multiple choice test than would the short answer quizzes. The intervening quizzes, for the present study, consisted of multiple choice, fill in the blank, essay, and application analysis questions. The majority of the questions were multiple choice. However, questions were intentionally picked to reduce the similarity from the pre-test/post-test question wording. The current study research population varied in age, gender, grade level, academic readiness, and academic courses. The student’s motivational dynamics, study habits, and behavior were relatively diverse. Absenteeism and class disruptions were much higher than in previous years. The variety in motivational dynamics with a majority of the students not studying beyond the classroom and students missing opportunities to interact with the content do to absenteeism one could assume a limited learning gains. Questions could be raised the testing effects ability to increase learning depending on the content material so the content coverage for this study varied from basic chemistry to chemical reactions in the chemistry course and basic chemistry to genetics in the biology course. In laboratory settings the quizzes were frequent and in short interval so, was the learning gains do to short term memory and would it remain over longer periods of time? The length of time between the pre-test and post-test ranged from 3 weeks to 4 weeks which is significantly longer than laboratory studies. Even with class variation, increased time between pre-test and post-test, different content coverage, modification of quizzing questions to reduce similarity, and small sample sizes, learning gains were higher in the experimental group in 7 of the 15 data sets.
The variability among chapters and individual classes learning gains could have been associated with the chapter content and variation in sample sizes from chapter to chapter. To address the issue of small sample sizes in high school settings the individual data were pooled by chapters. All the pooled chapter data indicate the positive effects of testing on learning gains. The biology course consisted of only one section with small sample sizes but again the positive effects of quizzing on learning gains were evident in 2 of the 3 chapters.

Analysis of the pooled data consistently resulted in a difference in learning between quizzed and non-quizzed topics. Frequent quizzing in all instances was at least as effective as the control method. The positive effects of quizzing on learning gains were also evident in half of the individual classes despite small sample sizes. The findings with this study in a high school setting appear similar to the findings with others studies involving the “Testing Effect”. The process of retrieving material through recognition or recall with various techniques has shown to improve overall learning (McDaniel et al., 2007) Reviewing, studying, feedback loop, and other educational techniques invoke the retrieval process like quizzing. The act of retrieval through quizzing however, adds a level of amplification the same as conventional methods and in several cases greater than conventional methods. Teachers do not always see the merits of testing as a teaching tool. However, the strengths of testing and quizzing as a teaching instrument and its ease of assimilation merit its attention as a method to increase student learning in a high school setting. America’s future as a world leader is incumbent upon our success in the STEM fields and this teaching instrument could assist secondary education teachers with this endeavor.

Formative assessment can be used as an evaluative tool but the benefits of frequent quizzing do not stop at evaluations. Testing can be used as a metacognitive tool for students to evaluate what they know and do not know. Frequent testing can reduce the anxiety associated with testing. Assessment can help students focus on what they do not know to improve the efficiency of study time. Testing can also serve as a motivational tool. Testing then becomes not only an assessment tool but a review tool that can give a performance advantage.
The present study focused on intervening quizzes between the pre-test and post-test. The experimental content was exposed to an additional recall opportunity with the quizzing. However, the study did not include an extra interaction (recall opportunity) for the control material. Future studies might add additional recall opportunities for the control question set in the form of a homework assignment. Short answers and essay questions were components of the intervening quizzes. Future studies might analyze the possible differences in learning gains between multiple choice quizzes and essay quizzes. Students initially disliked the quizzes but when the quizzing stopped they asked for the quizzes to resume. A questionnaire before the research started and after the research was completed might also provide insight into the students motivational changes associated with the implementation of the quizzing.

Difficulties arise when trying to perform research in high school setting which are evident in this paper. Sample size fluctuated considerably. The number of pre-test and post-test data sets can drastically change with class absenteeism due to many factors. Class interruptions for school functions, bad weather, and office calls resulted in delay of presenting material in the designed time. Finally, variability across a class and multiple classes can create variability in the performance of the study groups.

Teachers are continuously asked to implement new strategies to try and improve the overall effectiveness of their teaching. New teaching strategies are developed or old strategies are revised but little consideration is given to their ability to be fully implemented. The key to improving science education is to identify strategies that are effective and easily implemented into a high school classroom. No matter what the strength of a strategy if it is not implementable it will not be effective. Learning gains for the quizzed material were greater in all of the pooled chemistry chapters. For two of the three biology chapters quizzing resulted in greater learning gains. It is evident that the testing effect has the ability to increase learning gains. These results indicate that using tests and quizzes as an educational tool can have a positive impact on student learning gains. I recommend that
teachers take a second look at test as not just a summative tool but a formative and constructive tool. Quizzing can be used to increase student learning. Testing is an easily accessible tool that serves as diagnostic tool, feedback tool, efficient study device, a metacognitive device, and an easily implemented learning strategy.
REFERENCES


Kornell, Nate, and Sons K. Lisa. The online platform for Taylor & Francis Group content. Louisiana State University. 25 July 2013.


Stine, Deborah, and Christine Matthews. "DigitalCommons@ILR." Cornell University ILR School. 27 June 2012 <http://digitalcommons.ilr.cornell.edu/key_workplace/644/>.


APPENDIX A TEST QUESTION EXAMPLES

**Mass number (Tested)**

3) What is the approximate atomic mass of an atom with 16 neutrons, 15 protons, and 15 electrons?

A) 15 daltons  
B) 16 daltons  
C) 30 daltons  
D) 31 daltons  
E) 46 daltons  
Answer: D  
Topic: Concept 2.2  
Skill: Knowledge/Comprehension

**Atomic Number (Quizzed)**

3. Calcium has an atomic number of 20 and an atomic mass of 40. Therefore, a calcium atom must have

A) 20 protons.  
B) 40 electrons.  
C) 40 neutrons.  
D) A and B only  
E) A, B, and C  
Answer: A  
Topic: Concept 2.2  
Skill: Knowledge/Comprehension

**Electron Jump (TESTED NOT QUIZZED)**

4. Electrons exist only at fixed levels of potential energy. However, if an atom absorbs sufficient energy, a possible result is that

A) an electron may move to an electron shell farther out from the nucleus.  
B) an electron may move to an electron shell closer to the nucleus.  
C) the atom may become a radioactive isotope.  
D) the atom would become a positively charged ion, or cation.  
E) the atom would become a negatively charged ion, or anion.  
Answer: A  
Topic: Concept 2.2  
Skill: Knowledge/Comprehension
APPENDIX B PARENTAL PERMISSION FORMS

Project Title: Analysis of the Testing Effect on Retention of Acquired Information

Performance Site: Leesville High School

Investigators: Donell Evans

The following investigator is available for questions,
M-F, 9:00 a.m.-3:00 p.m.
Dr. Joseph F. Siebenaller 225-578-1746 225-578-5224
Donell Evans 337-353-778 337-239-3464

Purpose of the Study: The purpose of the research project is to develop effective teaching strategies for improving student knowledge retention. This study involves the testing effect. The study will evaluate the apparent positive effect of testing on retention of learned material by high school students. The magnitude of the testing effect will assessed by comparing the performance of students on pre-and post-test given before and after materials is presented with quizzing and without quizzing.

Inclusion Criteria: High School Science courses in Mrs. Evan’s classrooms

Exclusion Criteria: None

Description and Purpose of the Study: Over a period of one school year, the investigator, will analyze the test data from pre-test to post-test to analyze any significant different between materials quizzed verses material not quizzed during the course of a teaching unit.

Benefits: The students will have frequent opportunities for analysis of their test data, more feedback from the test data, and more opportunities to interact with the classroom material.

Risks: There are no known risks.

Right to Refuse: Participation is voluntary, and a child will become part of the study only if both child and parent agree to the child’s participation. At any time, either the subject may withdraw from the study or the subject’s parents may withdraw the subject from the study without penalty or loss of any benefit, which they might otherwise be entitled. The class procedures will not be altered if the student withdraws from study. The student will still be responsible for all test and quizzes involved in the daily procedures.

Privacy: Results of the study may be published, but no names or identifying information will be included for publication. Subject identity will remain confidential unless disclosure is required by law.
Financial Information: There is no cost for participation in the study, nor is there any compensation to the subjects for participation.

Signatures:

The study has been discussed with me and all my questions have been answered. I may direct additional questions regarding study specifics to the investigator. If I have questions about subjects' rights or other concerns, I can contact Robert C. Mathews, Chairman, Institutional Review Board, (225) 578-8692, irb@lsu.edu, www.lsu.edu/irb. I will allow my child to participate in the study described above and acknowledge the investigator's obligation to provide me with a signed copy of this consent form.

Parent's Signature:________________________________ Date:____________________

The parent/guardian has indicated to me that he/she is unable to read. I certify that I have read this consent form to the parent/guardian and explained that by completing the signature line above he/she has given permission for the child to participate in the study.
APPENDIX C STUDENT CONSENT FORM

Signature of Reader:______________________________ Date:____________________

Student’s Signature_______________________ Age__________ Date ____________

Witness____________________________ Date __________________________

(N.B. Witness must be present for the assent process, not just the signature of the minor)

Institutional Review Board
Dr. Robert Mathews, Chair
203 B-1 David Boyd Hall
Baton Rouge, LA 70803
P: 225.578.8692
F: 225.578.6792
irb@lsu.edu | lsu.edu/irb
APPENDIX D
IRB FORMS

Application for Exemption from Institutional Oversight

Unless qualified as meeting the specific criteria for exemption from Institutional Review Board (IRB) oversight, all LSU research projects involving human research subjects, or samples, or data obtained from humans, directly or indirectly, with or without their consent, must be approved or exempted in advance by the LSU IRB. This form helps the IRB determine if a project may be exempted and is used to request an exemption.

- Applicant: Please fill out the application in its entirety and include the completed application as well as parts A-E listed below, when submitting to the IRB. Once the application is completed, please submit two copies of the completed application to the IRB Office or to a member of the Human Subjects Review Committee. Members of this committee can be found at http://research.lsu.edu/compliance/policies/Procedures/InstitutionalReviewBoard/Forms/View_form?ID=24737.html

- A Complete Application Includes All of the Following:
  (A) Two copies of this completed form and two copies of part B thru E.
  (B) A brief project description (adequate to evaluate risks to subjects and to explain your responses to Parts 1,2,3)
  (C) Copies of all instruments to be used.
  *If this proposal is part of a grant proposal, include a copy of the proposal and all recruitment materials.
  (D) The consent form that you will use in the study (see part 3 for more information.)
  (E) Certificate of Completion of Human Subjects Protection Training for all personnel involved in the project, including students who are involved with testing or handling data, unless already on file with the IRB.
  Training link: http://phpsrtraining.com/course/login.php

1) Principal Investigator: Joseph F. Siebenaler
   Dept: Biological Sciences
   Title: Professor
   E-mail: josef@lsu.edu

2) Co-Investigator(s): please include department, rank, phone, and e-mail for each.
   DONNIE FOWLS, Department of Biological Sciences, Graduate Student (AMISTI MNS program) and High School Teacher.
   (J.F. Siebenaler is the supervising professor) phone: 225-578-7796, email: donnie@lsu.edu

3) Project Title: Effects of Frequent Testing

4) Proposal? (yes or no)
   If Yes, LSU Proposal Number
   Also, if YES, either
   ○ This application completely matches the scope of work in the grant
   OR
   ○ More IRB Applications will be filed later

5) Subject pool (e.g., Psychology students)
   High school science students
   *Circle any "vulnerable populations" to be used (e.g., children, the mentally impaired, pregnant women, the aged, etc.). Projects with inaccessible persons cannot be exempted.

6) PI Signature: Date: 17 May 2012
   Signature:

** I certify my responses are accurate and complete. If the project scope or design is later changed, I will re-submit for review. I will obtain written approval from the Authorized Representative of all non-LSU institutions in which the study is conducted. I also understand that it is my responsibility to maintain copies of all consent forms at LSU for three years after completion of the study. If I leave LSU before that time, the consent forms should be preserved in the Departmental Office.

Screening Committee Action: Exempted ✔ Not Exempted Category/Paragraph

Reviewer: Mathew Signature: Date: 6/5/12
VITA

Donell Leondra DeBacker Evans was born to Donald and Elizabeth DeBacker. She attended St. Mary’s Catholic primary school in Omaha, Nebraska and Washington elementary school in Fremont, Nebraska. She graduated from Evans High School in Evans, Louisiana. She entered Northwestern State University in Natchitoches, Louisiana and earned a Bachelor’s of Science Education in Biology with a cumulative GPA or 3.8. She was a National Science Foundation teaching scholar her last two years of college. Donell earned certification in Physics and Chemistry education in 2002. She entered Louisiana State University Graduate School in June 2011 and is a candidate for the Masters of Natural Science Degree. She earned her first Masters of Science degree in May of 2013 with a degree in Educational Leadership. She teaches advance placement chemistry, advance placement biology, and general chemistry at Leesville High School. She is currently the head of the science department and an AP Coordinator at Leesville High School in Leesville, Louisiana, where she was named Teacher of the Year 2009.