Pretesting in science: effect on unit test scores

Mary Catherine Johnson

Louisiana State University and Agricultural and Mechanical College

PRETESTING IN SCIENCE:
EFFECT ON UNIT TEST SCORES

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 Sciences

by
Mary Catherine Johnson
B.A., Louisiana State University, 2007
August 2012

ACKNOWLEDGEMENTS

I would like to express my extreme gratitude to my thesis advisor and committee chair, Dr. John Hopkins. Thank you so much for guiding and editing my thesis toward completion. I would also like to thank Dr. Dana Browne and Dr. Cyrill Slezak for simplifying statistics so that my fellow candidates and I could understand them. They both worked very closely with our group from the beginning dispensing sage advice.

To my family, thank you for supporting me with your words of encouragement. I could not have completed this thesis without all of you. To my wonderful mother and father, thank you for instilling in me the importance of knowledge and my education. I love you both very much.

To my MNS classmates, it has been a pleasure working with you all and getting to know you. I would like to especially thank Mandy LeBourgeois for making my LaMSTI experience one that I will treasure forever. I am proud to call you one of my best friends.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ii

LIST OF TABLES iv

LIST OF FIGURES v

ABSTRACT vi

INTRODUCTION 1

MATERIALS AND METHODS 8

RESULTS 11

DISCUSSION AND CONCLUSION 24

REFERENCES 30

APPENDIX A: BAR GRAPHS OF UNITS 3, 4, & 9 POSTTEST MEANS 31

APPENDIX B: BAR GRAPHS OF ALL THE UNITS' EXPERIMENTAL PRETEST MEANS (NO HONORS) 33

APPENDIX C: BAR GRAPHS OF ALL THE UNITS' EXPERIMENTAL NORMALIZED GAINS (NO HONORS) 35

APPENDIX D: INSTITUTIONAL REVIEW BOARD APPROVAL FORM 37

VITA 38

LIST OF TABLES

Table 1: Unit Edusoft posttest means……………………………………………………………………………………13

Table 2: Unit Edusoft pretest means……………………………………………………………………………………..15

LIST OF FIGURES

Figure 1: Bar graph of comprehensive pretest means 11

Figure 2: Bar graph of Unit 2 posttest means 12

Figure 3: Bar graph of all units posttest means 14

Figure 4: Bar graph of experimental pretest means for each unit 15

Figure 5: Bar graph of experimental normalized gains for each unit 17

Figure 6: Bar graph of posttest means for each unit 18

Figure 7: Bar graph of control and experimental posttest means for each unit 19

Figure 8: Bar graph of comprehensive posttest means 20

Figure 9: Bar graph of comprehensive normalized gains 21

Figure 10: Bar graph of EOC means 22

ABSTRACT

This study examines the effect that pretesting has on unit test scores in a high school Biology I classroom. The experimental classes were given a pretest before four units of curriculum and then taught the concepts of these units with a traditional lecture based methodology. Their unit test scores were compared to a control class that did not take pretests before the units. In addition, the End of Course (EOC) test scores of the control and experimental classes were compared to determine if pretesting improved transfer of knowledge to a different type of test.

Pretesting did not cause a statistically significant difference between the unit test scores of the control and regular experimental classes. However, the honors experimental class did have higher unit test scores than the control class as well as the regular experimental classes. It was also found that there was no significant difference between the control classes' EOC scores and the regular classes'. Again the honors experimental class did have higher EOC scores than the control class and the regular experimental classes.

Pretesting is not an effective tool to increase unit test scores for regular education students. It does not seem to affect learning gains or transfer of knowledge for these either for these students. Honors students performed better overall in this study maybe as a result of pretesting.

INTRODUCTION

A primary goal of teaching is to convey information to students in such a way as to ensure retention. There is a continuous search for best practices that will facilitate this learning. Beside the fact that teachers want their students to bring the knowledge that they impart upon them to future endeavors, teachers also need them to perform adequately on statewide unit exams, end of course testing, and other standardized tests. Throughout the school year, teachers have to strive to teach the entire state-mandated curriculum in a limited amount of time. That being said, teachers do not want to spend time on any method that does not prove itself to truly improve retention and test scores.

Many teachers give pretests to assess prior knowledge about a new topic being introduced. The analyses of these exams are used to find holes in students' understanding and use these gaps as guides for future lessons. The research outlined in this study was conducted to see if pretests could be even more helpful in the classroom. Can pretests help students to recall information from lectures and retain more concepts that they would see on exams, thus improving their scores?

Seemingly, pretests would help students to recognize the particular information that would be tested and remember it more easily. This study sought to find if the same would ring true with high school students in a Biology I class. In East Baton Rouge Parish, teachers are required to give standardized tests for each assigned unit called Edusoft tests. These tests consist of numerous multiple-choice questions and one or two constructed response questions. The questions are pre-determined for each Edusoft test and not chosen individually by each

teacher. The scores that the students receive on these tests are reviewed by students, teachers, administration, and school board employees. This research utilized these exams, because at this point, they are the main formal assessment for the Biology I curriculum.

Across Louisiana, a teacher's salary will soon depend on how their students perform on the Edusoft exams. A teacher's salary will fluctuate based on whether their students' scores rise or fall from year to year. Although teachers should strive to increase their students' abilities and be held accountable for their levels of comprehension, oftentimes these factors are difficult to control. For example, students must listen to what you are teaching and continue to study the concepts throughout the year to retain the material and perform adequately on their Edusoft exams and on end of course testing. Pretesting could be a tool to help the students to do just that.

To aid students in retaining information, teachers need to continually modify their methods and seek out best practices. It is extremely frustrating to try a new teaching method thinking that it will improve retention and later finding out that it was a waste of time. Time is precious when lengthy lessons must be piled into fifty minute class periods. By doing this study on the benefits of pretesting and improving students' test scores, teachers could gather insight on whether pretesting was an effective use of the limited available time. Other studies have looked at specific benefits of pretesting and the testing effect (tests being used as a learning and retention tool as opposed to purely assessing knowledge) in general. In other words, using tests to improve retention rather than just as an assessment (Karpicke & Roediger, 2006).

Some studies have sought to see if pretesting not only focused the learners' attentions to the tested material, but also improved their retention of the material. In the first trial done

by Kao, Kornell, and Richland (2009), which consisted of 63 undergraduate students, they compared a pretesting condition with an extended study condition to see which group scored higher on a posttest following the reading of a passage on colorblindness. The first trial found that, “although participants (pretest condition) largely failed on the initial test, the effect of those failures was to increase retention of studied content when compared with an extended opportunity to study the materials without being pretested.” This conclusion was reached after analyzing both conditions’ scores on a posttest that immediately followed. The pretesting condition scored better with a 95% confidence level with a mean score of M=75% compared to the extended study condition which scored M=56%. The second trial was set up the same, but this time the tested material in the passage was italicized to eliminate the effect of attention direction in order to determine if pretesting had an effect on retention. Again the pretesting condition students scored higher than the extended study students with a mean score of M=71% versus M=53% with a 95% confidence level. The findings of the second experiment showed that, “testing items created more potent learning opportunities than an extended study time of the same items, even when the key information in both conditions was italicized, equalizing attention direction. Thus, testing appears to provide a unique benefit above and beyond directing learners’ attention to content that has a high probability of being tested later” (Kao, Kornell, & Richland, 2009).

Focusing the students’ attentions on the most important information to study is another benefit of testing. Another study states, “the metacognitive use of tests lets students inform themselves about what they know and do not know so they can concentrate future study efforts on the information that they do not know” (Agarwal *et al*., 2011). This study did not

look specifically at pretesting, rather, it looked at the testing effect in general. The authors conducted three experiments in which they tested whether quizzing increased unit test scores compared to simply rereading the material. In the first two experiments 142 sixth grade Social Studies students were given a multiple-choice quiz immediately following a lecture on a certain topic. They were then given posttests a few days later that included pretested questions as well as questions they had never seen before. The students performed higher on the pretested questions than on the novel questions. The first experiment resulted in statistically significant different mean values of M=94% on tested items and M=81% on non-tested items with a confidence level of 95%. One reason for the higher performance which the authors concluded was that the students were more aware of the information that they needed to focus on during their independent study. Testing as a learning tool can help students to study more effectively and efficiently (Agarwal *et al*., 2011).

Multiple studies focus on pretesting as a tool to improve short term retention in which posttests are administered the same day or a few days later. In such a study conducted by Bjork and Little (2010), the authors considered the effect of taking a multiple-choice pretest on the future recall of pretested and non-pretested information. In one of their experiments, 24 participants were given two passages, one on Saturn, and one on Yellowstone National Park. Before one of the passages, 10 multiple-choice questions about that passage were answered with 4 minutes of time allotted. For the same passage, 10 minutes of study time were then given for a total of 14 minutes on the first passage. On the second passage, the participants were not given a multiple-choice pretest but allowed 14 minutes to study the passage. The participants were then given a five minute retention time during which they played Tetris (a

spatial-reasoning puzzle game). They then completed a final recall test which consisted of half pretested questions and half non-pretested questions. The authors concluded through a t-test that the p-value of pretested questions (M=61%) versus non-pretested (M=43%) questions was less than .05, indicating a statistically significant difference with a confidence level of 95%. This study shows that multiple-choice pretests improve short term retention where the retention time was only 5 minutes.

A study conducted by Roediger and Karpicke (2006) looked at the testing effect on long-term retention. The two experiments in their study used 120 and 180 Washington University undergraduate students, respectively. The students had to study a prose passage taken from the test-preparation book for the Test of English as a Foreign Language and then immediately take one or three recall tests or instead studied the passage one to three times. They then had to take final retention tests either 5 minutes, 2 days, or one week later. The results of the data analysis concluded that the students that took the recall tests immediately following the passage retained more information one week later than did the students that studied the passage. In the second experiment, students in the repeatedly tested condition scored a significantly different mean value of M=61% compared to the repeated study condition with a mean value of M=40% with a 95% confidence level (Karpicke & Roediger, 2006).

One particular study addressed the negative effects that multiple-choice testing could pose in the form of lures. Lures are answers to a multiple-choice question that students choose incorrectly and then endorse to be right. Butler and Roediger (2008) looked to see whether feedback could enhance the positive effects of multiple-choice testing and lower the negative effects. Seventy two undergraduate students participated in the study. The participants read 12

passages on a variety of historical topics. Five minutes later, they took a computerized multiple-choice recall test about the passages. The participants were either given immediate feedback after each question, delayed feedback after completion of the test, or no feedback at all. In the feedback, participants were given an indication of the accuracy of their answer, the question shown again along with their answer, and the correct response to the question. The participants returned one week later for the same multiple-choice test. Those that received delayed feedback performed statistically better than the ones that received immediate feedback. The participants that received immediate feedback performed statistically better than the ones that received no feedback. According to the authors, feedback should be given after a multiple-choice test to increase its positive effects (Butler & Roediger, 2008).

Finally, a study conducted by Butler (2010) looked to see whether the testing effect would enhance learning transfer compared to repeat studying. The author quotes a source to define transfer as “the influence of prior learning (retained until the present) upon the learning of, or response to, new material” (McGeoch, 1942). In the study, 48 undergraduate psychology students read 6 passages on a variety of topics. Half of the students then restudied the passages three times and the other half took three short answer tests about the passages. One week later the participants returned to take the final test which consisted of new inferential questions from the same domain. The repeat testing group scored significantly higher on the final test than the repeat studying group with a 95% confidence level. Butler concluded that the “benefits of test-enhanced learning are not limited to the retention of the specific response tested during initial learning but rather extend to the transfer of knowledge in a variety of contexts” (Butler, 2010).

This present study sought to determine the effect that pretesting had on unit Edusoft scores and a comprehensive Edusoft final in Biology I. It also looked to see if pretesting had an effect, through learning transfer, on the End of Course test taken by the same students.

MATERIALS AND METHODS

This study was designed to test the effect of pretesting on unit Edusoft exams in 10th grade Biology I classes. The students in the research group attended Broadmoor High School in Baton Rouge, Louisiana. Broadmoor had an enrollment of 1099 students. The school's student population was made up of 78% African Americans, 9% Caucasian, 8% Asian, and 5% Hispanic. Of these students, 80% were on free and reduced lunch. The sample population of students was made up of 170 students that participated in some portion of the study. The sample demographics were very similar to the school wide demographics with 78% of participants being African American, 9% Caucasian, 9% Asian, and 4% Hispanic. A total of five different classes participated in the study. The first period class, which acted as the control, consisted of 31 students. The experimental classes were the remaining third, fourth, fifth, and sixth period classes. The third period class, which was an honors class, consisted of 31 students. The fourth, fifth, and sixth period classes consisted of 30, 30 and 32 students respectively. (These counts were determined at the end of the school year.) The units tested in the study were Unit 2 "Balance in Nature", Unit 3 "Cellular Reproduction/Genetics", Unit 4 "Changes Over Time/Adaptations for Survival", and Unit 9 "Excretory, Endocrine, and Reproductive Systems." These units were chosen at random.

During the first week of school all of the students in the study were given a 71 question multiple-choice comprehensive Edusoft test. The test covered concepts that would be taught over the course of the entire school year. This test served as an inventory of the students' prior knowledge of Biology I concepts and the scores were analyzed to determine that the sample population was homogenous based on a p-value of greater than 0.05. East Baton Rouge Parish

requires that the students take standardized multiple-choice Edusoft exams that the parish provides at the end of each unit of study. In the fall semester, before any instruction for Unit 2, the Unit 2 Edusoft exam was given as a pretest to the four experimental classes. As an incentive, the students were instructed that the three highest pretest test scores in each class would receive bonus points. The Unit 2 Edusoft exam contained 29 multiple-choice questions and 2 constructed response questions. The “Balance in Nature” material which consisted of concepts on ecology, cellular respiration, and photosynthesis was then taught via power points, guided notes, and lectures over a period of four weeks to all five classes. Feedback was given to the experimental classes during the lecture that pertained to the pretest. The feedback was not specific to whether the students answered correctly but instead addressed misconceptions more generally. For example, “Do you remember this idea from the pretest? Which question did it relate to? Did you put the right answer?” At the end of the four weeks, the Unit 2 Edusoft exam was administered to all five classes to assess their knowledge of Unit 2 concepts. The same procedure was completed for Unit 3 over a six week period. The Unit 3 Edusoft exam contained 27 multiple-choice questions and 2 constructed response questions. The procedure was then repeated for Unit 4 over a period of four weeks. The Unit 4 Edusoft exam contained 29 multiple-choice questions and 1 constructed response question. In the spring semester, the process was repeated with Unit 9 over a period of two weeks. The Unit 9 Edusoft exam contained 28 multiple-choice questions and 2 constructed response questions. The Edusoft posttest scores were analyzed and compared between the control class and experimental classes. The experimental pretests and gains of each unit were also compared.

The state of Louisiana required that all Biology I students completed a computerized End of Course test. The test was comprised of three parts. The first and third parts were each 25 multiple-choice questions and the second part consisted of 3 short answer or constructed response questions. The test scores were based on an 800 point total value. A score of 800-739 was given a rating of excellent. A score of 738-700 was given a rating of good. A score of 699-668 was given a rating of fair. A score of 667 and lower was ranked as needs improvement. The students' total points and percent scores were recorded and analyzed to test the effect pre-testing had on the transfer of knowledge to a different type of test on related material.

Finally, during the last week of school, the same 71 question comprehensive Edusoft test was given to all five classes. The scores were recorded to compare the control class to the experimental classes and to analyze the overall gains of each student.

RESULTS

The results were first analyzed to determine if all five classes entered the study with the same prior knowledge of Biology I concepts. An analysis of variance (ANOVA) on the comprehensive Edusoft pretest scores showed no statistically significant difference between the means based on the p-value of $p = 0.486$. A 95% confidence level was used for this comparison and all future comparisons in this research. See Figure 1.

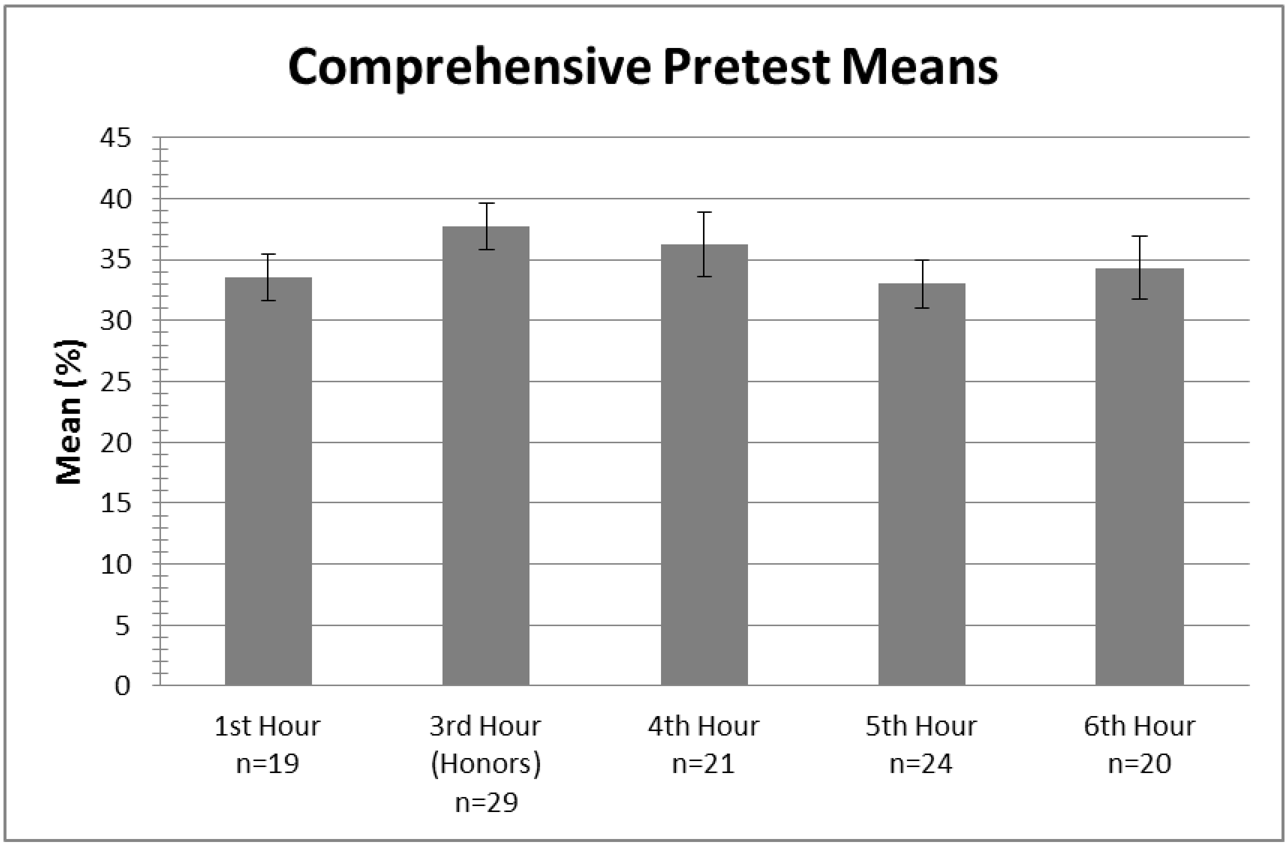


Figure 1: Bar graph of comprehensive pretest means

The data were then analyzed to determine whether the pretested classes scored higher on each unit's Edusoft posttest than the control class. An ANOVA test showed that there was a significant difference between the five classes' Unit 2 (Balance in Nature) posttest means with a p-value of p=1.29E-05. See Figure 2.

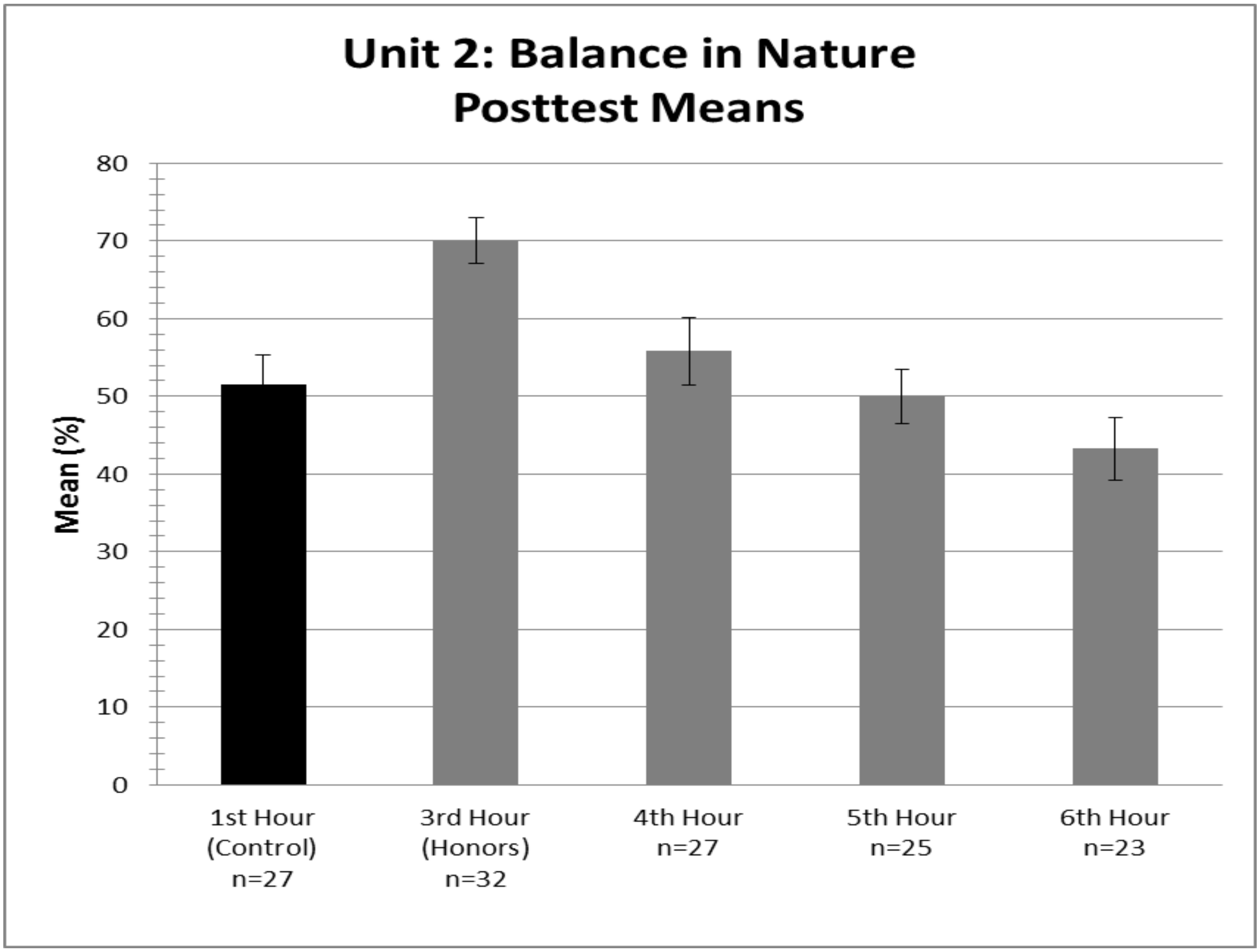


Figure 2: Bar graph of Unit 2 posttest means

Similar results were found when running an ANOVA for Unit 3 (Cellular Reproduction/Genetics) with a p-value of p=0.002, Unit 4 (Changes over Time/Adaptations for Survival) with a p-value of p=0.0003, and Unit 9 (Excretory, Endocrine, and Reproductive

Systems) with a p-value of p=0.001. See Appendix A for bar graphs of the posttest means of these units. Table 1 shows the five classes' mean scores on each units' Edusoft posttest along with the p-values from the ANOVA.

Table 1: Unit Edusoft posttest means

POSTTEST MEAN SCORES				
CLASS	**UNIT 2**	**UNIT 3**	**UNIT 4**	**UNIT 9**
1st Hour	52±4	54±4	57±2	64±4
3rd Hour	70±3	67±3	72±3	77±3
4th Hour	56±4	52±4	59±5	64±4
5th Hour	50±4	49±4	53±4	64±4
6th Hour	43±4	49±4	52±4	54±4
P-VALUES	1.3E-05	0.002	<0.001	0.001

Upon analyzing the posttest data for each unit, it was apparent that the experimental 3rd hour honors class' means were consistently the highest. See Figure 3.

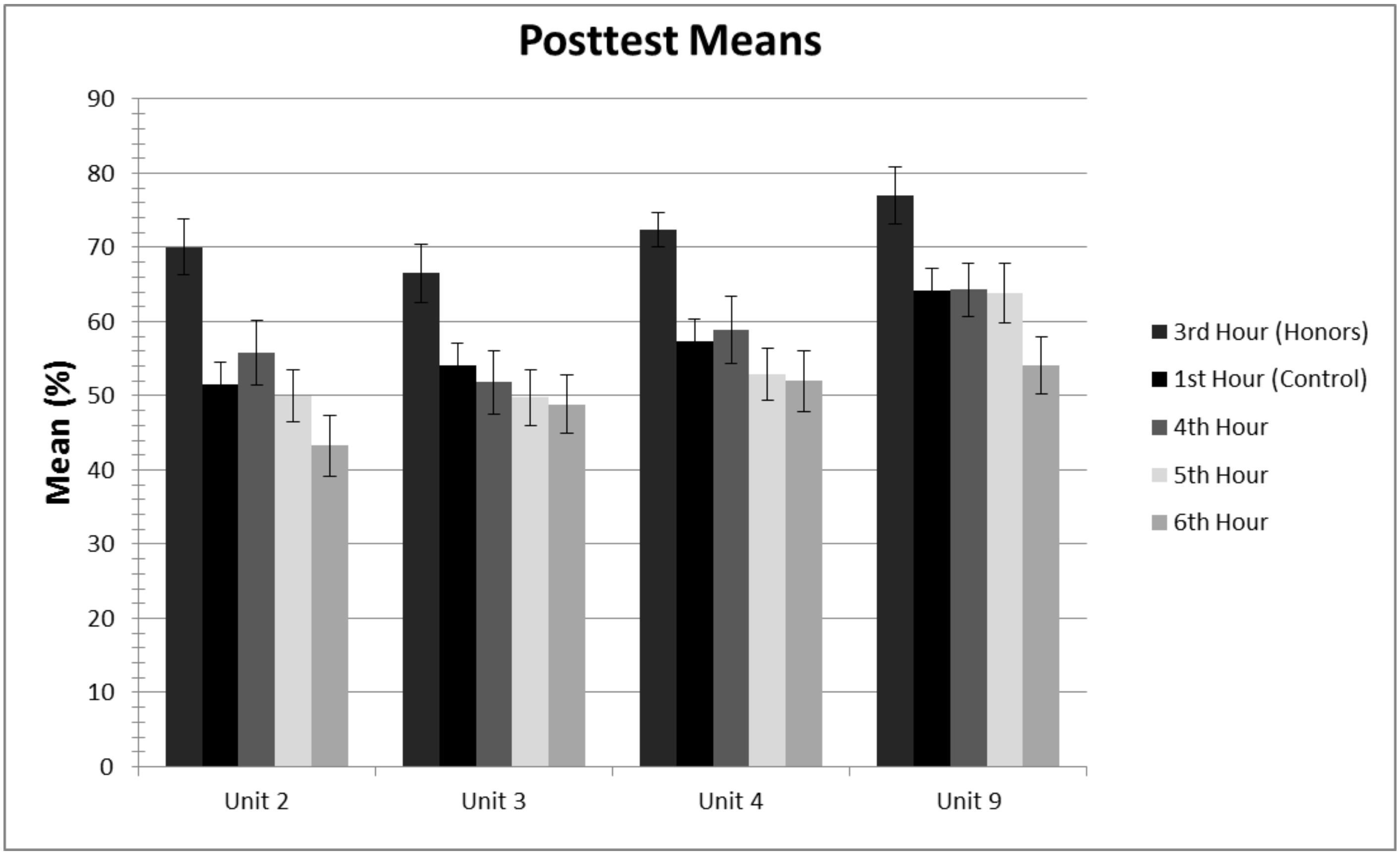


Figure 3: Bar graph of all units posttest means

After noticing this trend, an ANOVA was run on the pretests for each unit to determine if there was a difference between the four experimental groups at the beginning of each unit. The ANOVA for Unit 2 (Balance in Nature) resulted in a p-value of p=0.001 when comparing experimental pretest means showing that the four classes were significantly different. Similar results were found when running an ANOVA on experimental pretest means for Unit 3 (Cellular Reproduction/Genetics) with a p-value of p=0.007, Unit 4 (Changes over Time/Adaptations for Survival) with a p-value of p=4.46E-05, and Unit 9 (Excretory, Endocrine, and Reproductive Systems) with a p-value of p=0.009. See Figure 4. Table 2 shows the four experimental classes' mean scores on each unit's Edusoft pretest along with the p-values from the ANOVA.

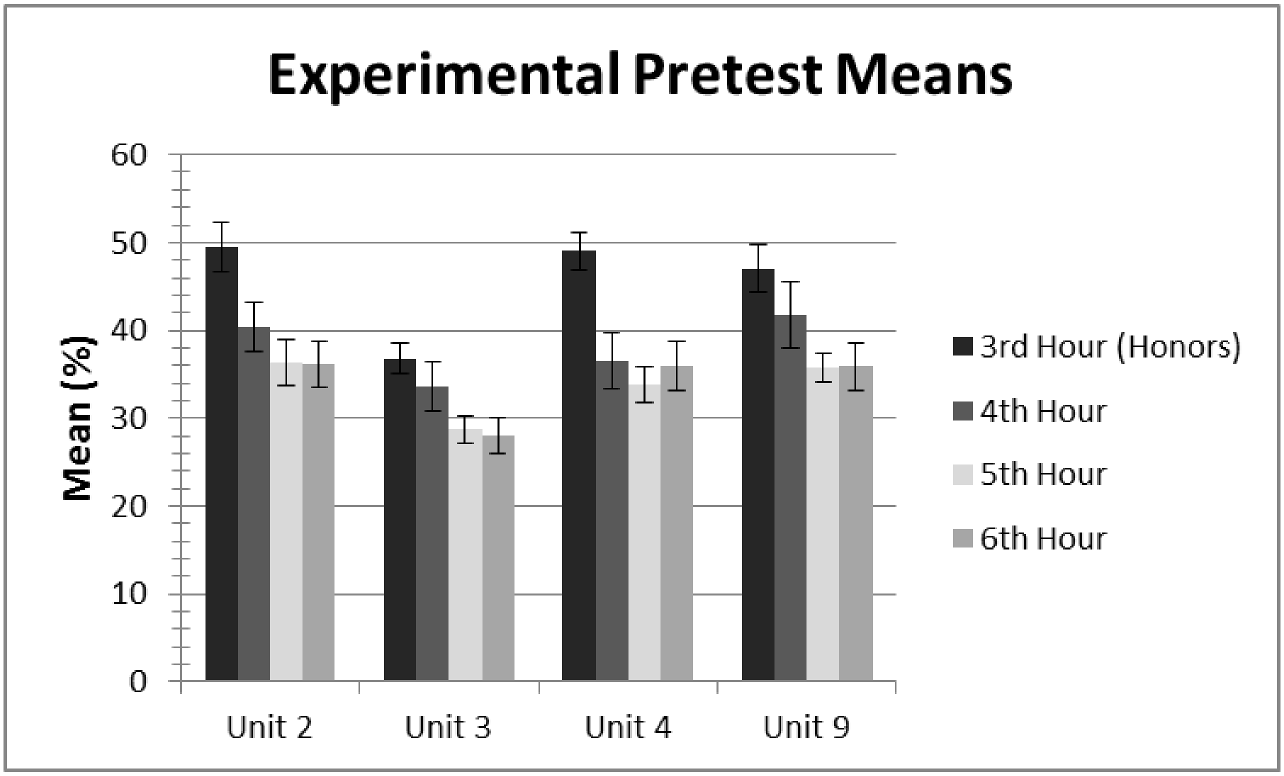


Figure 4: Bar graph of experimental pretest means for each unit

Table 2: Unit Edusoft pretest means

EXPERIMENTAL PRETEST MEAN SCORES				
CLASS	**UNIT 2**	**UNIT 3**	**UNIT 4**	**UNIT 9**
3rd Hour	50±3	37±2	49±2	47±3
4th Hour	40±3	34±3	37±3	42±4
5th Hour	36±3	29±2	34±2	36±2
6th Hour	36±3	28±2	36±3	36±3
P-VALUES	0.001	0.007	4.5E-05	0.009

Unlike the comprehensive Edusoft pretest means, the experimental unit Edusoft pretest means show a statistically significant difference. It is clear that the 3rd honors class' means began the highest for each unit. To see if the three regular experimental classes' pretests means began statistically the same, an ANOVA was run on the means excluding the 3rd hour honors class. The p-value for Unit 2 (Balance in Nature) of p=0.447 showed that the three regular experimental groups' pretest means were not significantly different. Similar results were found when running an ANOVA on experimental pretest means excluding honors for Unit 3 (Cellular Reproduction/Genetics) with a p-value of p=0.150, Unit 4 (Changes over Time/Adaptations for Survival) with a p-value of p=0.725, and Unit 9 (Excretory, Endocrine, and Reproductive Systems) with a p-value of p=0.240. The three regular experimental classes' pretest means show no statistically significant difference for each unit. See Appendix B for bar graphs of the pretest means excluding the honors class for all the units.

To provide further evidence that the 3rd hour honors experimental class was different from the three regular experimental classes, the experimental normalized gains for each unit were compared. Did the honors class not only start and end higher on the unit Edusoft tests, but did they also have greater gains for each unit? An ANOVA was used to for this comparison which showed a significant difference between the normalized gains for each unit between the honors experimental class and the three regular experimental classes. The p-values for each unit were as follows: p=5.73E-05 (Unit 2), p=0.003 (Unit 3), p=0.022 (Unit 4), and p=0.008 (Unit 9). See Figure 5.

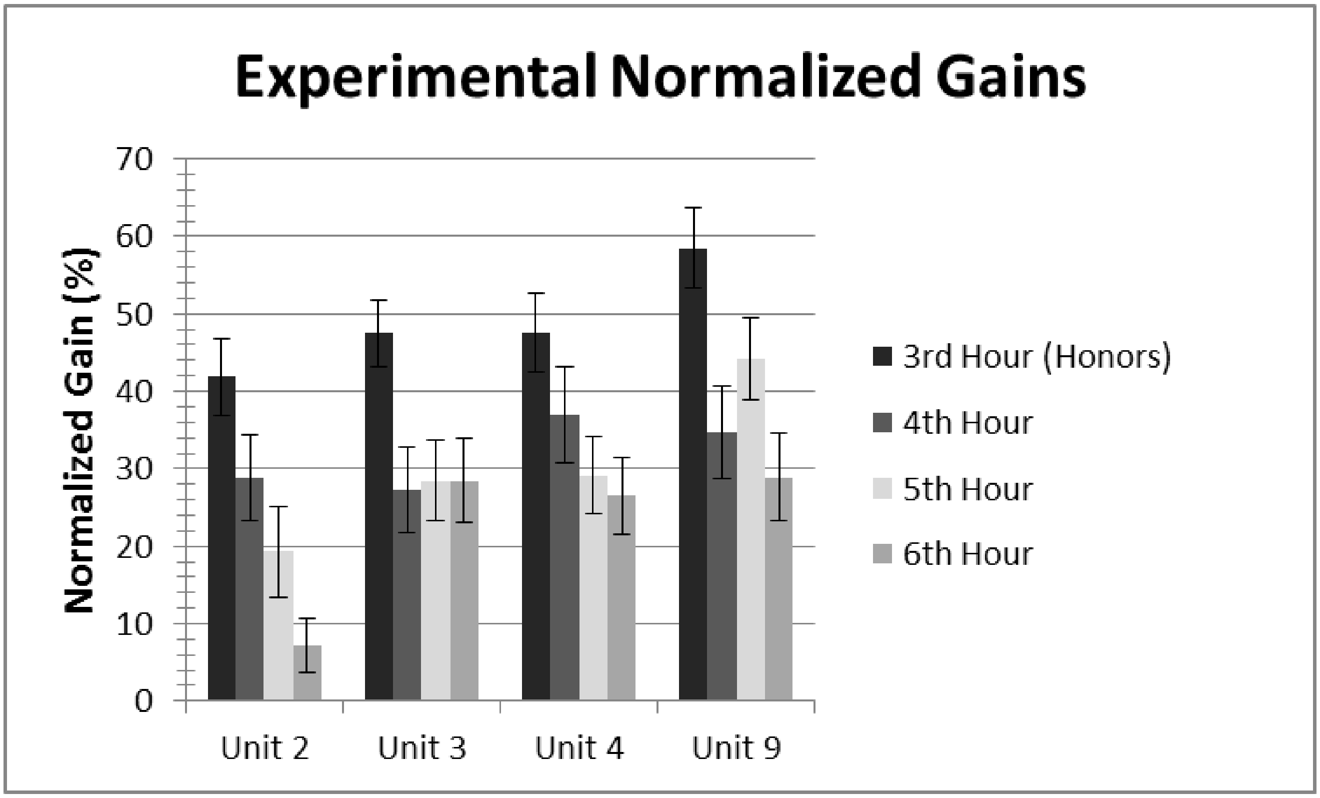


Figure 5: Bar graph of experimental normalized gains for each unit

Again it was obvious that the honors class outperformed the three regular experimental classes. An ANOVA was run to determine whether the three regular experimental classes had the same gains for each unit. The three regular experimental classes show no statistically significant difference when comparing normalized gains for Units 3, 4 and 9. The p-values were as follows: p=0.985 (Unit 3), p=0.389 (Unit 4), and p=0.256 (Unit 9). There was one exception to this trend. When comparing normalized gains for Unit 2 (Balance in Nature), the three regular experimental classes showed a significant difference with a p-value of p=0.018. It was apparent from analyzing the three normalized gain mean values (M=28.89±5.59 (4th hour), M=19.27±5.95 (5th hour), and M=7.22±3.48 (6th hour)) that the 6th hour class had particularly low gains for this

unit. See Appendix C for bar graphs comparing the three regular experimental classes' normalized gains for each unit.

Because the 3rd hour honors class' means stood out amongst all the units' posttest scores, pretest scores, and normalized gains, it became obvious that the experimental groups could be divided into two, 4th ,5th ,and 6th hours in one and 3rd hour honors in another. This was done to see if increasing the sample size of the experimental group would show different results. After this realization, the posttest scores for each unit were compared using an ANOVA between the control group, regular experimental group, and the honors experimental group. A significant difference was found between the three groups' posttest means for each unit. The p-values were as follows: p=1.09E-05 (Unit 2), p<0.001 (Unit 3), p=7.11E-05 (Unit 4), and p=0.001 (Unit 9). See Figure 6.

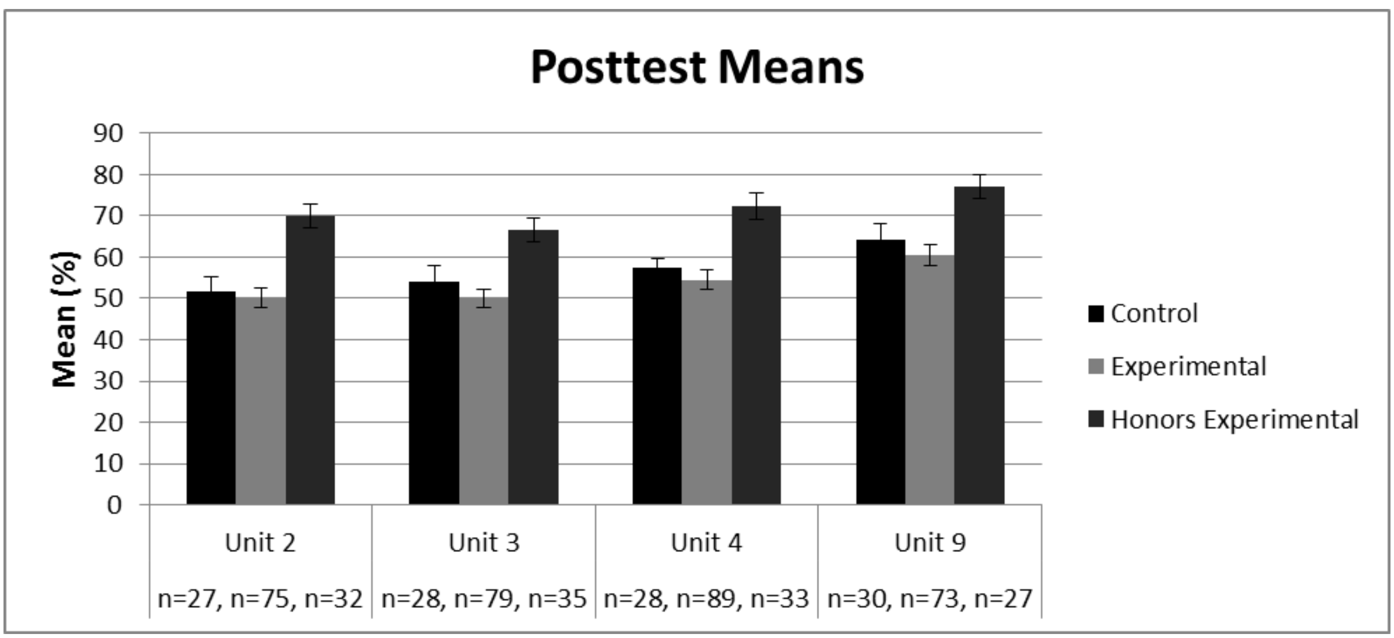


Figure 6: Bar graph of posttest means for each unit

A t-test was done on each unit's posttest means to compare the control group to the regular experimental group. Because it was apparent that the honors experimental group outperformed the regular experimental group for each unit, the honors experimental group was removed from the this comparison. The following p-values for each unit were found with this method: Unit 2 (Balance and Nature) p=0.732, Unit 3 (Cellular Reproduction/Genetics) p=0.360, Unit 4 (Changes Over Time/Adaptations for Survival) p=0.510, and Unit 9 (Excretory, Endocrine, and Reproductive Systems) p=0.406. These p-values showed no significantly statistical differences between the posttest means of the control and the experimental groups for each unit. See Figure 7.

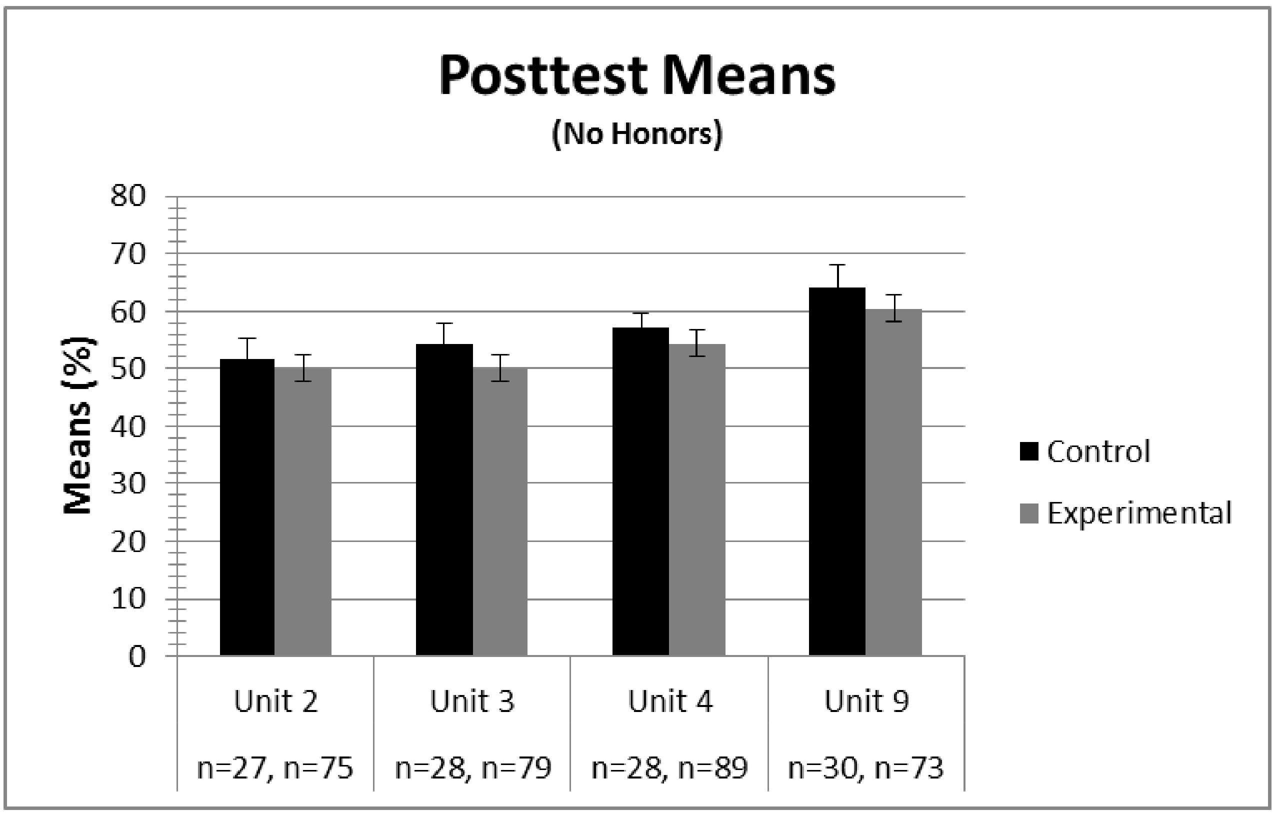


Figure 7: Bar graph of control and experimental posttest means for each unit

To see if there was a difference between cumulative knowledge, an ANOVA was done to compare the control group's, regular experimental group's, and honors experimental group's posttest means of the comprehensive Edusoft exam. Any student in the four experimental classes that did not complete a pretest for each unit (about 40% of experimental students) was excluded from this comparison. There was a difference between the three groups with a p-value of p=0.022. The honors experimental group again caused the difference as seen in Figure 8.

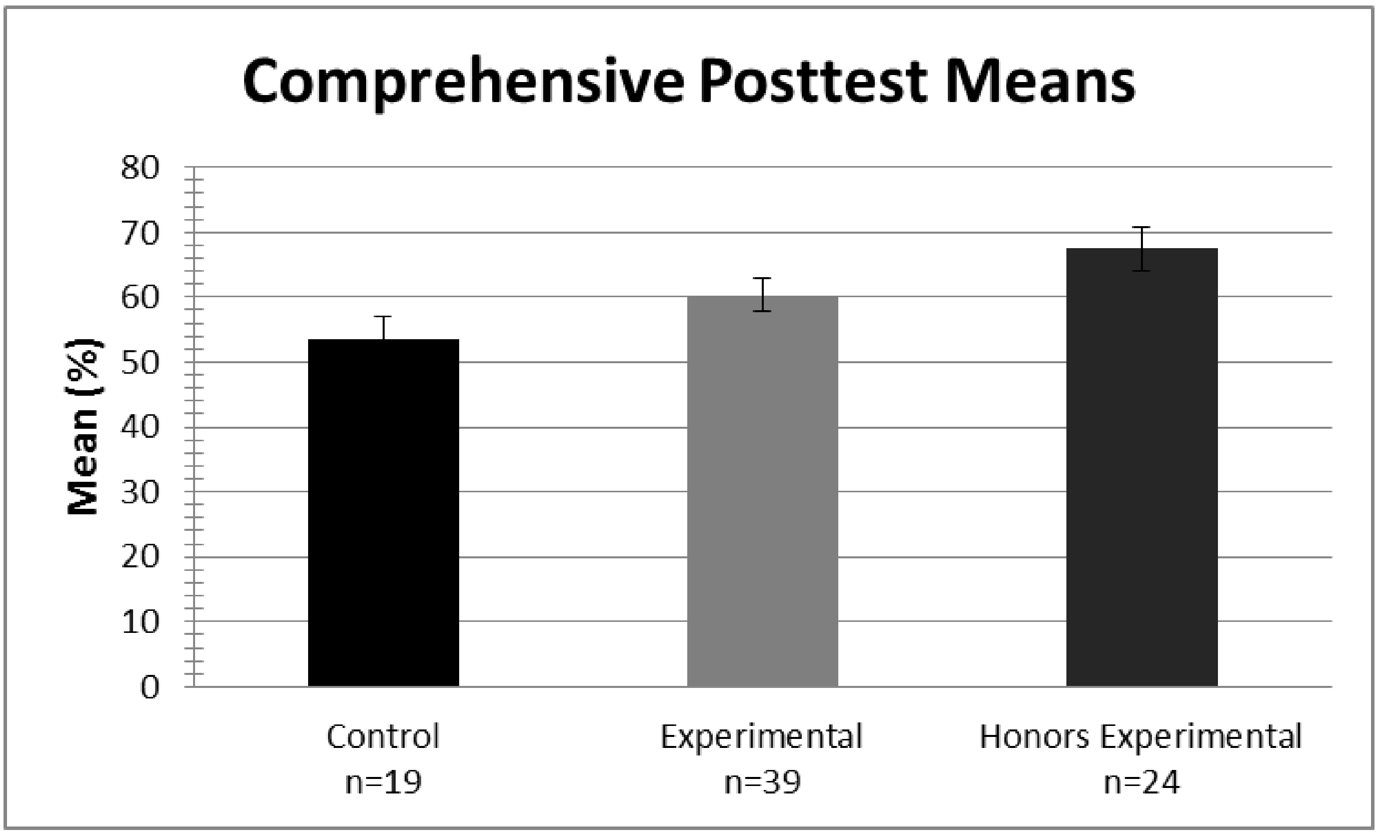


Figure 8: Bar graph of comprehensive posttest means

The honors experimental group was then removed from the comparison and a t-test was done to analyze the difference between the control group's and the regular experimental group's comprehensive posttests. There was no significant difference found between the means with a p-value of p=0.122.

To further validate the previous trends, an ANOVA was done to compare the comprehensive normalized gains between the control group, regular experimental group, and the honors experimental group. Again, any student in the four experimental classes that did not complete a pretest for each unit (about 40% of experimental students) was excluded from this comparison. There was a significant difference between the comprehensive normalized gains of these three groups as shown by the p-value p=0.018. See Figure 9.

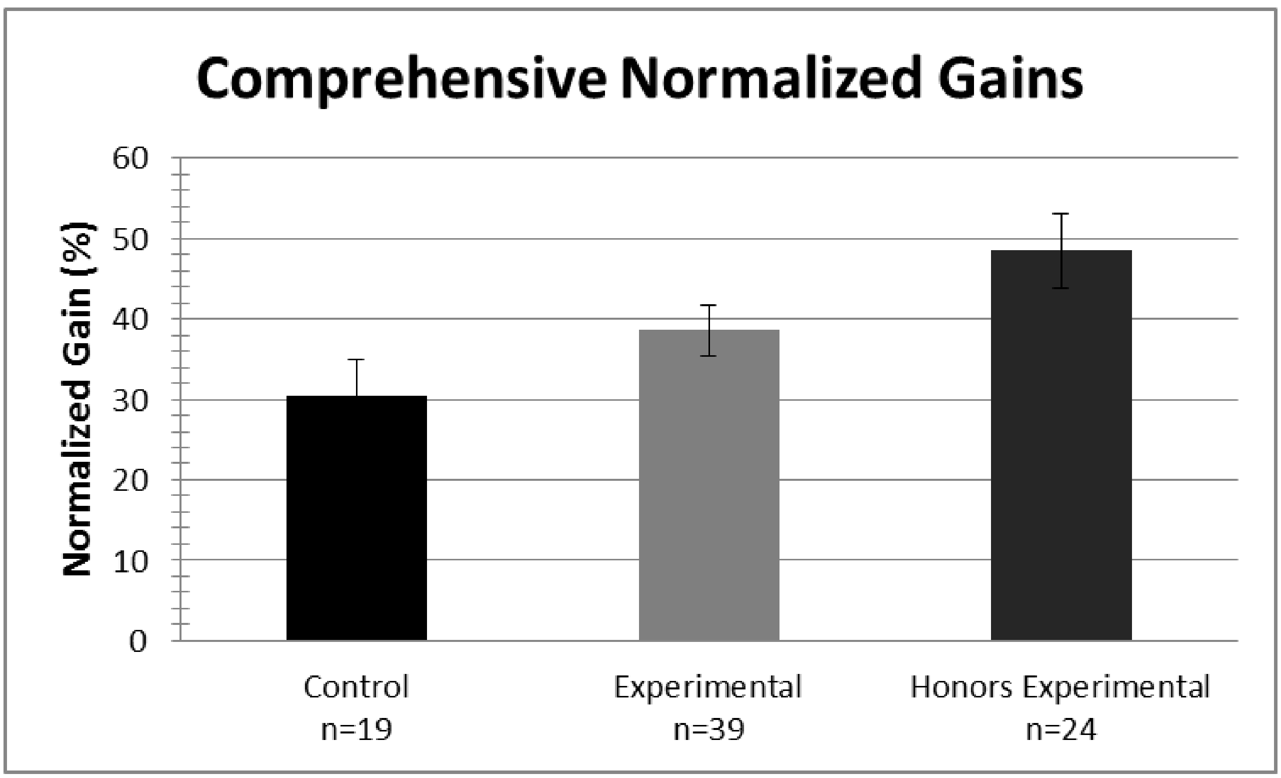


Figure 9: Bar graph of comprehensive normalized gains

When the honors experimental group was removed from this comparison and a t-test was run between the control group's and the regular experimental group's comprehensive normalized gains, there was no statistically significant difference as shown by the p-value p=0.140.

Finally, an ANOVA was performed on the End of Course (EOC) test score means of the control group, regular experimental group, and the honors experimental group. Any student in the four experimental classes that did not complete a pretest for each unit (about 46% of experimental students) was excluded from this comparison. The test resulted in a p-value of p=0.001 showing a significant difference between the three groups. See Figure 10.

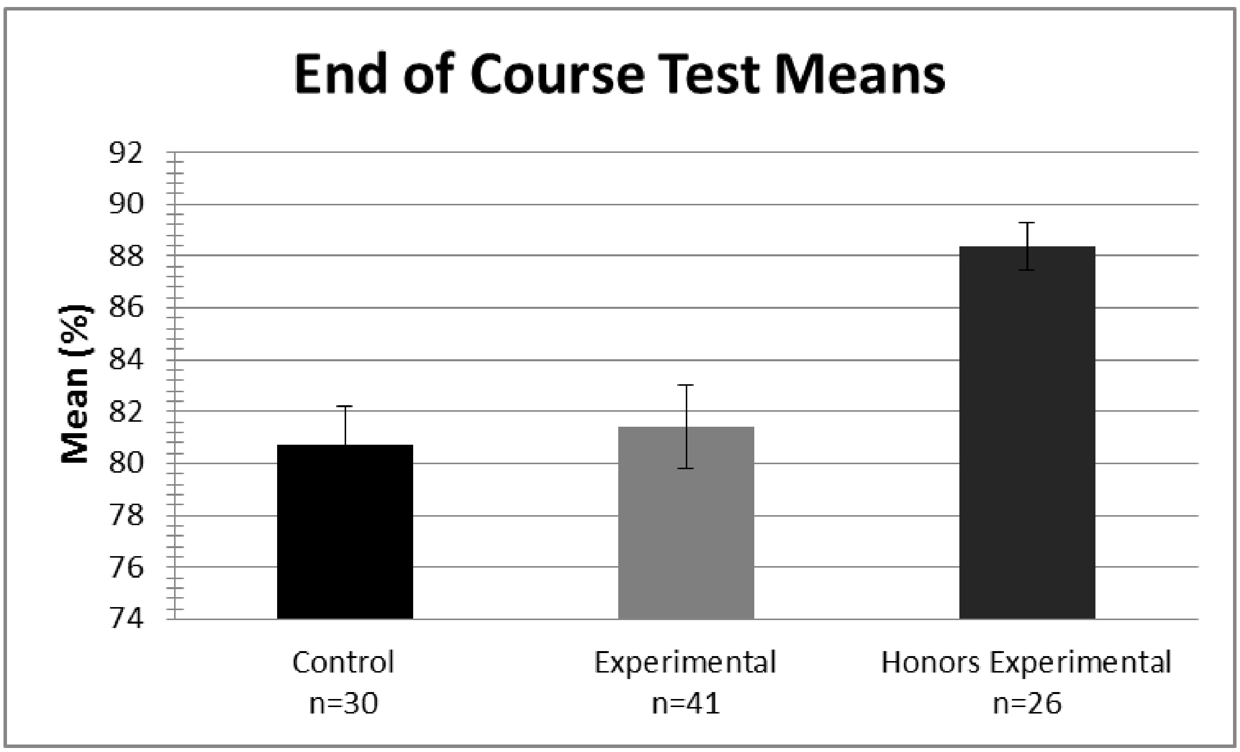


Figure 10: Bar graph of EOC means

After compiling evidence that the honors experimental group caused the statistically significant difference, a t-test was run to compare the control group's and regular experimental group's EOC mean scores. There was no significant difference between the two groups as shown by the p-value of $p=0.753$.

DISCUSSION AND CONCLUSION

The purpose of this study was to determine if pretesting increased unit Edusoft test scores in the high school Biology I classroom. Other studies have examined whether unsuccessful retrieval attempts on pretests would increase learning gains (Kao, Kornell, & Richland, 2009) or whether multiple-choice pretests in particular improve retention (Bjork & Little, 2011). Other studies have looked at different aspects of the testing effect on learning and retention (Butler & Roediger, 2008; Butler, 2010; Roediger & Karpicke, 2006; Roediger *et al*., 2011). If pretesting increased unit test scores, it could be used as another tool by teachers to help students pass required classes.

By analyzing the comprehensive Edusoft pretests for all of the classes, it was determined that the research began on a group of students with similar knowledge of Biology I concepts. No class started out with an advantage over the others.

The posttest means of each unit showed a significant difference between the five classes. The honors experimental class always had the highest posttest means (Figure 3 & Table 1). Furthermore, the posttest means of the four regular classes seemed to decrease as the day went on. Because there was no honors control class, it was not for certain that the higher posttest scores were caused by pretesting. It could be the case. It appeared that the honors students tended to spend a longer time taking the pretest and really focusing on the questions, while the regular classes would speed through it. Butler and Roediger (2008) stressed the importance of feedback to enhance the positive effects of multiple-choice testing. The honors students could have benefitted more from pretesting because they listened more intently to the lectures given after the test, in particular feedback given about the pretest questions and

answers. Another possibility could be that the honors students were more motivated by the bonus points offered to the three highest pretest scores. From personal observations, honors students, by nature, are more competitive and more concerned with their grades. On the other hand, the higher scores could just be attributed to greater effort, more studying, or a home environment that stressed the importance of school performance.

Even though the comprehensive pretests showed no statistically significant differences among the classes, the posttest means showed that the experimental pretests for each unit should be analyzed to investigate whether the four experimental classes started off the same for each unit.

Could the factors pertaining to honors students discussed previously start these students off with an advantage? The pretest analysis of each unit showed a definite trend with the experimental honors class always starting off with the highest scores (Figure 4 & Table 2). When the three regular experimental classes' posttest means were compared to each other, they showed no significant difference for each unit.

Since the honors experimental class began highest for each unit the normalized gains were analyzed to determine whether the honors experimental class really outperformed the three regular experimental classes. The normalized gains for each unit showed a significant difference between the experimental classes. It was again obvious that the honors class stood out among the rest (Figure 5). To prove this observation, the three regular experimental classes' normalized gains were compared. There was no significant difference between the regular classes for Units 3, 4, and 9. The only exception was Unit 2. In this unit, the 6th hour

class had particularly low learning gains. This supported the trend shown earlier of the decreasing posttest scores throughout the school day. Some differences that could have caused this were the student check-outs that most often happen at the end of the day or the students' attention spans that seemed to grow shorter as the day went on. Both of these factors could have contributed to the lower scores.

Because of the previous observations, it was apparent that the honors experimental class should be separated from the three regular experimental classes which could be grouped together because of their similar scores. The experimental classes were grouped this way for four more comparisons. The first of which was to again look at the posttest scores for each unit in the hopes that a higher regular experimental sample size might yield different results. The ANOVA of the posttest means between the control group, regular experimental group, and honors experimental group showed the same results. There was a significant difference when the honors class was in the comparison but no significant difference when they were not (Figure 6 &7).

Next the comprehensive posttest means and normalized gains were analyzed to see if pretesting helped with long-term retention and learning cumulative concepts. Any student that did not complete a pretest for each unit was eliminated from these comparisons. The same trend that was obvious in all of the previous comparisons held up in these two analyzations. It appeared that the honors class might have benefitted cumulatively from the pretests, but the regular experimental group did not.

Butler (2010) found that the testing effect could increase learning transfer or the transfer of knowledge to material related to concepts previously tested on in a new context. To see if pretesting increased the transfer of knowledge to related material on the End of Course (EOC) test, an ANOVA was done to compare the control and experimental scores on this test. The honors group had significantly higher scores than the control and regular experimental groups. Pretesting could promote learning transfer in honors class but does not in the regular classes.

In conclusion, pretesting showed no statistically significant effect on the regular classes' unit Edusoft scores. Essentially teaching to the test did not appear to work. Pretesting might have had an effect on the honors class, but further research is needed to say whether this is true. Pretesting did not help the regular classes learn more as evidenced by the normalized gains both for each unit and comprehensively. However the honors class did learn more than the control class, but not necessarily strictly due to pretesting. Pretesting did not increase learning transfer on the EOC test for the regular classes. The honors class did have higher EOC scores but not for certain attributable to pretesting.

In the future, a study such as this would benefit from having an honors control class to determine if pretesting truly had an effect on these students. This group of students seemed to benefit more from pretesting. They spent a longer time taking the pretest and really tried to do their best. The honors class seemed to always want to pay attention to anything that would give them an academic advantage. More than the other classes, the honors students would refer back to material that they remembered from the pretest during lectures. If there was an

honors control class to compare to, it would be more definitive that pretesting benefited honors students.

Another change that could be made for future research would be to shorten the pretests and retention time. The pretests that were used in the present study were on average about thirty questions. In comparison with previous studies, it seemed that dividing the pretests up into smaller sections could be more effective. The test could be divided into three pretests around ten questions each. It could be divided by closely related topics. The shortened tests could be given at the beginning of the hour followed by a lecture on those topics and then the same test given as a posttest at the end of class. By shortening the pretests, students would not be overwhelmed by the amount of new information. The pretests would have a greater chance of improving their retention.

Students tend to respond better to routine. For this fact, giving pretests for every unit rather than just a few might show better results. If the students could get used to the tests as just part of the course, they should be less likely to complain about taking it and maybe realize that focusing on the pretest was in their best interest.

Another factor that seemed to hurt the effects of pretesting was the lack of motivation from the students, especially the regular experimental classes. Bonus points were offered as an incentive to the three highest scores on the pretests in each class. For this sample of students, this did not seem to be the right type of reward. Many of the students sped through the pretest by filling in random guesses without even reading the questions. They did not care whether they received bonus points or not or whether the pretest would let them know what was on

their unit test. From many observations, it appears that these students are motivated more by food than anything, candy in particular. It might be advantageous to offer candy as a reward along with bonus points for high scores on the pretests. This would incite the less academically motivated students to put forth more effort on the pretest, thus possibly improving the results in regular classes.

Many assume that teaching to the test is a quick fix for lagging test scores; it gives teachers an easy way out. The analyses of the results of this study prove that this is an over simplified statement. While pretesting might be effective with some study groups, not all groups of students benefit from this technique. There are many factors that contribute to the ability of pretesting to improve retention. The design of this study did not lend itself to gains solely due to pretesting. The aforementioned modifications would likely produce positive results.

REFERENCES

Agarwal, P. K., McDaniel, M. A., McDermott, K. B., & Roediger, H. L. (2011). Test-Enhanced Learning in the Classroom: Long-Term Improvements From Quizzing. *Journal of Experimental Psychology: Applied*, 382-394.

Bjork, E. L., & Little, J. L. (2011, July 20). Pretesting with Multiple-choice Questions Facilitate Learning. 294-299. Boston, Massachusetts, USA.

Butler, A. C. (2010). Repeated Testing Produces Superior Transfer of Learning Relative to Repeated Studying. *Journal of Experimental Pschology: Learning, Memory, and Cognition*, 1118-1133.

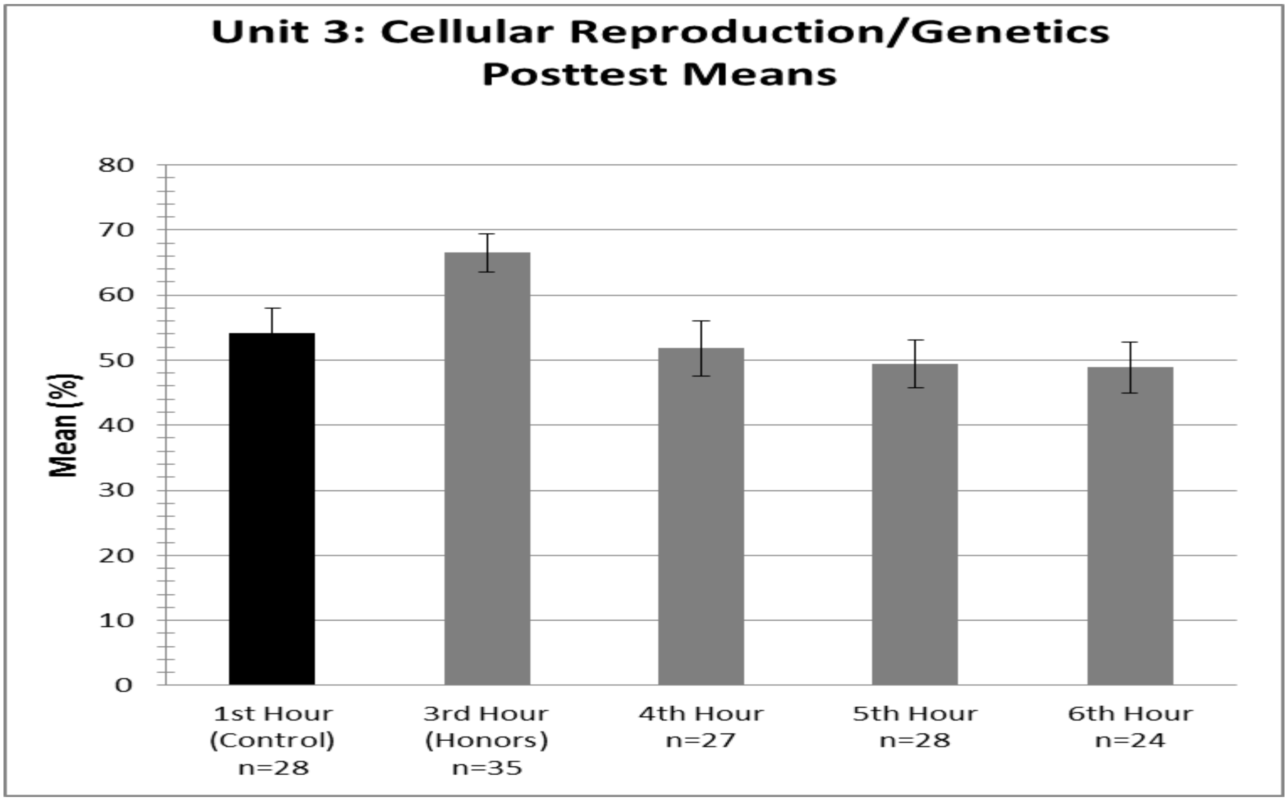
Butler, A. C., & Roediger, H. L. (2008). Feedback Enhances the Postive Effects and Reduces the Negative Effects of Mutilple-choice Testing. *Memory & Cognition*, 604-613.

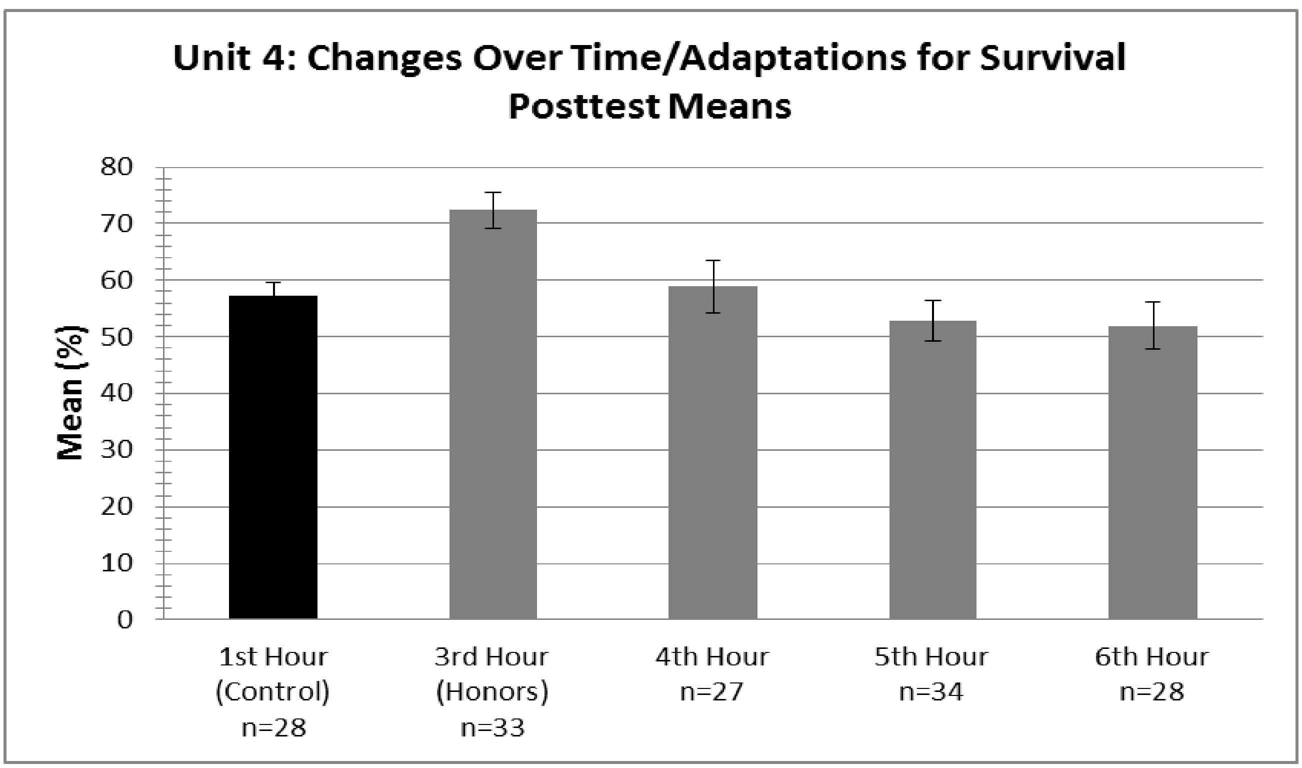
Kao, L. S., Kornell, N., & Richland, L. E. (2009). The Pretesting Effect: Do Unsuccessful Retrieval Attempts Enhance Learning? *Journal of Experimental Pschology: Applied*, 243-257.

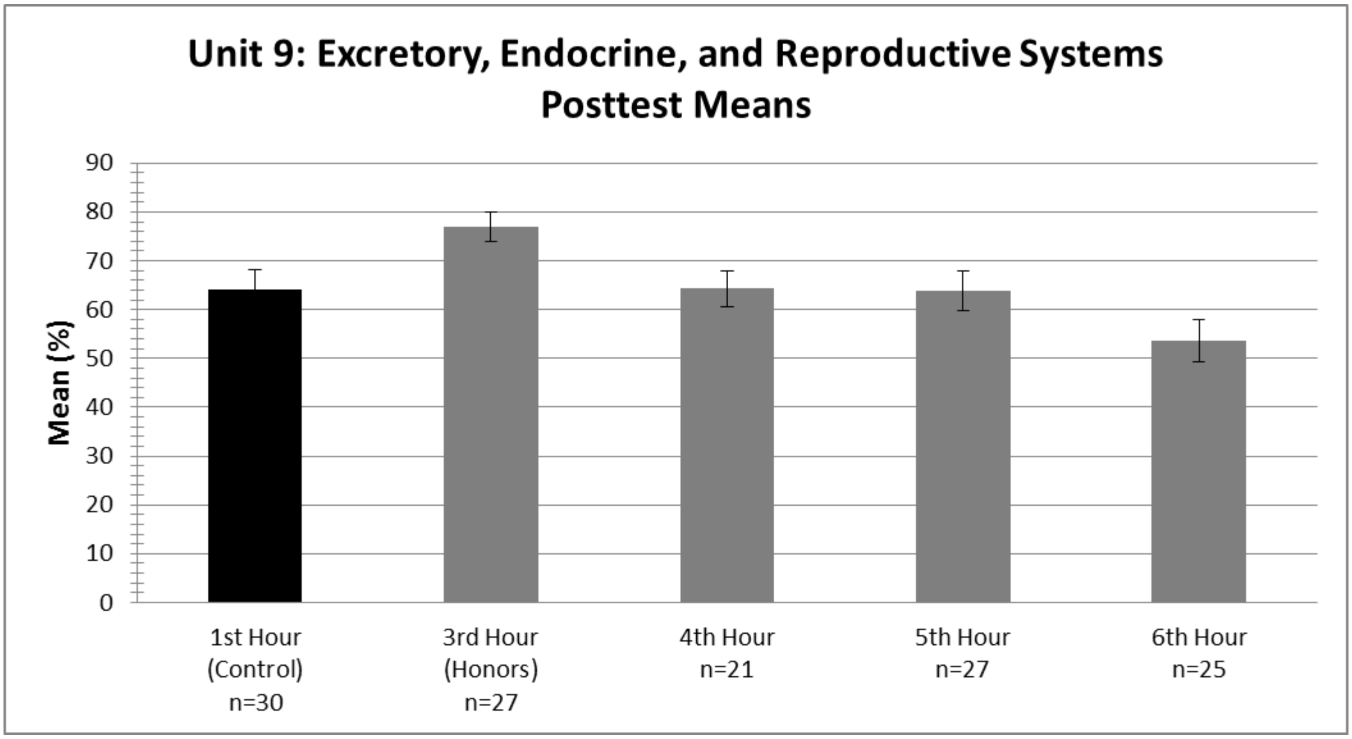
Karpicke, J. D., & Roediger, H. L. (2006). Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention. *Psychological Science*, 249-255.

McGeoch, J. A. (1942). The psychology of human learning: An introduction. New York, NY: Longmans, Green.

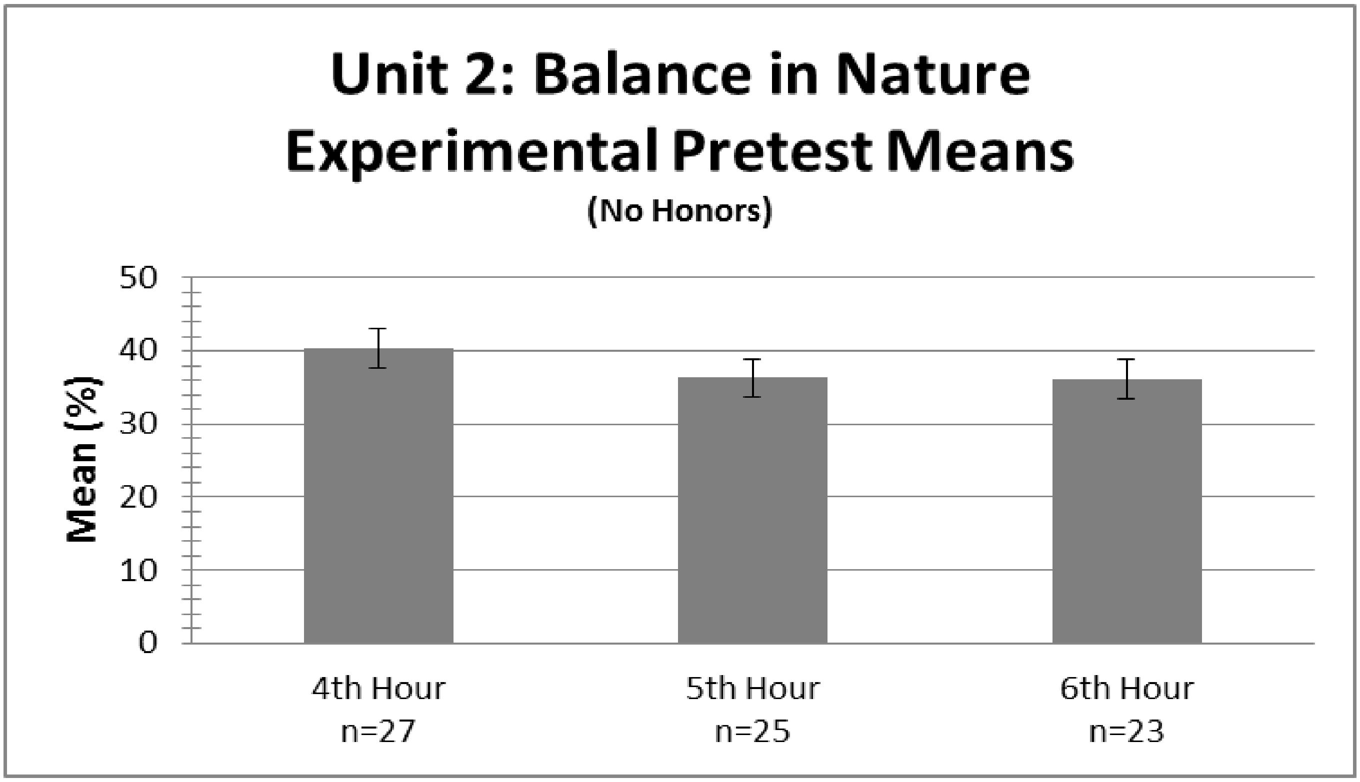
APPENDIX A: BAR GRAPHS OF UNITS 3, 4, & 9 POSTTEST MEANS

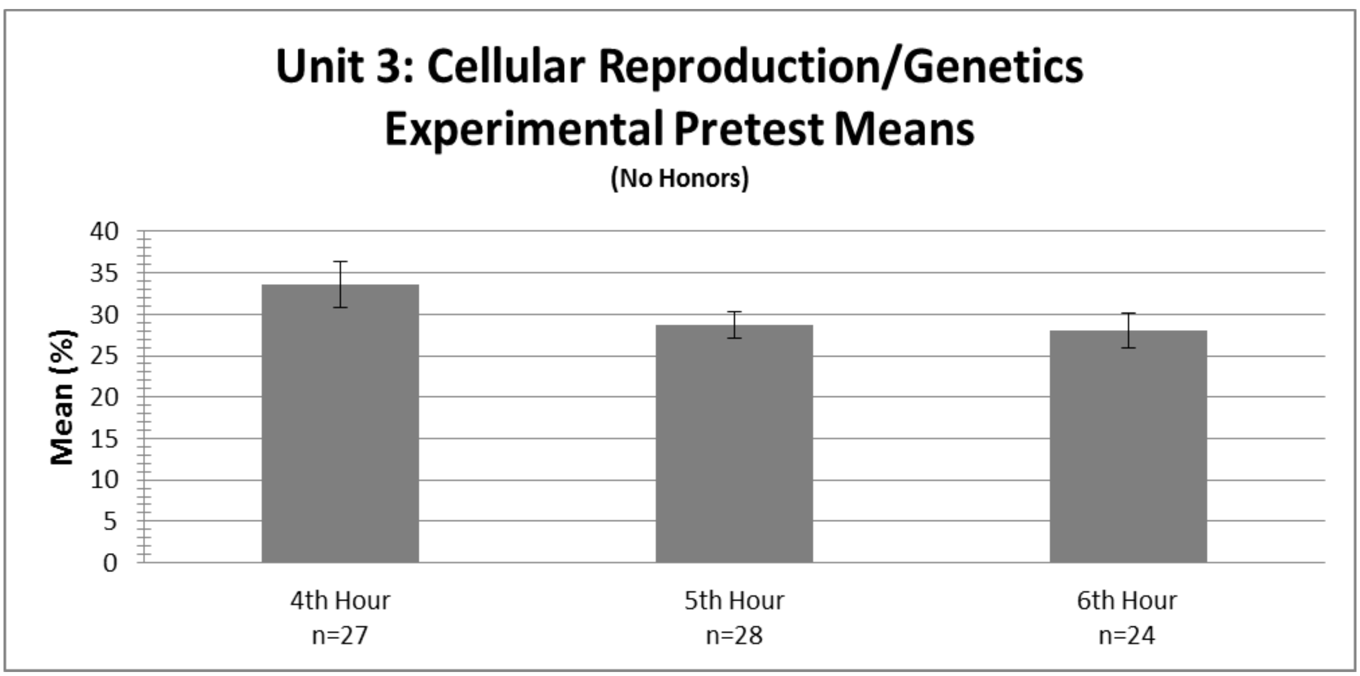


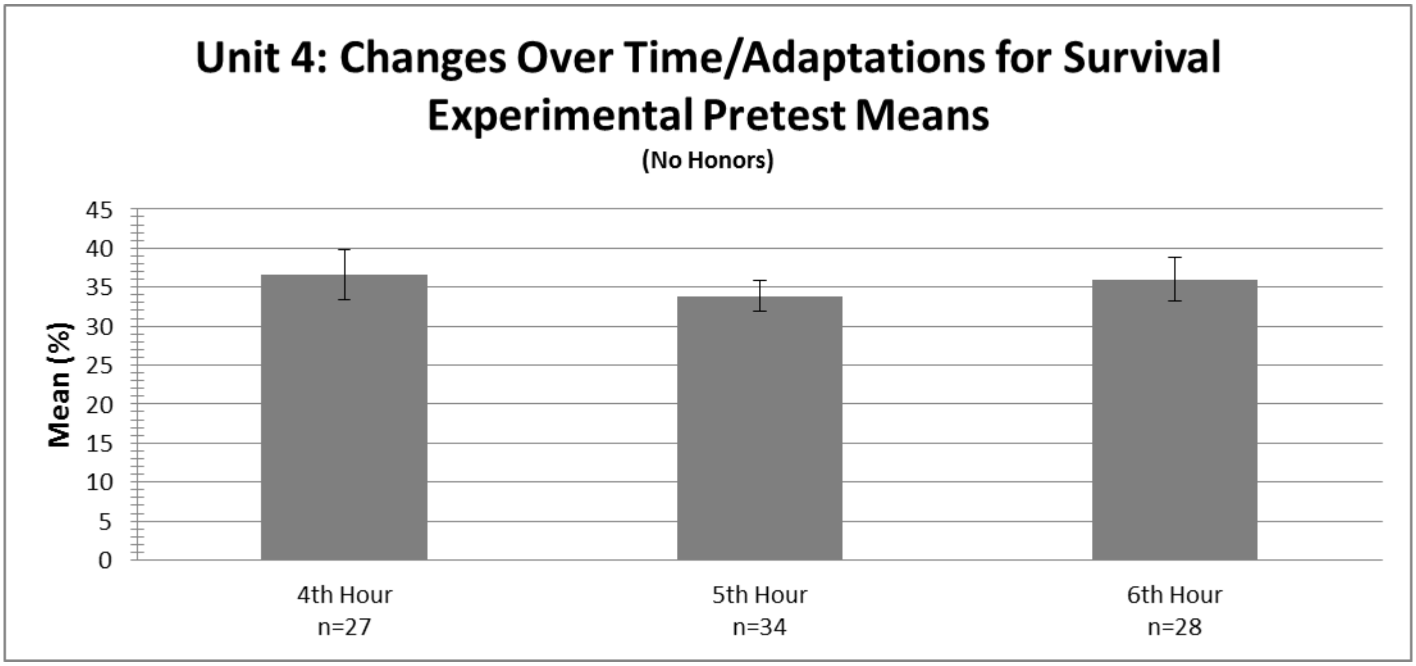


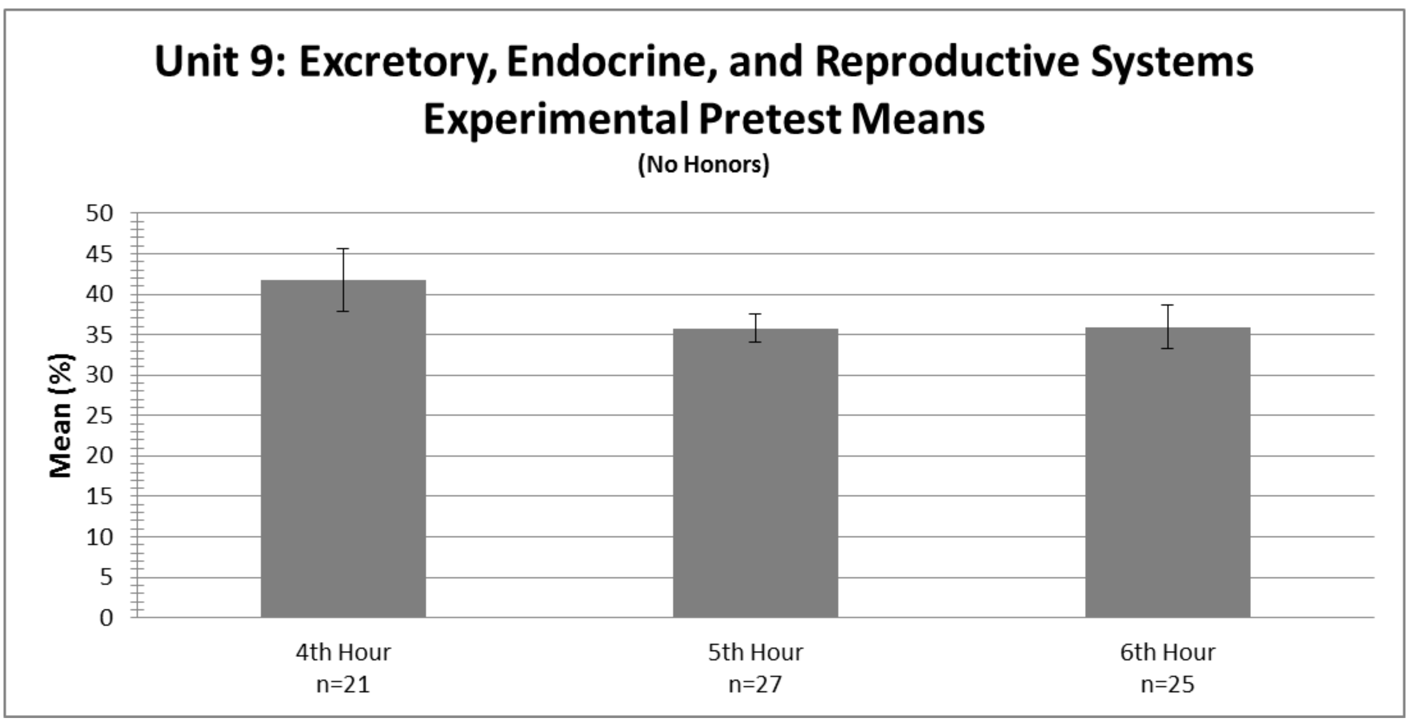

Unit 9: Excretory, Endocrine, and Reproductive Systems
Posttest Means
Mean (%)
90
80
70
60
50
40
30
20
10
0
1st Hour
(Control)
n=30
3rd Hour
(Honors)
n=27
4th Hour
n=21
5th Hour
n=27
6th Hour
n=25

APPENDIX B: BAR GRAPHS OF ALL THE UNITS' EXPERIMENTAL PRETEST MEANS (NO HONORS)

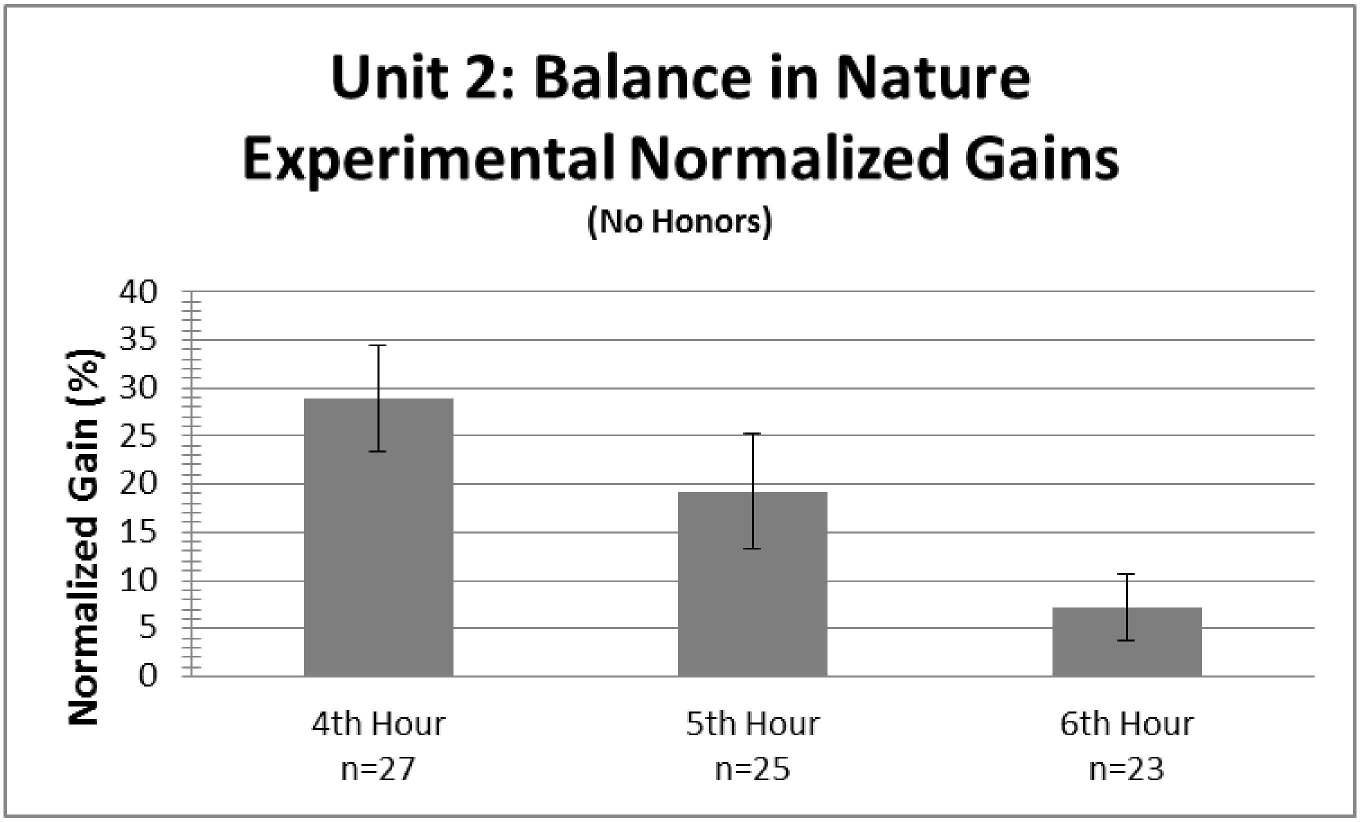


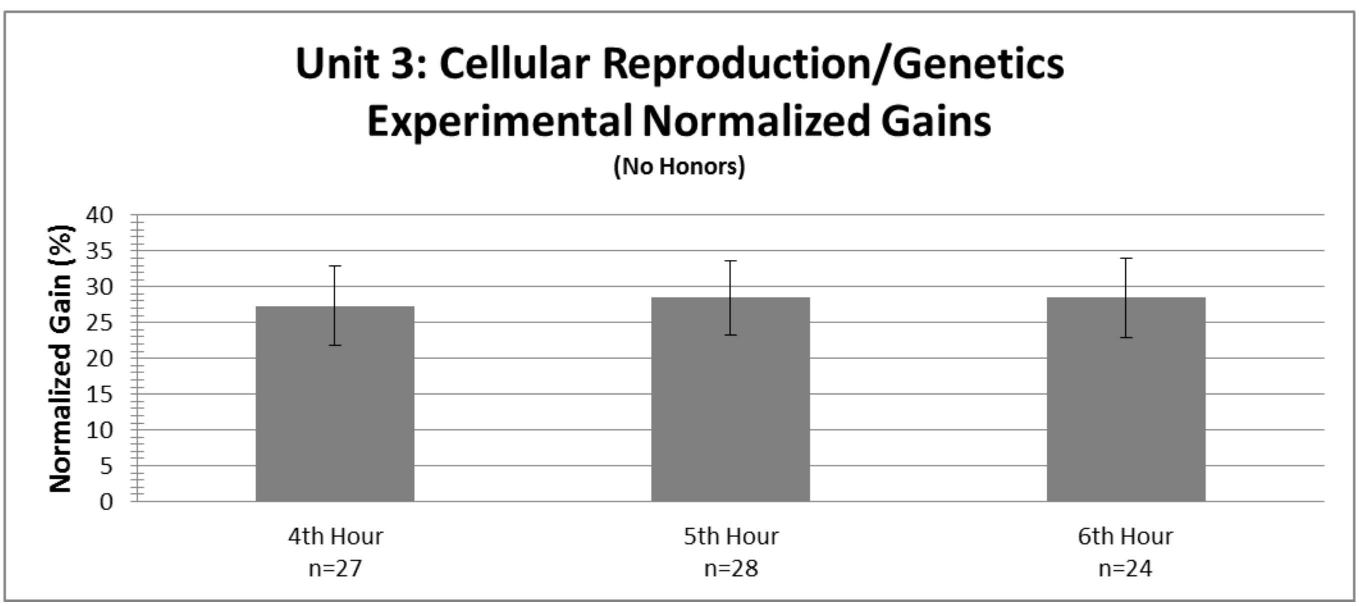


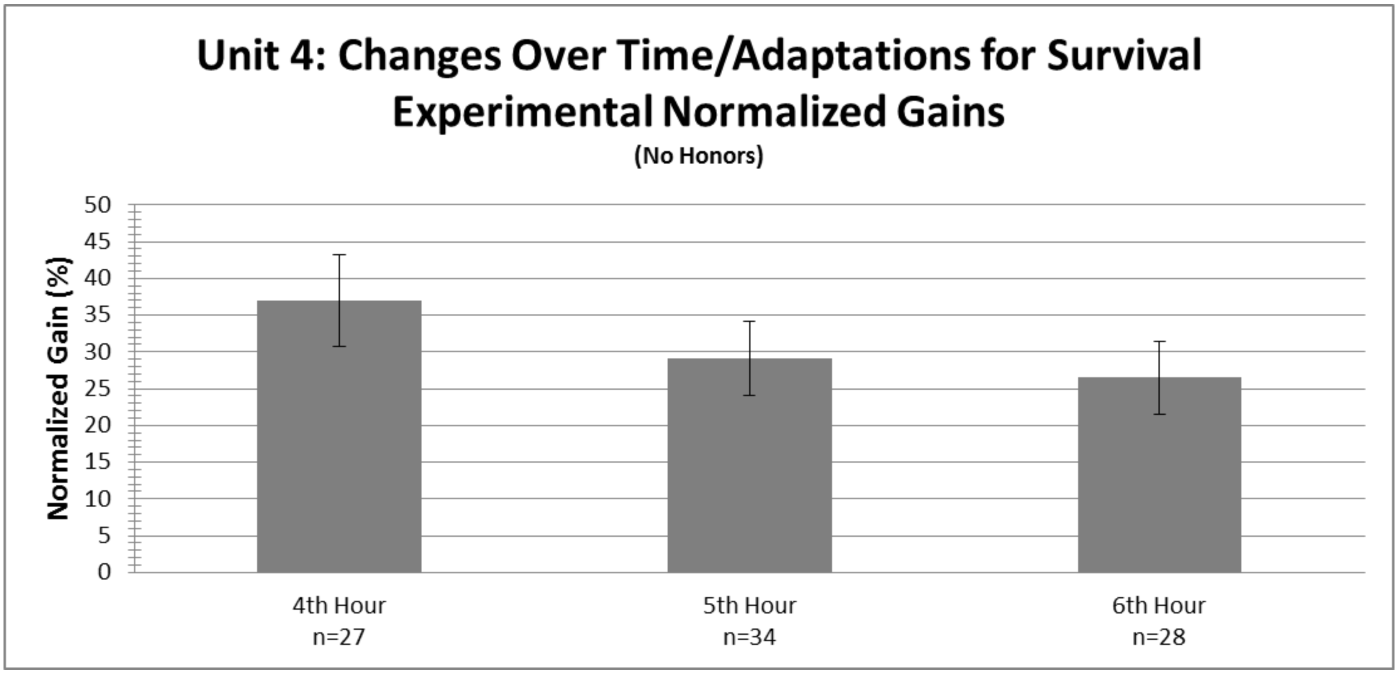

Unit 4: Changes Over Time/Adaptations for Survival
Experimental Pretest Means
(No Honors)
Mean (%)
45
40
35
30
25
20
15
10
5
0
4th Hour
n=27
5th Hour
n=34
6th Hour
n=28

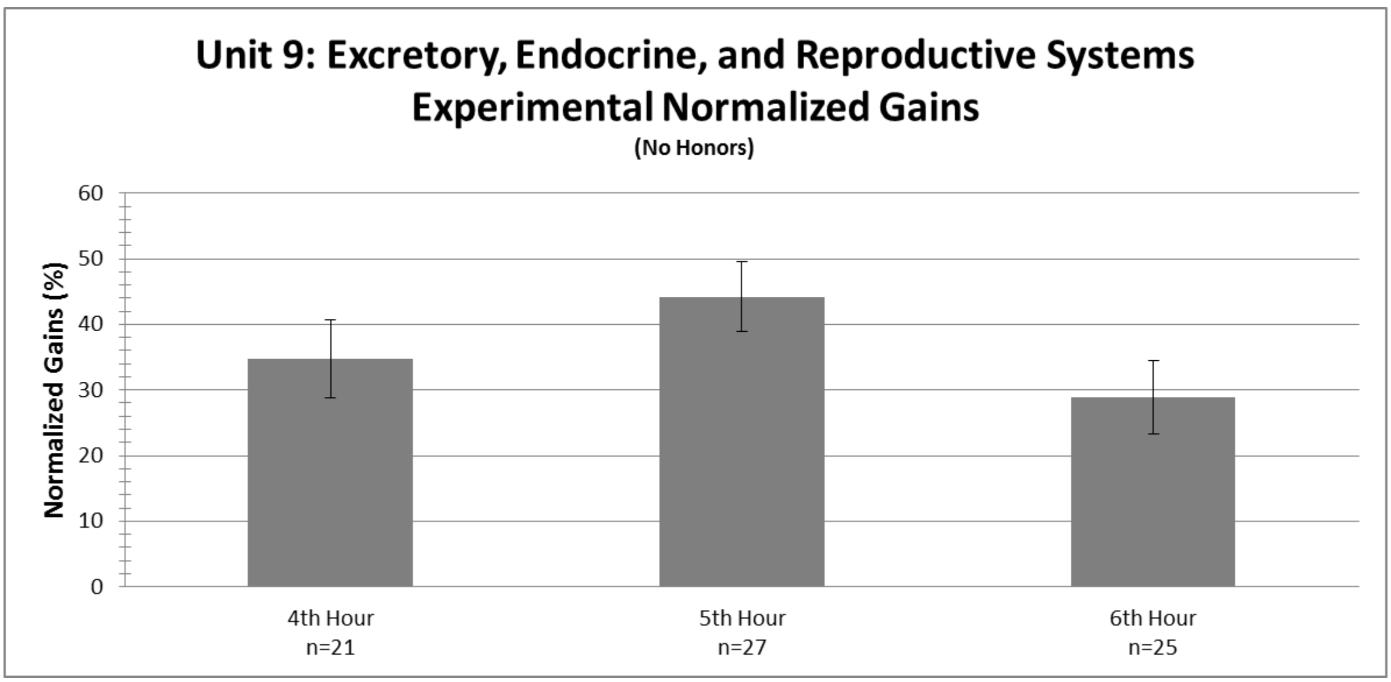

Unit 9: Excretory, Endocrine, and Reproductive Systems
Experimental Pretest Means
(No Honors)
Mean (%)
50
45
40
35
30
25
20
15
10
5
0
4th Hour
n=21
5th Hour
n=27
6th Hour
n=25

APPENDIX C: BAR GRAPHS OF ALL THE UNITS' EXPERIMENTAL NORMALIZED GAINS (NO HONORS)






Unit 4: Changes Over Time/Adaptations for Survival
Experimental Normalized Gains
(No Honors)
Normalized Gain (%)
0
5
10
15
20
25
30
35
40
45
50
4th Hour
n=27
5th Hour
n=34
6th Hour
n=28


Unit 9: Excretory, Endocrine, and Reproductive Systems
Experimental Normalized Gains
(No Honors)
Normalized Gains (%)
0
10
20
30
40
50
60
4th Hour
n=21
5th Hour
n=27
6th Hour
n=25

APPENDIX D: INSTITUTIONAL REVIEW BOARD APPROVAL FORM

Application for Exemption from Institutional Oversight

LSU
Institutional Review Board
Dr. Robert Mathews, Chair
131 David Boyd Hall
Baton Rouge, LA 70803
P: 225.578.8692
F: 225.578.6792
irb@lsu.edu
lsu.edu/irb

Unless qualified as meeting the specific criteria for exemption from Institutional Review Board (IRB) oversight, ALL LSU research/ projects using living humans as 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 PI 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 Screening Committee. Members of this committee can be found at http://www.lsu.edu/screeningmembers.shtml

– 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)
 (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 material.
 (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://phrp.nihtaining.com/users/login.php.)
 (F) IRB Security of Data Agreement: (http://www.lsu.edu/irb/IRB%20Security%20of%20Data.pdf)

1) Principal Investigator: Dr. Dana Browne **Rank:** Professor

Dept: Physics **Ph:** 578--6843 **E-mail:** phowne@lsu.edu

2) Co Investigator(s): please include department, rank, phone and e-mail for each

Mary Catherine Johnson - graduate student, 337-304-8444, mchammer305@hotmail.com

IRB# E5604 LSU Proposal #
Complete Application
Human Subjects Training

3) Project Title: Pretesting in Science: Effect on Unit Test Scores

Study Exempted By:
Dr. Robert C. Mathews, Chairman
Institutional Review Board
Louisiana State University
203 B-1 David Boyd Hall
225-578-8692 | www.lsu.edu/irb
Exemption Expires: 8-17-2014

4) Proposal? (yes or no) 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 Biology I students
 *Circle any **"vulnerable populations" to be used:** (children <18; the mentally impaired, pregnant women, the ages, other). Projects with incarcerated persons cannot be exempted.

6) PI Signature [signature] **Date** 30 JUNE 2011 (no per signatures)

** **I certify my responses are accurate and complete.** If the project scope or design is later changes, I will resubmit 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 1

Reviewer Mathews **Signature** [signature] **Date** 8/18/11

VITA

Mary Catherine Johnson was born in Lake Charles, Louisiana in March of 1985. She graduated in 2003 from Alfred M. Barbe High School in the Calcasieu Parish school district. Following graduation from high school, Mary Catherine entered Louisiana State University. She graduated with honors in December of 2007 with a Bachelor of Arts Degree in marketing. She entered Louisiana State University Graduate School in May of 2010. She is presently employed as a Biology I teacher by Broadmoor High School in the East Baton Rouge Parish school district.