Assessing the potential to improve the basic quantitative skills of undergraduate biology students

Liliane E. (Liliane Elyse) Dethier
Western Washington University
ASSESSING THE POTENTIAL TO IMPROVE THE BASIC QUANTITATIVE SKILLS OF UNDERGRADUATE BIOLOGY STUDENTS

By

Liliane Elyse Dethier

Accepted in Partial Completion

Of the Requirements for the Degree

Master of Science

Kathleen L. Kitto, Dean of the Graduate School

ADVISORY COMMITTEE

Chair, Dr. Deborah Donovan

Dr. Emily Borda

Dr. Merrill Peterson
MASTER’S THESIS

In presenting this thesis in partial fulfillment of the requirements for a master’s degree at Western Washington University, I grant to Western Washington University the non-exclusive royalty-free right to archive, reproduce, distribute, and display the thesis in any and all forms, including electronic format, via any digital library mechanisms maintained by WWU.

I represent and warrant this is my original work, and does not infringe or violate any rights of others. I warrant that I have obtained written permissions from the owner of any third party copyrighted material included in these files.

I acknowledge that I retain ownership rights to the copyright of this work, including but not limited to the right to use all or part of this work in future works, such as articles or books.

Library users are granted permission for individual, research and non-commercial reproduction of this work for educational purposes only. Any further digital posting of this document requires specific permission from the author.

Any copying or publication of this thesis for commercial purposes, or for financial gain, is not allowed without my written permission.

Liliane Elyse Dethier

February 18, 2014
ASSESSING THE POTENTIAL TO IMPROVE THE BASIC QUANTITATIVE SKILLS OF UNDERGRADUATE BIOLOGY STUDENTS

A Thesis
Presented to
The Faculty of
Western Washington University

In Partial Fulfillment
Of the Requirements for the Degree
Master of Science

By
Liliane Elyse Dethier
January 2014
ABSTRACT

Collaboration between biology and mathematics has led to important advances in both fields. Recent advances in technology have made biology more quantitative, which makes interdisciplinary work increasingly important. In contrast, biology education has historically been less quantitative than the other sciences despite the need for biology education to prepare students to be scientifically literate and to do research. Biology educators must prepare students with appropriate skills to understand, communicate, collaborate, and compete for funding in this increasingly quantitative field. There is a growing awareness at many universities that undergraduate biology students frequently lack basic quantitative skills. This project followed models from other institutions to modify an introductory-level biology course with the goal of improving the basic quantitative skills of students in that course. The faculty identified basic quantitative skills that they considered the most important for undergraduate biology students in the core major courses: (1) using and manipulating fractions, proportions, ratios, and percentages; (2) reading and analyzing data, including graphical presentation and interpretation; (3) using and manipulating logarithms; (4) using and manipulating exponents; and (5) translating biological problems into quantitative problems and vice versa.

I worked with students in Biology 205, the second quarter of a three-course introductory series. Biology 205 is entitled Introduction to Cell and Molecular Biology. I created an assessment to evaluate these five skills in Biology 205 students. Based on the assessment results I created three activities targeted to improve the areas of deficiency within the identified skills. I administered these treatment activities to half of Biology 205 lab sections during the winter and spring quarters 2013. I also administered control activities to the other half of Biology 205 sections to control for the potentially confounding effects of (1) skills naturally improving over the quarter and (2) skills improving due to increased effort time. I administered the assessment as a pre-test and post-test before and after the activities to evaluate if the activities improved students’ basic quantitative skills. To answer my research questions, I used statistical analyses that, to my knowledge, have not been used for biology education research before: generalized linear mixed models (GLMMs). Treatment was not included in the best-fit model predicting whether or not students answered questions correctly. However, the treatment group appeared to perform better on the post-test while the control group scored similarly on the pre- and post-tests. I also administered a population demographics survey to identify the important factors that predicted students’ incoming basic quantitative skills, and how those skills changed throughout the quarter. The important factors were: highest college-level mathematics course completed, transfer history, major, year in school and gender. I offer recommendations for future work that should be more effective at improving the basic quantitative skills of biology undergraduates.
ACKNOWLEDGEMENTS

I would first like to thank my advisor, Dr. Deborah Donovan, and other thesis committee members, Drs. Emily Borda and Merrill Peterson, for their seemingly endless guidance and feedback in this project. This project was truly interdisciplinary; I relied tremendously on the assistance of many individuals in other departments at this institution and in other institutions. Without their help this project would have been impossible. I am very grateful to Dr. Katerina Thompson (University of Maryland), who corresponded with me on multiple aspects of the project, and graciously offered the assessment she developed to help me develop the assessment for this project. Kelly Matthews, an instructor at the University of Queensland, also shared some tools that she developed for a similar purpose. I would like to thank the Biology 205 students who participated in this project, and the Biology 205 graduate teaching assistants and professors who also supported the project.

I would also like to thank the following people who helped me with validity and reliability statistics: Dr. Jay Teachman (Western Washington University), Dr. Sara Kim (University of Washington), and Sarah Sleddy (University of Washington). Thank you to the team of experts that helped validate the assessment: Dr. Kimberly Markworth (Western Washington University), my committee members, and my dad: Dr. Jean-Louis Dethier. Another large thanks goes to my dad for writing a computer program to download and organize the assessment results. Thanks to Lily Baxter and Arin Anthon, undergraduate students who helped with data entry. Enormous thanks goes to Dr. Benjamin Miner (Western Washington University) who helped me analyze my data and learn how to use generalized linear mixed effect models (GLMMs). Last, but certainly not least, I thank Western’s Biology Department and the NASA Washington Space Grant Consortium for funding this research. I received approval for this project from the University’s Human Subjects Review Committee and conducted the project under protocol number 13-011.
# TABLE OF CONTENTS

ABSTRACT ............................................................................................................................... iv

ACKNOWLEDGEMENTS .......................................................................................................... v

TABLE OF CONTENTS ........................................................................................................... vi

LIST OF FIGURES .................................................................................................................. vii

LIST OF TABLES ..................................................................................................................... viii

INTRODUCTION ...................................................................................................................... 1
  Mathematics and biology ....................................................................................................... 1
  Mathematics and undergraduate biology education ............................................................... 2
  Study system ........................................................................................................................ 6
  Summary of experimental approach .................................................................................... 7
  Research objectives ............................................................................................................ 8
    Improving quantitative skills ............................................................................................ 9
    Population demographics and quantitative skills ............................................................ 9

METHODS .............................................................................................................................. 11
  Identifying basic quantitative skills to target ....................................................................... 11
  Developing the quantitative skills assessment .................................................................. 11
  Assessment validity and reliability ..................................................................................... 16
  Targeting quantitative skills improvement ......................................................................... 17
  Spring quarter changes ...................................................................................................... 22
  Statistical analyses ............................................................................................................ 23

RESULTS ................................................................................................................................ 32
  Improving quantitative skills (research objectives one and two) ......................................... 32
  Activity results .................................................................................................................. 37
  Population demographics and quantitative skills (research objectives three and four) .... 41
    Pre-test score (research objective three) .......................................................................... 41
    Difference between pre- and post- test scores (research objective four) ......................... 44

DISCUSSION .......................................................................................................................... 54
  Improving quantitative skills ............................................................................................. 54
  Ways to improve the treatment activities ........................................................................... 56
  Ways to improve the treatment design .............................................................................. 58
  An additional skill to include: statistics ............................................................................. 61
  Population demographics and quantitative skills analysis ............................................... 62
  Study contributions ........................................................................................................... 65
  Future work ....................................................................................................................... 66

REFERENCES ......................................................................................................................... 69

APPENDIX A ........................................................................................................................... 74

APPENDIX B ........................................................................................................................... 75

APPENDIX C ........................................................................................................................... 80

APPENDIX D ........................................................................................................................... 85
APPENDIX E ........................................................................................................................................... 88
APPENDIX F ........................................................................................................................................... 93
APPENDIX G ........................................................................................................................................... 95
APPENDIX H ........................................................................................................................................... 98
APPENDIX I ........................................................................................................................................... 101
APPENDIX J ........................................................................................................................................... 103
LIST OF FIGURES

Figure 1. Pre- and post-test scores of students divided by achievement-level for purely quantitative questions compared to those requiring quantitative skills to answer biological questions. This figure includes students in both the control and treatment groups. Horizontal lines represent the median score, and box limits represent the first quartile (Q1) and third quartile (Q3). Dashed whiskers indicate the most extreme data points within 1.5*IQR where IQR is the interquartile range: Q3-Q1. The points represent outliers which are outside 1.5*IQR. 34

Figure 2. Overall pre- and post-test scores for control and treatment groups. Horizontal lines represent the median score, and box limits represent the first quartile (Q1) and third quartile (Q3). Dashed whiskers indicate the most extreme data points within 1.5*IQR where IQR is the interquartile range: Q3-Q1. The points represent outliers which are outside 1.5*IQR. 35

Figure 3. Comparison of pre-test scores of control and treatment groups for each quarter and corresponding t-test results. Horizontal lines represent the median score, and box limits represent the first quartile (Q1) and third quartile (Q3). Dashed whiskers indicate the most extreme data points within 1.5*IQR where IQR is the interquartile range: Q3-Q1. The points represent outliers which are outside 1.5*IQR. Numbers in the boxes represent sample sizes. 36

Figure 4. Pre and post-test scores of students in both treatment groups divided by achievement-level for the winter quarter only. Horizontal lines represent the median score, and box limits represent the first quartile (Q1) and third quartile (Q3). Dashed whiskers indicate the most extreme data points within 1.5*IQR where IQR is the interquartile range: Q3-Q1. The points represent outliers which are outside 1.5*IQR. 39

Figure 5. Graded subset of activities from both quarters. Horizontal lines represent the median score, and box limits represent the first quartile (Q1) and third quartile (Q3). Dashed whiskers indicate the most extreme data points within 1.5*IQR where IQR is the interquartile range: Q3-Q1. The points represent outliers which are outside 1.5*IQR. T-tests comparing the winter and spring quarter grades revealed no significant differences between the quarters: (Exponents: $t=0.53$, df=36.2, $p=0.60$; Logs: $t=0.96$, df=24.7, $p=0.35$; Graphing: $t=-1.05$, df=37.8, $p=0.30$). 40

Figure 6. Pre- and post-test scores of quantitative skill areas for each treatment group. Horizontal lines represent the median score, and box limits represent the first quartile (Q1) and third quartile (Q3). Dashed whiskers indicate the most extreme data points within 1.5*IQR where IQR is the interquartile range: Q3-Q1. The points represent outliers which are outside 1.5*IQR. 42

Figure 7. Pre- and post-test scores of students with the highest college-level mathematics class they have completed, and the change between pre and post-test scores for these groups, represented by normalized gains. Horizontal lines represent the median score, and box limits represent the first quartile (Q1) and third quartile (Q3). Dashed whiskers indicate the
most extreme data points within 1.5*IQR where IQR is the interquartile range: Q3-Q1. The points represent outliers which are outside 1.5*IQR. Numbers in boxes represent sample sizes.

Figure 8. Pre- and post-test scores of students with their transfer history, and the change between pre and post-test scores for these groups, represented by normalized gains [(post-pre)/(16-pre)]. Horizontal lines represent the median score, and box limits represent the first quartile (Q1) and third quartile (Q3). Dashed whiskers indicate the most extreme data points within 1.5*IQR where IQR is the interquartile range: Q3-Q1. The points represent outliers which are outside 1.5*IQR. Numbers in boxes represent sample sizes.

Figure 9. Pre- and post-test scores of students arranged by major groups, and the change between pre and post-test scores for these groups, represented by normalized gains. Horizontal lines represent the median score, and box limits represent the first quartile (Q1) and third quartile (Q3). Dashed whiskers indicate the most extreme data points within 1.5*IQR where IQR is the interquartile range: Q3-Q1. The points represent outliers which are outside 1.5*IQR. Numbers in boxes represent sample sizes.

Figure 10. Pre- and post-test scores of students with their year, and the change between pre and post-test scores for these groups, represented by normalized gains [(post-pre)/(16-pre)]. Horizontal lines represent the median score, and box limits represent the first quartile (Q1) and third quartile (Q3). Dashed whiskers indicate the most extreme data points within 1.5*IQR where IQR is the interquartile range: Q3-Q1. The points represent outliers which are outside 1.5*IQR. Numbers in boxes represent sample sizes.

Figure 11. Pre- and post- test scores for males and females. Horizontal lines represent the median score, and box limits represent the first quartile (Q1) and third quartile (Q3). Dashed whiskers indicate the most extreme data points within 1.5*IQR where IQR is the interquartile range: Q3-Q1. The points represent outliers which are outside 1.5*IQR. Numbers in boxes represent sample sizes.
LIST OF TABLES

Table 1. Results from the faculty survey (N=19) asking them to identify the three most critical basic quantitative skills for biology undergraduates. Asterisks indicate skills selected for this study. ......................................................... 12

Table 2. Project timeline ........................................................................................................................................ 14

Table 3. Sample sizes of all participant groups. Includes only full study participants, defined as students who completed the pre-test, all three activities, and post-test. ................................. 15

Table 4. Cronbach's alpha results from all assessment administrations. $\alpha > 0.60$ indicates a reliable assessment for this study.................................................................................................................. 18

Table 5. Second pilot results used to determine the three skill areas to target with the quantitative activities. SD represents Standard Deviation.......................................................................................... 19

Table 6. Lecture and lab sections assigned to each treatment group for both quarters. .............................. 21

Table 7. Full model for each set of GLMM analyses. ................................................................. 28

Table 8. Top five models evaluating the factors that predicted whether students answered questions correctly or incorrectly on the pre- and post-tests. The gray box highlights the best-fit model, and the top five other models are listed in order from best to worst fit; lowest to highest Akaike Information Criterion (AIC) value. All models tested are included in Appendix J. df represents degrees of freedom for each model, delta AIC is the difference between the given model’s AIC score and the best model’s AIC score. Evidence ratios compare the best-fit model to the other top five models; these values indicate how much more likely the best-fit model is than the given model................................................................. 33

Table 9. Top five models evaluating the factors that predicted whether students answered questions correctly or incorrectly on the pre- and post-tests during the winter quarter only. The gray box highlights the three models that are similarly “best-fit” ranked from lowest to highest Akaike Information Criterion (AIC). df represents degrees of freedom for each model, delta AIC is the difference between the given model’s AIC score and the best model’s AIC score. Evidence ratios compare the best-fit model to the other top five models; these values indicate how much more likely the best-fit model is than the given model..... 38

Table 10. Top five models for population demographics and quantitative skills analysis evaluating the factors that predicted pre-test score. The gray box highlights the four models that are similarly “best-fit” ranked from lowest to highest Akaike Information Criterion (AIC). All models tested are included in Appendix J. df represents degrees of freedom for each model, delta AIC is the difference between the given model’s AIC score and the best model’s AIC score. Evidence ratios compare the best-fit model to the other top five models; these values indicate how much more likely the best-fit model is than the given model..... 43
Table 11. Top five models for population demographics and quantitative skills analysis evaluating the factors that predict differences between pre- and post-test scores. The gray box highlights the top three models averaged together to make the best-fit model. The relative importance of each variable in the averaged model is included at the bottom of the table; higher numbers indicate more importance. The top four other models are listed in order from best to worst fit; lowest to highest Akaike Information Criterion (AIC) value. All models tested are included in Appendix J. df represents degrees of freedom for each model, delta AIC is the difference between the given model’s AIC score and the best model’s AIC score. Evidence ratios compare the best-fit model to the other top five models; these values indicate how much more likely the best-fit model is than the given model. 

Table 12. Contingency table comparing student responses to opinion questions. I used G-tests of independence to make comparisons between the pre- and post- test responses for each treatment group for the first two questions, and between the treatment and control group for the third question because it was only on the post-test. There were two degrees of freedom for each analysis and I adjusted alpha to 0.01 to correct for multiple comparisons. Significant results are highlighted with the gray box.

Table 13. All models for hypothesis testing analysis evaluating factors that predict whether or not a student answered a question correctly. The best model is highlighted with the gray box. df represents degrees of freedom for each model and ΔAIC is the difference between the given model’s AIC score and the best model’s AIC score. Models are listed in order from best to worst fit (lowest to highest AIC value).

Table 14. All models for hypothesis testing analysis evaluating factors that predict whether or not a student answered a question correctly during the winter quarter. The gray box highlights the top three models averaged together to make the best-fit model. df represents degrees of freedom for each model and ΔAIC is the difference between the given model’s AIC score and the best model’s AIC score. Models are listed in order from best to worst fit (lowest to highest AIC value).

Table 15. All models for population demographics and quantitative skills analysis evaluating factors that predict pre-test score. The gray box highlights the four models averaged together to make the best-fit model. df represents degrees of freedom for each model and ΔAIC is the difference between the given model’s AIC score and the best model’s AIC score. Models are listed in order from best to worst fit (lowest to highest AIC value).

Table 16. All models for population demographics and quantitative skills analysis evaluating factors that predict changes between the pre- and post-test scores. The gray box highlights the two models averaged together to make the best-fit model. df represents degrees of freedom for each model and ΔAIC is the difference between the given model’s AIC score and the best model’s AIC score. Models are listed in order from best to worst fit (lowest to highest AIC value).
INTRODUCTION

Mathematics and biology

Collaboration between biologists and mathematicians has been essential for advancing both fields (Cohen 2004; Wortman 2008). For example, the foundational principles of genetics resulted from simple mathematical observations; Gregor Mendel tracked pea plant traits over multiple generations to infer the laws of inheritance (Mendel 1865; Cohen 2004). Most medical equipment, including x-ray machines, CT scanners, and MRI machines have been developed using mathematics (Glydon 1995; Cohen 2004). Biological problems have also driven the development of new mathematical techniques (Cohen 2004). For example, the famed biologist and mathematician Sir Ronald Fisher developed Analysis of Variance, a family of statistical analyses that are commonly used to answer biology research questions today (Whitlock 2013).

Though mathematicians and biologists have collaborated in the past, interdisciplinary work is becoming even more important because biology research is becoming increasingly quantitative. New technologies, often created using mathematics, allow biology researchers to collect large datasets at multiple biological levels (NRC 2005; Wortman 2008). These large datasets become overwhelming and incomprehensible without using statistics and mathematical modeling to draw meaning from them (Bloom et al. 2008). For example, ecologists often use data logging instruments to continuously record abiotic features of the environment, such as temperature, and then use statistics and modeling to interpret global patterns from these datasets (Bloom et al. 2008; Wolkovich et al. 2012). Molecular biology researchers can identify potentially cancer-causing genetic mutations in large genomic datasets with the mathematical modeling tools of bioinformatics (Wortman 2008). Thus, using quantitative skills to interpret large biological datasets is crucial in understanding complex issues from climate change to
healthcare. Biologists recognize that the quantitative nature of their field is increasing and many believe that questions requiring both mathematics and biology methods will be at the root of the greatest scientific discoveries in this century (Steen 2005). In 2013 at least two conferences were held in the United States focused on specific topics in Quantitative Biology (first annual winter q-bio meeting, Hawaii February 2013; seventh q-bio conference, New Mexico August 2013). To continue making discoveries at the forefront of biology, new researchers must be equipped with the skills to communicate and collaborate in this increasingly quantitative field.

**Mathematics and undergraduate biology education**

Despite important collaborations between mathematics and biology, undergraduate biology education has historically been less quantitative than the other sciences, and has struggled to keep up with the rapidly changing field of biology research (Steen 2005; Depelteau et al. 2010;). Undergraduate biology education should produce scientifically literate adults who will be able to understand the results of research on biologically related problems facing the entire world (AAAS 2011). It should also enable students to become medical professionals and/or researchers who help address these problems (NRC 2003). All students who graduate with a bachelor’s degree in biology need the appropriate skills to understand and participate in this increasingly quantitative field (NRC 2003; AAAS 2011). In 2003 a division of the National Research Council published the influential document *BIO2010*, which highlighted the need to improve biology undergraduate education. Among other recommendations, the document counseled institutions to improve the quantitative components of biology undergraduate curricula especially at the introductory level (NRC 2003; Depelteau et al. 2010). This document and others
have spurred institutions to search for effective ways to produce more mathematically literate biology students (NRC 2003; Gross 2004).

*BIO2010* included specific curricular recommendations for institutions to work towards achieving this goal. It recommended that institutions require mathematics and physical science courses for all biology undergraduates, and that biology courses incorporate quantitative content from these other courses (NRC 2003). The document also recommended that institutions modify existing courses and research labs, or create new ones, that are interdisciplinary between mathematics and biology departments (NRC 2003). With this model students from mathematics and biology would work together, learn from each other, and make contributions to both fields. Collaboration in courses and labs at the undergraduate level might also help make students aware that there is a great potential for jobs in overlapping areas of these two fields (NRC 2003).

*BIO2010* also made several recommendations to institutions about how they can improve the quantitative skills of their biology undergraduates. Likely because biology undergraduate education has traditionally been less quantitative than the other sciences, some biology faculty have weaker quantitative skills than faculty in the other sciences (Katz 2003; Hoy 2004; Miller & Walston 2010). Biology professors also have probably not received training in how to effectively teach quantitative skills (Katz 2003). Many mathematicians feel similarly about biology so they do not see the biological applications of what they teach or use biological examples in their mathematics classes (Katz 2003). To overcome these challenges, *BIO2010* encouraged the creation of workshops for biology faculty to improve their own quantitative skills and learn how to teach these skills effectively (Katz 2003; NRC 2003). Some such workshops do exist, but they are limited and not always accessible because another challenge facing faculty is that many institutions do not reward high-quality or innovative teaching (AAAS 2011).
Consequently many faculty lack the resources and motivation to take advantage of opportunities to improve their teaching, even when those limited opportunities are available (Hoy 2004). *BIO2010* asked institutions to reward biology faculty for strong and innovative teaching, especially with the goal of improving students’ quantitative skills (NRC 2003).

In response to *BIO2010*, some institutions have worked to improve the quantitative skills of their undergraduate biology students following several different strategies: (1) creating mathematics-biology research programs, (2) redesigning introductory-level biology courses, and (3) adding supplementary quantitative material to biology courses. The University of Tennessee and Truman State University have responded to *BIO2010* and addressed their students’ deficit of quantitative skills by introducing students to interdisciplinary research earlier with undergraduate mathematics-biology research programs (Duncan et al. 2010; Miller & Walston 2010). The programs at both schools are similar, and have both been successful. They accepted elite students from the biology and mathematics departments into research labs to work on projects with two faculty mentors; one from biology and one from mathematics or computer science. Students attended lectures about the background material for their projects and collaborated with students from the other field to collect data, analyze results, and present their findings. Student reviews of both programs were positive, and many students went on to graduate programs or research positions (Duncan et al. 2010; Miller & Walston 2010). This strategy was successful, is most directly applicable to students pursuing research, and targeted improving advanced quantitative skills in high-achieving students.

Several other institutions have responded to *BIO2010* by attempting to improve the basic quantitative skills of all their biology students, within the context of biology. Montana State University (MSU) and East Tennessee State University (ETSU) both redesigned their
introductory-level biology courses to better integrate quantitative concepts. MSU added the requirement of a pre- or co-requisite statistics course with the majors’ introductory biology course and added statistical components to inquiry-based labs (Metz 2008). Researchers found that students’ statistics skills improved, and students retained the material even a full year after the course (Metz 2008). ETSU developed a new biology-mathematics integrated curriculum to replace the existing series of introductory courses (Depelteau et al. 2010). Biology and mathematics faculty team-taught the two-week labs, which included data collection and analysis. The team-teaching gave students access to both fields of knowledge, and helped the faculty become more comfortable with the subject material from the others’ field (Depelteau et al. 2010). Pilot results convinced this institution to implement the newly developed courses (Depelteau et al. 2010). Like the strategy of interdisciplinary research programs, this strategy was successful at improving students’ quantitative skills. It also targeted all biology students, and emphasized basic quantitative skills.

While many institutions might wish to develop interdisciplinary mathematics-biology research programs and/or redesign their introductory-course curricula, faculty time and resources are frequently limited. At the University of Maryland (UMD), a research team added supplementary quantitative material to an existing introductory-level course. Thompson et al. (2010) developed MathBench Biology Modules, which are interactive online activities designed to instruct students in basic quantitative skills within a biological context. In Thompson and colleagues’ (2010) study evaluating the efficacy of this strategy, students worked through the web-based modules outside of class, at their own pace, and conducted complementary labs designed to reinforce the concepts. Students’ quantitative skills improved significantly over a semester of completing nine of the 36 available modules, and lower-achieving students (those
who scored lower on the pre-test) improved the most (Thompson et al. 2010). Thompson et al. (2010) made these modules freely available online so other institutions can access them and use any desired number of them to supplement their own courses.

**Study system**

In response to the increasingly quantitative nature of biology research and recommendations from *BIO2010*, the biology major at Western Washington University (Western) has become more quantitative at the upper levels. Most biology majors take an upper-level biostatistics course using the statistical software package “R”. This program is becoming increasingly popular among biology researchers because it is available for free, flexible in which analyses it can perform, and has easily accessible online help resources. However, it is a difficult program to learn because it requires computer programming in R-code, and therefore undergraduates often do not learn how to use it well. The computer science department also offers an interdisciplinary bioinformatics course in which chemistry, biology, and computer science students attend interdisciplinary lectures and collaborate on group projects. Biology students learn how the relevant quantitative skills are necessary for biology research, and students from all the disciplines practice communicating and working together. This course is similar to the mathematics-biology interdisciplinary research programs created at other institutions, but on a smaller scale.

Biology faculty have anecdotally noticed a lack of basic quantitative skills in students entering the major, and in upper-level courses. The department wanted to address this deficit to give their students a strong foundation of quantitative skills. This study attempted to address this
need by adding some of the MathBench Biology Modules (Nelson et al. 2009; Thompson et al. 2010) to an existing introductory-level course.

The Biology Department offers a core set of courses that all undergraduate biology majors must take, a suitable system for testing a strategy in which we add quantitative activities to supplement existing course material. I decided to conduct this study in Biology 205: Introduction to Cell and Molecular Biology, the second course in the introductory-level series. This course is one of the first core courses in which students must use quantitative skills to answer biological questions. The course is also an appropriate study system because the department offers multiple sections every quarter, taught by a number of different professors and lab teaching assistants. The variety of instructors and lab TAs allowed us to generalize the results from this experiment to the entire Biology 205 course.

**Summary of experimental approach**

I first surveyed the Biology faculty to determine what they thought were the most critical basic quantitative skills for their undergraduate biology students. I then developed a quantitative skills assessment and population demographics survey to evaluate these skills in Biology 205 students. I piloted this assessment and used the results to identify apparent deficiencies within these skill areas. With these results, I developed three activities targeted to improve these deficiencies and three control activities to control for (1) skills naturally improving over the quarter and (2) skills improving due to increased effort time. Following methods used by others, I administered the assessment in a pre-/post- test style, with the treatment and control activities supplementing regular course material between the two assessment administrations, to evaluate if the treatment improved the students’ assessment scores.
I simultaneously conducted a population demographics survey to evaluate if any population demographic factors predicted students’ initial performance on the basic quantitative skills assessment and if any population demographic factors predicted their change in score between the pre- and post-tests. I considered the highest level of math class a student had completed because a previous study showed that students eligible for higher-level college math courses scored higher on a quantitative skills assessment in biology (Thompson et al. 2010). I considered whether a student had transferred and if so, from what type of institution, because previous studies have showed that students who transferred from two-year colleges to four-year universities initially had lower GPAs and performed worse in some courses than students who were continuing education (Laband & Hanby 2003; Asarta et al. 2013). I considered a student’s declared major because biology education has historically been less quantitative than the other sciences, which means that students in other science majors would likely have had more experience using quantitative skills in their coursework. I considered a student’s year in school assuming that students who had taken more college-level courses would have better quantitative skills. I considered a student’s gender because some studies have shown that women perform worse on science and mathematics tests, while some have shown that males and females do not perform differently (Hake 1998; McCullough 2004; Metz 2008). I considered a student’s race/ethnicity because studies have shown that underrepresented minority students perform worse in science and mathematics courses than Caucasian students (Goins et al. 2010).

**Research objectives**

I divided my research objectives, data collected and analyses performed into two categories: (1) improving quantitative skills and (2) population demographics and quantitative
skills. The first category evaluated the primary question of whether or not the treatment worked to improve students’ quantitative skills. The second category evaluated how population demographic factors might have predicted students’ basic quantitative skills.

**Improving quantitative skills**

1. **Question**: Did the treatment activities improve students’ basic quantitative skills, their ability to use quantitative skills to solve biological problems, or both?
   
   **Prediction**: Students who completed the treatment activities would improve in both these areas relative to students who completed the control activities.

2. **Question**: Did the treatment activities affect high-achieving students and low-achieving students differently?
   
   **Prediction**: Assessment scores of low-achieving students in the treatment group would improve more than those of high-achieving students in the treatment group because the treatment targeted basic quantitative skills.

**Population demographics and quantitative skills**

3. **Question**: What population demographic factors predicted students’ incoming quantitative skills and biological application of those skills (pre-test score)? I considered the factors discussed above.

   **Predictions:**
   
   - Students who had completed higher-level college mathematics courses would perform better on the assessment.
• Students who had transferred from two-year colleges would perform worse on the assessment.

• Students in the more quantitative majors (physical and natural sciences besides biology) would perform better on the assessment.

• Students in later years in school would perform better on the assessment.

• Women and men would perform similarly on the assessment.

• Racial/ethnic minority students would perform worse on the assessment.

4. **Question:** What population demographic factors predicted students’ changes in quantitative skills and biological application of those skills throughout the quarter (both pre- and post-test scores)? I considered the same factors discussed above.

**Predictions:**

• Students who had completed lower-level college mathematics courses would improve more from pre- to post-testing.

• Students who had transferred from two-year colleges would improve more from pre- to post-testing.

• Students in the biology majors and social science majors would improve more from pre- to post-testing.

• Students in earlier years in school would improve more from pre- to post-testing.

• Women and men would perform similarly from pre- to post-testing.

• Minority students would improve more from pre- to post-testing.
METHODS

Identifying basic quantitative skills to target

To identify the needs of the Biology Department, I began with an informal email to the faculty, asking them to list the most important basic quantitative skills for undergraduate biology majors they felt were lacking in their students. I consolidated this list into a survey which asked the faculty to choose the three most critical skills for students in core undergraduate biology courses, and rank them from most to least critical (Appendix A). I assigned points to the skills from each survey response according to their rank: three points to skills ranked as most critical, two points to skills ranked as second most critical, and one point to skills ranked as third most critical. Survey results indicated the following most critical basic quantitative skills: (1) using and manipulating fractions, proportions, ratios, and percentages; (2) reading and analyzing data, including graphical presentation and interpretation; (3) using and manipulating logarithms; (4) using and manipulating exponents; and (5) translating biological problems into quantitative problems and vice versa (Table 1).

Developing the quantitative skills assessment

Using resources from previous studies (Thompson per. comm.¹), and writing some questions myself, I developed an assessment to evaluate the identified skills in Biology 205 students (Appendix B). The assessment evaluated the first four skills with four questions each. It evaluated the fifth skill, transferring knowledge of quantitative skills to solve biological problems, with a paired-question design. One question in each pair of questions tested a quantitative skill and the other tested that skill within a biological context. I wrote two pairs of

¹ Katerina Thompson University of Maryland College of Computer, Mathematical, and Natural Sciences, 2300 Symons Hall, University of Maryland, College Park, MD 20742, October 2012.
Table 1. Results from the faculty survey (N=19) asking them to identify the three most critical basic quantitative skills for biology undergraduates. Asterisks indicate skills selected for this study.

<table>
<thead>
<tr>
<th>Skills</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Using and manipulating fractions/proportions/ratios/percentages</td>
<td>28*</td>
</tr>
<tr>
<td>Reading and analyzing data, including graphical presentation and interpretation</td>
<td>16*</td>
</tr>
<tr>
<td>Using and manipulating logarithms</td>
<td>13*</td>
</tr>
<tr>
<td>Using scientific exponents, including scientific notation</td>
<td>11*</td>
</tr>
<tr>
<td>Translating biological problems into quantitative problems and vice versa</td>
<td>11*</td>
</tr>
<tr>
<td>Using the metric system</td>
<td>10</td>
</tr>
<tr>
<td>Using basic algebra</td>
<td>5</td>
</tr>
<tr>
<td>Enough calculus to know that dN/dt is instantaneous rate of change</td>
<td>5</td>
</tr>
<tr>
<td>Sense of the size of numbers</td>
<td>4</td>
</tr>
<tr>
<td>Using probability to make predictions</td>
<td>3</td>
</tr>
<tr>
<td>Basic numeracy/arithmetic, order of operations, measured numbers</td>
<td>0</td>
</tr>
<tr>
<td>Using dimensional analysis</td>
<td>0</td>
</tr>
</tbody>
</table>
questions for each skill to evaluate each skill more accurately than one pair would, but also to keep the assessment relatively short. I wanted to be able to administer the assessment at the end of a lab period, and easily persuade students to take it. I initially combined the two skills “using and manipulating exponents” and “using and manipulating logarithms,” into one skill (with only two pairs of questions) because they are more closely related than the other skills, but after piloting the assessment, I learned I could better evaluate both skills by separating them. The final assessment consisted of 8 pairs of questions, 16 questions total. I included a population demographics survey with the quantitative skills assessment. The survey gathered information about the highest college-level math class a student had completed, transfer history, major, year in school, gender and ethnicity.

I piloted the quantitative skills assessment and associated population demographic survey (Appendix C) in the fall of 2012 to test the validity and reliability of the assessment (Phelan & Wren 2006). I first piloted the assessment in November by recruiting student volunteers (Tables 2-3) to take the test in an open-response format. I used the most common wrong answers from these responses to write multiple-choice distracter options so that I could later administer the assessment as a multiple-choice test. I piloted the revised multiple-choice assessment in December 2012 (Tables 2-3). I administered this test on Blackboard Learning Management System (a web-based course management system) and randomized the question order, thus controlling for the possibility of the first question in each pair of questions becoming a “treatment” for the second. Student volunteers received one extra credit point in compensation for their time. If students wanted to earn extra credit, but did not wish to participate in the study, they had the option of an alternate assignment reading essays and answering follow-up questions.
Table 2. Project timeline

<table>
<thead>
<tr>
<th></th>
<th>Fall 2012</th>
<th>Winter 2013</th>
<th>Spring 2013</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>November 13-14</td>
<td>January 15-25</td>
<td>April 9-12</td>
</tr>
<tr>
<td></td>
<td>First pilot</td>
<td>Pre-test</td>
<td>Pre-test</td>
</tr>
<tr>
<td></td>
<td>December 4-6</td>
<td>January 29-February 8</td>
<td>April 23-26</td>
</tr>
<tr>
<td></td>
<td>Second pilot</td>
<td>Activity 1</td>
<td>Activity 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>February 5-15</td>
<td>May 7-10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Activity 2</td>
<td>Activity 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>February 26-March 8</td>
<td>May 21-24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-test</td>
<td>Activity 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>March 12-15</td>
<td>June 4-6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3. Sample sizes of all participant groups. Includes only full study participants, defined as students who completed the pre-test, all three activities, and post-test.

<table>
<thead>
<tr>
<th>PILOT</th>
<th>Number of Students</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Completed Assessment</td>
<td>Interview Participants</td>
<td></td>
</tr>
<tr>
<td>Fall 2012</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First pilot</td>
<td>33</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Second pilot</td>
<td>90</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>123</td>
<td>17</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EXPERIMENT</th>
<th>Treatment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low-achieving</td>
<td>36</td>
<td>43</td>
</tr>
<tr>
<td>High-achieving</td>
<td>55</td>
<td>57</td>
</tr>
<tr>
<td>Treatment Group Total</td>
<td>91</td>
<td>100</td>
</tr>
</tbody>
</table>
Assessment validity and reliability

I evaluated the assessment for construct validity (Phelan & Wren 2006), which is how well the assessment questions measure a student’s underlying understanding of the targeted quantitative skills, by giving the assessment to a team of experts (Phelan & Wren 2006). The experts included my thesis committee members, Dr. Kimberly Markworth, a mathematics education researcher, and Dr. Jean-Louis Dethier, another mathematician. The experts were able to identify whether or not the questions tested the underlying skills I intended to test. I also evaluated the assessment for face validity, which is how well students understand the assessment questions and if students use the intended targeted skills to answer the questions (Phelan & Wren 2006). I evaluated face validity by recruiting student volunteers from both pilot test administrations for follow-up interviews (Phelan & Wren 2006). The interviews lasted approximately 10-20 minutes, during which time the students looked through their tests and explained their thinking behind their chosen process for solving each problem. I made audio recordings of these interviews, and later analyzed the interviews by comparing student responses to an answer key that detailed how to solve each problem. I marked whether students found (1) the correct answer via the intended approach, (2) the correct answer via an unintended approach, (3) the incorrect answer via the intended approach, or (4) the incorrect answer via an unintended approach. I noted confusing wording and when questions were answered correctly or incorrectly for the wrong reasons (2 or 4 above) multiple times, and used this information to re-write some questions. Students who participated in these interviews received small gift certificates in compensation for their time.

Based on validity results from the team of experts and student interviews, I adjusted unclear or ambiguous wording. At the same time I changed the skill “using and manipulating
logarithms and exponents” into two separate skills: “using and manipulating logarithms” and “using and manipulating exponents.” I wrote four additional questions addressing these skills for a balanced study design. I used knowledge from the two pilots and interviews to write the new questions clearly and very similarly to the existing questions. I then administered this assessment and the population demographics survey as pre- and post- tests at the beginning and end of the winter and spring quarters 2013 (Appendix B-C, E-F; Tables 2-3).

I evaluated both the winter and spring assessments for reliability, also known as internal consistency, which measures how well an assessment gives consistent results (Phelan & Wren 2006). I used the Chronbach’s alpha analysis, which is very commonly used for this purpose, and measures how well each test question measures the same underlying concept: basic quantitative skills in this case (Tavakol & Dennick 2011). I required an alpha equal to or greater than 0.60 for this assessment because it was relatively short and had a relatively small number of test takers (Thompson per. comm.2). From the results of the Cronbach’s alpha analysis, team of experts, and interviews I concluded that the revised assessment was reliable and valid (Table 4).

**Targeting quantitative skills improvement**

I used the results from the second pilot assessment and all interviews to identify four more specific skills within the identified basic quantitative skills with which students struggled most (Table 5). I calculated mean percent scores for each question and selected those that were below average, compared to the grand mean for all the mean question scores. Based on the multiple-choice responses and interview statements, I inferred more specifically where students

---

2 Katerina Thompson University of Maryland College of Computer, Mathematical, and Natural Sciences, 2300 Symons Hall, University of Maryland, College Park, MD 20742, January 2013.
Table 4. Cronbach's alpha results from all assessment administrations. $\alpha > 0.60$ indicates a reliable assessment for this study.

<table>
<thead>
<tr>
<th>Test</th>
<th>Items</th>
<th>Number of Students</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Winter pre-test</td>
<td>16</td>
<td>189</td>
<td>0.634</td>
</tr>
<tr>
<td>Winter post-test</td>
<td>16</td>
<td>155</td>
<td>0.743</td>
</tr>
<tr>
<td>Spring pre-test</td>
<td>16</td>
<td>140</td>
<td>0.686</td>
</tr>
<tr>
<td>Spring post-test</td>
<td>16</td>
<td>141</td>
<td>0.824</td>
</tr>
</tbody>
</table>
Table 5. Second pilot results used to determine the three skill areas to target with the quantitative activities. SD represents Standard Deviation.

<table>
<thead>
<tr>
<th>Question</th>
<th>Mean Score (Percent ± SD)</th>
<th>Comments</th>
<th>Skill to Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>*1</td>
<td>52 ± 0.50</td>
<td></td>
<td>Exponential growth</td>
</tr>
<tr>
<td>2</td>
<td>77 ± 0.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>*3</td>
<td>32 ± 0.47</td>
<td></td>
<td>Log rules</td>
</tr>
<tr>
<td>*4</td>
<td>54 ± 0.50</td>
<td></td>
<td>Log rules</td>
</tr>
<tr>
<td>5</td>
<td>74 ± 0.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>85 ± 0.36</td>
<td>Interviews revealed that students were unclear about why they chose the correct response; they seemed to have an intuitive understanding of the concept.</td>
<td>Explicit understanding of what types of graphs are appropriate for what types of data and vice versa</td>
</tr>
<tr>
<td>7</td>
<td>94 ± 0.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>84 ± 0.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>*9</td>
<td>64 ± 0.48</td>
<td>56% of students who answered incorrectly did not follow directions and divide. 31% estimated instead of calculating the exact answer</td>
<td>Following directions</td>
</tr>
<tr>
<td>10</td>
<td>70 ± 0.46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>75 ± 0.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>*12</td>
<td>36 ± 0.48</td>
<td>93% of students who answered incorrectly did not follow directions and round.</td>
<td>Following directions</td>
</tr>
</tbody>
</table>

Grand mean of all questions’ mean scores: 66 ± 0.47

*Represents questions whose average score was below average compared to the rest of the question averages.
lacked quantitative skills. I selected the following skills to target with the treatment activities: (1) exponential growth; (2) using and manipulating logarithms; (3) types of graphs appropriate for different types of data, and (4) following mathematics-specific directions (rounding, simplifying, etc.). I developed three treatment activities targeted to improve the first three skills: exponential growth, using and manipulating logarithms, and types of graphs appropriate for different types of data. All activities targeted the fourth skill, following mathematics-specific directions, by increasing the amount of time biology students spent using quantitative skills (Appendix G). Following methods used by others (Metz 2008; Thompson et al. 2010; Colon-Berlingeri & Burrowes 2011), I administered the assessment before and after the treatment, which included all three targeted activities. The activities included selected MathBench biology modules, and follow-up practice problems (MathBench; Appendix G). I also developed three control activities to control for the potentially confounding effects of students improving due to: (1) the course, and (2) more time spent doing schoolwork. The control activities took approximately the same amount of time as the treatment activities, and involved reading articles about mathematics and biology and answering follow-up questions (Appendix H).

I administered the treatment activities in half of Biology 205 lab sections and the control activities in the other half during the winter and spring quarters 2013. Each professor taught at least two lab sections of Biology 205 so I assigned one treatment and one control section per professor (Table 6). I haphazardly assigned the remaining winter quarter lab sections to treatment or control groups by flipping a coin, and the remaining spring quarter lab section to create a balanced study design (Table 6). I did not inform students or lab instructors of which treatment group I assigned them to prevent this knowledge from biasing their efforts. I never revealed the
Table 6. Lecture and lab sections assigned to each treatment group for both quarters.

<table>
<thead>
<tr>
<th>Quarter</th>
<th>Lecture Instructor</th>
<th>Lab Section</th>
<th>Treatment Group</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Winter 2013</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professor A</td>
<td>Tuesday 8-11</td>
<td>Control</td>
<td></td>
</tr>
<tr>
<td>Professor A</td>
<td>Tuesday 12-3</td>
<td>Treatment</td>
<td></td>
</tr>
<tr>
<td>Professor A</td>
<td>Thursday 2-5</td>
<td>Treatment</td>
<td></td>
</tr>
<tr>
<td>Professor B</td>
<td>Wednesday 2-5</td>
<td>Treatment</td>
<td></td>
</tr>
<tr>
<td>Professor B</td>
<td>Thursday 8-11</td>
<td>Control</td>
<td></td>
</tr>
<tr>
<td>Professor B</td>
<td>Thursday 11-2</td>
<td>Control</td>
<td></td>
</tr>
<tr>
<td>Professor C</td>
<td>Wednesday 8-11</td>
<td>Control</td>
<td></td>
</tr>
<tr>
<td>Professor C</td>
<td>Wednesday 11-2</td>
<td>Treatment</td>
<td></td>
</tr>
<tr>
<td>Professor C</td>
<td>Friday 1-2</td>
<td>Treatment</td>
<td></td>
</tr>
<tr>
<td><strong>Spring 2013</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professor D</td>
<td>Tuesday 8-11</td>
<td>Treatment</td>
<td></td>
</tr>
<tr>
<td>Professor D</td>
<td>Tuesday 2-5</td>
<td>Control</td>
<td></td>
</tr>
<tr>
<td>Professor D</td>
<td>Wednesday 1-4</td>
<td>Control</td>
<td></td>
</tr>
<tr>
<td>Professor D</td>
<td>Friday 1-4</td>
<td>Treatment</td>
<td></td>
</tr>
<tr>
<td>Professor E</td>
<td>Wednesday 8-11</td>
<td>Control</td>
<td></td>
</tr>
<tr>
<td>Professor E</td>
<td>Thursday 8-11</td>
<td>Control</td>
<td></td>
</tr>
<tr>
<td>Professor E</td>
<td>Thursday 11-2</td>
<td>Treatment</td>
<td></td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8 sections</td>
<td>Control</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>8 sections</td>
<td>Treatment</td>
</tr>
</tbody>
</table>
correct answers to the assessment questions or activities to discourage students from cheating. I administered the assessments, demographic surveys and activities using Blackboard LMS.

To motivate students to take the assessments and activities seriously they received a small number of course points based on completion (1% of their overall lab grade). Students who did not wish to participate in the study completed the assessments and assigned activities for course points, but I did not include their results in the study. Students submitted activities electronically or by hand. I allowed students to submit late activities for partial credit until 3-4 days before the post-test. I graded a subset of the treatment activities submitted electronically (20 per activity) to evaluate how effectively each module enabled students to answer the follow-up questions. I also used these results to identify areas for improvement in the treatment activities. I performed t-tests comparing grades on the winter and spring quarter activities and found no significant differences so I pooled results from both quarters for graphical evaluation.

**Spring quarter changes**

I intended to use the same tools and follow the same design for the experiment in the spring quarter as the winter quarter, but discovered some problems while implementing the experiment in the winter quarter so I made changes for the spring quarter. After administering the assessment during the winter quarter I discovered some formatting problems with Blackboard LMS so students could not clearly read two of the logarithm questions. I revised the assessment slightly by inserting these questions as pictures instead of as text so the formatting would be clear. I also removed the multiple-choice option “I do not know how to approach this problem” to encourage students to put more effort into the assessment. I administered the revised assessment at the beginning (pre-test) and end (post-test) of the spring quarter (Appendix E).
After receiving results from the pre- and post-test surveys I also discovered that some students changed their responses to questions when they should not have: for example, “what is the highest college-level mathematics class you have completed?” For the spring quarter, I revised the post-test population demographics survey to exclude these questions, though I still gathered the same information from every student on the pre-test. On the spring population demographics survey I also included a self-reported measure of effort invested in the activities and post-assessment to potentially identify outliers due to low effort, but did not find enough patterns to use this measure (Appendix F).

In the winter quarter I created one Blackboard LMS site for all study participants to take the assessments and demographic surveys, and access the activities. I instructed students to only complete the activities for their assigned treatment section (A or B), but they had the ability to access the activities for the other group. During the spring quarter I created two separate Blackboard LMS sites, one for the treatment sections and one for the control sections so that students would be much less likely to intentionally or unintentionally look at the activities for the group that was not their assigned treatment. While grading a subset of the activities after the winter quarter I discovered that one treatment activity had some slightly ambiguous wording so I altered the instructions for the spring quarter (Appendix I). I also made minor changes to the schedule of activities to space them out and better fit them into the Biology 205 lab schedule (Table 2).

**Statistical analyses**

For all analyses I only included students who participated fully in the study, defined as students who completed both assessments (pre- and post-tests) and all three activities.
students enrolled in Biology 205 multiple times during the 2012-2013 academic year so I removed them from the dataset the second time they participated because they took the assessment and/or completed at least one activity during their first quarter, which may have affected their performance during their second quarter. To assess whether the treatment affected high- and low-achieving students differently, I classified each student as high- or low-achieving based on his or her pre-test score. I calculated the mean pre-test score for all study participants and students who scored below that I classified as low-achieving, while students who scored above that I classified as high-achieving. I chose the mean instead of the median for the cut-off in this grouping because the two numbers were very similar (the median was 11 and the mean was 10.6). Also, all scores were integers so if I had used the median I would have had to choose whether to classify students at the median as high- or low-achieving, put them in a separate category, or remove them from the analysis.

I used generalized linear mixed-effect models (GLMMs) to answer all my research questions. GLMMs are robust, powerful, and flexible tests that allow for violations of many traditional hypothesis-testing assumptions, such as non-normally distributed data, non-independent data points, and correlated factors (Bolker et al. 2008). GLMMs are becoming increasingly common in some fields of biology research, namely ecology and evolution, because these fields have notoriously “messy” datasets that often violate assumptions of traditional hypothesis-testing (Bolker et al. 2008). Science education datasets can be similarly “messy,” but to my knowledge these analyses have never before been used for biology education research (Freeman per. comm.\textsuperscript{3}).

\textsuperscript{3} Scott Freeman University of Washington Biology Department, Box 351800 Seattle, WA 98195, June 2012.
GLMMs were the most appropriate statistical analysis for my dataset for multiple reasons. First, the response variables (dependent variables) were not normally distributed. The response variable for the analysis asking if the treatment improved students’ quantitative skills was whether a question was answered correctly or incorrectly. I chose this response variable instead of the commonly used normalized gains score, because I was still able to consider the differences between pre- and post-test scores by including test as a predictor variable in the analyses, and normalized gains scores cannot be computed for students who score 100% on the pre-test (Metz 2008). The response variable of score had two possible outcomes: 0 or 1. These data fit a binomial distribution rather than a normal distribution. The response variables for the analyses asking about population demographic factors were students’ overall test scores. These data also fit a binomial distribution better than a normal one because their range was restricted, students could only score 0-16, and the mean did not fall in the middle of this range (as it must for a normal distribution). Traditional hypothesis-testing requires normally distributed data, but GLMMs allowed me to choose a binomial distribution for all analyses.

GLMMs were also the most appropriate statistical analysis for my dataset because there were two random factors in my study: quarter and lab section. Random factors are usually not the factors of primary interest, and often researchers cannot collect data from all existing levels, but these factors may still have an effect on the response. For example, I collected data during two academic quarters: the winter and spring of 2013. There may have been differences in students who took the course in the winter as compared to the spring. In traditional hypothesis-testing these additional factors would weaken the statistical Power of the analysis for the fixed factors in which I was primarily interested. GLMMs allow for fixed and random factors and their associated errors to be evaluated independently, and additional factors do not weaken the
statistical power of the analysis (Bolker et al. 2008, Zuur et al. 2009). GLMMs allowed me to include the random factors of “quarter” and “lab section” in all analyses without weakening the statistical power of the analysis. In contrast, fixed factors are the factors of primary interest; researchers usually collect data from all possible levels or groups within that factor, and often the researcher defines those levels. For example, in my study the factor “treatment” is the primary factor under study; I collected data from the two levels of that factor that I defined: “treatment” and “control”.

Unlike traditional hypothesis-testing, GLMMs do not result in a p-value. Instead, they determine the simplest model that includes the most important predictor variables that best explain the response. I began with full models including all possible predictor variables, and relevant interactions (two factors interact when one factor’s effect on the response depends upon the other factor). I compared full models to alternative models with fewer predictor variables and different combinations of these predictors. In GLMM analysis there are several different measurements that can evaluate how well models fit the response variable. One of the most common is the Akaike Information Criterion (AIC) score, which I used. Lower AIC scores indicate better-fitting models (Zuur et al. 2009). There are also several different methods for estimating a model’s parameters, mathematically representing the contributions of fixed and random factors (Zuur et al. 2009). I used Maximum Likelihood (ML) estimates because the models contained different combinations of fixed factors (Zuur et al. 2009). Sometimes several models have AIC scores that are very close to each other, which was the case for some of my analyses. In such cases, statisticians recommend calculating evidence ratios which measure how much more likely the “best-fit” model is than the subsequent models. Statisticians recommend averaging all models that have evidence ratios less than 2.7, which is the protocol I followed.
(Mazerolle 2004, Ross 2013). The averaged model gives the relative importance of each predictor variable from all the contributing models. I validated all best-fit models using residual plots of the fitted values and each factor (Zuur et al. 2009). I used the R project for statistical computing for all analyses.

**Improving quantitative skills: research objectives one and two**

The following analyses evaluated the primary question of whether or not the treatment improved students’ quantitative skills. I ran a series of models for the response variable score, defined as whether a student answered a question correctly or incorrectly. These models included the fixed predictor variables of: (1) treatment: whether a student was in the control or treatment group; (2) construct: whether the question tested a quantitative skill or applying that skill to a biological context; (3) achievement: whether a student was low- or high-achieving; and (4) test: whether the question was from the pre- or post-test (Table 7). These models also included the random predictor variables of (1) quarter, and (2) lab section. To visually examine the summarized data, I made boxplots of the pre- and post- test scores for each treatment group, and pre- and post-test scores in each quantitative skill area for each treatment group. I also performed a t-test on the pre-test scores of both treatment groups for each quarter because I assumed that on average, the groups would initially have similar levels of basic quantitative skills and abilities to apply those skills to biological contexts. Because these analyses revealed that students in the treatment group during the spring quarter scored significantly higher on the pre-test than students in the control, I also performed the same GLMM analysis described above on only the winter quarter dataset.
Table 7. Full model for each set of GLMM analyses.

<table>
<thead>
<tr>
<th></th>
<th>Factor</th>
<th>Factor Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questions 1-2:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>improving</td>
<td></td>
<td></td>
</tr>
<tr>
<td>quantitative skills</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response</td>
<td>Correct vs. incorrect</td>
<td></td>
</tr>
<tr>
<td>answer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fixed factors</td>
<td>Treatment</td>
<td>Treatment vs. control</td>
</tr>
<tr>
<td></td>
<td>Construct</td>
<td>Math vs. math-biology</td>
</tr>
<tr>
<td></td>
<td>Achievement</td>
<td>Low vs. high</td>
</tr>
<tr>
<td></td>
<td>Test</td>
<td>Pre vs. post</td>
</tr>
<tr>
<td>Random factors</td>
<td>Quarter</td>
<td>Winter vs. spring</td>
</tr>
<tr>
<td></td>
<td>Lab section</td>
<td></td>
</tr>
<tr>
<td>Question 3:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>population</td>
<td></td>
<td></td>
</tr>
<tr>
<td>demographics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>predicting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pre-test score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response</td>
<td>Test score (combined</td>
<td></td>
</tr>
<tr>
<td></td>
<td>number of correct and</td>
<td></td>
</tr>
<tr>
<td></td>
<td>incorrect)</td>
<td></td>
</tr>
<tr>
<td>Fixed factors</td>
<td>Year</td>
<td>Freshmen, sophomore, etc.</td>
</tr>
<tr>
<td></td>
<td>Transfer history</td>
<td>If transferred and from what</td>
</tr>
<tr>
<td></td>
<td></td>
<td>type of institution</td>
</tr>
<tr>
<td></td>
<td>Highest college-level</td>
<td></td>
</tr>
<tr>
<td></td>
<td>math completed</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gender</td>
<td>Male vs. female</td>
</tr>
<tr>
<td>Random factors</td>
<td>Quarter</td>
<td>Winter vs. spring</td>
</tr>
<tr>
<td></td>
<td>Lab section</td>
<td></td>
</tr>
<tr>
<td>Question 4:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>population</td>
<td></td>
<td></td>
</tr>
<tr>
<td>demographics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>predicting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>change from</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pre- to post-test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response</td>
<td>Test score (combined</td>
<td></td>
</tr>
<tr>
<td></td>
<td>number of correct and</td>
<td></td>
</tr>
<tr>
<td></td>
<td>incorrect)</td>
<td></td>
</tr>
<tr>
<td>Fixed factors</td>
<td>Year</td>
<td>Freshmen, sophomore, etc.</td>
</tr>
<tr>
<td></td>
<td>Transfer history</td>
<td>If transferred and from what</td>
</tr>
<tr>
<td></td>
<td></td>
<td>type of institution</td>
</tr>
<tr>
<td></td>
<td>Highest college-level</td>
<td></td>
</tr>
<tr>
<td></td>
<td>math completed</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gender</td>
<td>Male vs. female</td>
</tr>
<tr>
<td></td>
<td>Test</td>
<td>Pre vs. post</td>
</tr>
<tr>
<td>Random factors</td>
<td>Quarter</td>
<td>Winter vs. spring</td>
</tr>
<tr>
<td></td>
<td>Lab section</td>
<td></td>
</tr>
</tbody>
</table>
Population demographics and quantitative skills: research objectives three and four

The following analyses evaluated the questions of (1) which population demographics factors predicted students’ incoming assessment scores (pre-test scores) and (2) which of these factors predicted changes in assessment scores (pre- and post-test scores). For both groups of models the response variable was a combined factor of two values: the number of questions a student answered correctly and the number of questions that student answered incorrectly. Though these values are related to each other, they were both necessary to indicate the range restriction (0-16) for the response variable’s distribution. The population demographics models included the fixed predictor variables of: (1) previous math: highest college-level math class completed; (2) transfer history: if a student transferred and from what type of institution; (3) major: what major a student declared; (4) test: whether it was the pre- or post-test (only for the analysis answering research question four); (5) year: a student’s year in school; (6) gender; and (7) race/ethnicity. These models also included the same random predictor variables as for the above analysis: (1) quarter, and (2) lab section. I ran all possible combinations of fixed factors for both groups of models, but did not include any interactions because of the complexity of the models, and relatively small sample size of some groups within the interactions.

After data collection, I discovered that some levels of the fixed population demographics factors had very small sample sizes so I pooled some of these groups. For previous math, I pooled all students whose highest college-level mathematics class taken was above Math 125 except for students who had taken Math 240: statistics. The math 240 group was larger than all those for the other upper-level classes and though statistics is a higher-level course, it has lower-level pre-requisites so it does not follow as well in the same succession as the other math courses. For major I pooled all students with different biology emphases into the overarching
major “biology”: general biology, marine biology, ecology evolution and organismal biology, cellular and molecular biology, and biochemistry. I classified all students within the majors chemistry, kinesiology, environmental studies, and behavioral neuroscience as “other physical/natural sciences.” I classified all students within the majors psychology, community health, and anthropology as “social sciences” because at this university they are not included in the College of Science and Technology. I classified students who declared English or other majors as “other”. For ethnicity, I pooled the groups with the smallest sample sizes (only two or three students each) into the category “Minorities”: Hispanic, African American, Native American, and Hawaiian/Pacific Islander students. For transfer history, I pooled two groups so that the remaining levels were: students who had not transferred, students who had transferred from a two-year college, and students who had transferred from a two-year college or four-year university, to identify the students who had no previous experience of a four-year institution (Appendix J includes all models tested).

To visually examine the summarized data, I made boxplots of pre- and post-test scores for all population demographics factors that were include in the averaged best-fit models. I also graphed normalized gains scores, though I did not include them in the analyses, because I did evaluate differences between pre- and post- test scores and normalized gains helped visualize these results.

To evaluate student impressions and opinions of the experiment, I made contingency tables of their responses to several questions from the population demographics survey, including: (1) Are quantitative skills important for biology? (2) Should Biology 205 include more math? And (3) Do you think the activities helped improve your quantitative skills? I used G-tests of independence to evaluate the differences between each treatment group’s responses on
the post-test as compared to the pre-test for the first two questions. The third question was only applicable on the post-test, so I compared student responses in the treatment group to the control group. I adjusted alpha to 0.01 to correct for multiple (five) comparisons.
RESULTS

Improving quantitative skills (research objectives one and two)

The best-fit model from the GLMM analysis evaluating factors influencing students’
assessments scores from both quarters included three terms. These terms were: (1) an interaction
between the fixed factors test, achievement, and construct, and the random factors (2) quarter
and (3) lab section (Table 8). This model was 6.5 times more likely than the next closest best-fit
model, which also included an interaction between treatment and test (Table 8).

The three-way interaction between test, achievement, and construct indicated that a
student’s relative performance on the pre- and post- tests depended upon which achievement
group s/he was in (low- vs. high-achieving), and which construct the question addressed (purely
quantitative skills or using quantitative skills to solve biological problems). Low-achieving
students performed similarly on the quantitative questions on both the pre- and post- tests, while
high-achieving students scored lower on the quantitative questions on the post-test as compared
to the pre-test (Fig. 1). Low-achieving students scored higher on questions that required
quantitative skills for biological questions on the post-test as compared to the pre-test, while
high-achieving students scored similarly on these questions on both tests (Fig. 1).

If the treatment had improved students’ assessment scores I would expect the best-fit
model to include an interaction between treatment and one or more of the other fixed factors.
The best-fit model did not include any of these terms, but graphical representation of the data
revealed that students in the treatment group scored higher on the post-test as compared to the
pre-test while students in the control group scored similarly on both tests (Fig. 2). In the winter
quarter students in the treatment and control groups performed similarly on the pre-test (Fig. 3).
However, during the spring quarter, students in the treatment group scored significantly
Table 8. Top five models evaluating the factors that predicted whether students answered questions correctly or incorrectly on the pre- and post-tests. The gray box highlights the best-fit model, and the top five other models are listed in order from best to worst fit; lowest to highest Akaike Information Criterion (AIC) value. All models tested are included in Appendix J. df represents degrees of freedom for each model, delta AIC is the difference between the given model’s AIC score and the best model’s AIC score. Evidence ratios compare the best-fit model to the other top five models; these values indicate how much more likely the best-fit model is than the given model.

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>AIC</th>
<th>Delta AIC (ΔAIC)</th>
<th>Akaike Weight (w_i)</th>
<th>Evidence Ratio (w_1/w_i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M14 = score ~ construct<em>achievement</em>test + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>10</td>
<td>7431.0</td>
<td>0</td>
</tr>
<tr>
<td>M16 = score ~ treatment<em>test + construct</em>achievement*test + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>12</td>
<td>7434.7</td>
<td>3.7</td>
</tr>
<tr>
<td>M10 = score ~ achievement*test + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>6</td>
<td>7435.7</td>
<td>4.7</td>
</tr>
<tr>
<td>M13 = score ~ treatment<em>achievement</em>test + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>10</td>
<td>7439.6</td>
<td>8.6</td>
</tr>
<tr>
<td>M15 = score ~ treatment<em>construct</em>achievement*test</td>
<td>18</td>
<td>7439.9</td>
<td>8.9</td>
<td>0.01</td>
<td>78</td>
</tr>
</tbody>
</table>
Figure 1. Pre- and post-test scores of students divided by achievement-level for purely quantitative questions compared to those requiring quantitative skills to answer biological questions. This figure includes students in both the control and treatment groups. Horizontal lines represent the median score, and box limits represent the first quartile (Q1) and third quartile (Q3). Dashed whiskers indicate the most extreme data points within 1.5*IQR where IQR is the interquartile range: Q3-Q1. The points represent outliers which are outside 1.5*IQR.
Figure 2. Overall pre- and post-test scores for control and treatment groups. Horizontal lines represent the median score, and box limits represent the first quartile (Q1) and third quartile (Q3). Dashed whiskers indicate the most extreme data points within 1.5*IQR where IQR is the interquartile range: Q3-Q1. The points represent outliers which are outside 1.5*IQR.
Figure 3. Comparison of pre-test scores of control and treatment groups for each quarter and corresponding t-test results. Horizontal lines represent the median score, and box limits represent the first quartile (Q1) and third quartile (Q3). Dashed whiskers indicate the most extreme data points within 1.5*IQR where IQR is the interquartile range: Q3-Q1. The points represent outliers which are outside 1.5*IQR. Numbers in the boxes represent sample sizes.
higher on the pre-test than students in the control group (Fig. 3).

Because of these differences, I conducted the same overall GLMM analysis on the winter quarter dataset only. The top three models competed as “best-fit” so I averaged them to determine the relative importance of each variable (Table 9). One of the contributing models contained the interaction of test and achievement; the second contributing model contained the interaction of test, treatment and achievement; and the third contributing model was the same as for the analysis including both quarters. All models included the same random factors as above: quarter and lab section. Because the second contributing model contained a factor most different from the analysis including both quarters, I graphed the interaction in that model to further explore it graphically (Fig. 4). Both groups of low-achieving students appeared to improve from pre- to post-testing. High-achieving students in the control group scored worse on the post-test as compared to the pre-test, while high-achieving students in the treatment group scored similarly on both tests (Fig. 4). Though treatment was included as a contributing factor in some interactions in this model, these interactions were moderately important compared to the other variables (Table 9). It is difficult to see trends related to treatment, but treatment did contribute a little explanatory power in the winter when it did not for the full dataset.

Activity results

To better understand how students responded to each treatment activity, I graded a random subset of them. Visual comparison of the trends revealed that student performance varied with skill type: students scored relatively high on the activity about exponential growth and decay, lower on the activity about graphing, and lowest on the activity about logarithms (Fig. 5). Though I did not statistically compare the specific quantitative skill areas (exponents, logs,
Table 9. Top five models evaluating the factors that predicted whether students answered questions correctly or incorrectly on the pre- and post-tests during the winter quarter only. The gray box highlights the three models that are similarly “best-fit” ranked from lowest to highest Akaike Information Criterion (AIC). df represents degrees of freedom for each model, delta AIC is the difference between the given model’s AIC score and the best model’s AIC score. Evidence ratios compare the best-fit model to the other top five models; these values indicate how much more likely the best-fit model is than the given model.

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>AIC</th>
<th>ΔAIC</th>
<th>Akaike Weight (w_i)</th>
<th>Evidence Ratio (w_1/w_i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M10 = score (winter) ~ achievement*test + (1</td>
<td>section)</td>
<td>5</td>
<td>4130.6</td>
<td>0.3755</td>
<td>1</td>
</tr>
<tr>
<td>M13 = score (winter) ~ treatment<em>achievement</em>test + (1</td>
<td>section)</td>
<td>9</td>
<td>4131.1</td>
<td>0.5</td>
<td>0.2925</td>
</tr>
<tr>
<td>M14 = score (winter) ~ construct<em>achievement</em>test + (1</td>
<td>section)</td>
<td>9</td>
<td>4132.0</td>
<td>1.4</td>
<td>0.1865</td>
</tr>
<tr>
<td>M16 = score (winter) ~ treatment<em>test + construct</em>achievement*test + (1</td>
<td>section)</td>
<td>11</td>
<td>4133.0</td>
<td>2.4</td>
<td>0.1131</td>
</tr>
<tr>
<td>M6 = score (winter) ~ treatment*achievement + (1</td>
<td>section)</td>
<td>5</td>
<td>4135.5</td>
<td>4.9</td>
<td>0.0324</td>
</tr>
</tbody>
</table>

Relative importance of variables in averaged model
achievement 1.00
test 1.00
achievement*test 1.00
treatment 0.34
achievement*treatment 0.34
test*treatment 0.34
achievement*test*treatment 0.34
construct 0.22
achievement*construct 0.22
construct*test 0.22
achievement*construct*test 0.22
Figure 4. Pre and post-test scores of students in both treatment groups divided by achievement-level for the winter quarter only. Horizontal lines represent the median score, and box limits represent the first quartile (Q1) and third quartile (Q3). Dashed whiskers indicate the most extreme data points within 1.5*IQR where IQR is the interquartile range: Q3-Q1. The points represent outliers which are outside 1.5*IQR.
Figure 5. Graded subset of activities from both quarters. Horizontal lines represent the median score, and box limits represent the first quartile (Q1) and third quartile (Q3). Dashed whiskers indicate the most extreme data points within 1.5*IQR where IQR is the interquartile range: Q3-Q1. The points represent outliers which are outside 1.5*IQR. T-tests comparing the winter and spring quarter grades revealed no significant differences between the quarters: (Exponents: t=0.53, df=36.2, p=0.60; Logs: t=0.96, df=24.7, p=0.35; Graphing: t=-1.05, df=37.8, p=0.30).
graphs, and fractions) in the GLMM analysis, I graphed student assessment scores on these questions for both the treatment and control groups to visually evaluate the effectiveness of the activities that students in the treatment group completed (Fig. 6). Students in both the control and treatment groups apparently performed similarly on the exponent questions from pre- to post-testing. Students in the control group evidently improved on the post-test questions about logarithms, while students in the treatment group performed similarly on both tests. However, students in the control group initially scored lower than students in the treatment group on the logarithm questions. Students in the control and treatment groups both appeared to perform better on the graphing questions from pre- to post-tests. This graphical comparison also suggested that both groups performed similarly on the questions about fractions (Fig. 6).

**Population demographics and quantitative skills (research objectives three and four)**

*Pre-test score (research objective three)*

The GLMM analysis identifying population demographic factors that predicted students’ incoming basic quantitative skills and abilities to apply them to biological problems determined that the top four models competed as “best-fit” (Table 10). I averaged these four models to determine the relative importance of each factor. The averaged model included the fixed factors of: previous math, transfer history, major, year and gender. It also included the same random factors as for the analysis evaluating treatment: quarter and lab section. Previous math (highest college-level math completed) and transfer history (if students transferred to this university and from what type of institution) were the most important variables. Major was the next most important variable, followed by year and gender (Table 10). I based the following descriptions
Figure 6. Pre- and post-test scores of quantitative skill areas for each treatment group. Horizontal lines represent the median score, and box limits represent the first quartile (Q1) and third quartile (Q3). Dashed whiskers indicate the most extreme data points within $1.5 \times \text{IQR}$ where IQR is the interquartile range: $Q_3 - Q_1$. The points represent outliers which are outside $1.5 \times \text{IQR}$. 
Table 10. Top five models for population demographics and quantitative skills analysis evaluating the factors that predicted pre-test score. The gray box highlights the four models that are similarly “best-fit” ranked from lowest to highest Akaike Information Criterion (AIC). All models tested are included in Appendix J. df represents degrees of freedom for each model, delta AIC is the difference between the given model’s AIC score and the best model’s AIC score. Evidence ratios compare the best-fit model to the other top five models; these values indicate how much more likely the best-fit model is than the given model.

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>AIC</th>
<th>ΔAIC</th>
<th>Akaike Weight (wi)</th>
<th>Evidence Ratio (w1/wi)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEx47 = score ~ transfer + p.math + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>11</td>
<td>340.86</td>
<td>0.00</td>
</tr>
<tr>
<td>MEx29 = score ~ transfer + p.math + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>14</td>
<td>341.76</td>
<td>0.90</td>
</tr>
<tr>
<td>MEx22 = score ~ year + transfer + p.math + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>12</td>
<td>341.99</td>
<td>1.13</td>
</tr>
<tr>
<td>MEx24 = score ~ transfer + gender + p.math + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>12</td>
<td>342.69</td>
<td>1.83</td>
</tr>
<tr>
<td>MEx8 = score ~ year + transfer + p.math + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>15</td>
<td>343.42</td>
<td>2.56</td>
</tr>
</tbody>
</table>

Relative importance of each variable
p.math: 1.00
transfer: 1.00
Pmajor: 0.24
year: 0.22
gender: 0.15
on visual comparisons from the following graphs, not from statistical analyses, to interpret the best-fit model described above. In general, students who had completed higher-level mathematics classes performed better on the pre-test, but this pattern did not hold true for students who had completed Math 240: statistics (Fig. 7). It appears that students who had been at Western for their entire undergraduate careers scored generally lower on the pre-test than students who had transferred from another four-year university or two-year college (Fig. 8). Evidently students in all majors scored similarly on the pre-test except for English majors and students who selected “other.” These students scored lower than those in the other majors (Fig. 9). Sophomore and juniors seemed to perform slightly worse on the pre-test than freshmen and seniors, but the sample sizes of freshmen and seniors were quite small and this factor did not contribute strongly compared to the others described above (Fig. 10). Gender was the least important variable in the model and graphical representation of these data revealed very little difference between pre-test performances of the two groups (Fig. 11).

Difference between pre- and post-test scores (research objective four)

The GLMM analysis identifying population demographic factors that predicted students’ changes in basic quantitative skills and abilities to apply them to biological problems determined that the top three models competed as “best-fit” (Table 11). I averaged these three models to determine the relative importance of each factor. The averaged model included the same suite of fixed factors as for the above analysis: previous math, transfer history, major, test, year and gender (test was the only variable included in this analysis that was not included in the previous analysis), but these variables differed in their importance. It also included the same random
Figure 7. Pre- and post-test scores of students with the highest college-level mathematics class they have completed, and the change between pre and post-test scores for these groups, represented by normalized gains. Horizontal lines represent the median score, and box limits represent the first quartile (Q1) and third quartile (Q3). Dashed whiskers indicate the most extreme data points within 1.5*IQR where IQR is the interquartile range: Q3-Q1. The points represent outliers which are outside 1.5*IQR. Numbers in boxes represent sample sizes.
Figure 8. Pre- and post-test scores of students with their transfer history, and the change between pre and post-test scores for these groups, represented by normalized gains \([\text{post-pre}]/(16-\text{pre})\). Horizontal lines represent the median score, and box limits represent the first quartile (Q1) and third quartile (Q3). Dashed whiskers indicate the most extreme data points within 1.5*IQR where IQR is the interquartile range: Q3-Q1. The points represent outliers which are outside 1.5*IQR. Numbers in boxes represent sample sizes.
Figure 9. Pre- and post-test scores of students arranged by major groups, and the change between pre and post-test scores for these groups, represented by normalized gains. Horizontal lines represent the median score, and box limits represent the first quartile (Q1) and third quartile (Q3). Dashed whiskers indicate the most extreme data points within 1.5*IQR where IQR is the interquartile range: Q3-Q1. The points represent outliers which are outside 1.5*IQR. Numbers in boxes represent sample sizes.
Figure 10. Pre- and post-test scores of students with their year, and the change between pre and post-test scores for these groups, represented by normalized gains \([(\text{post-pre})/(16-\text{pre})]\). Horizontal lines represent the median score, and box limits represent the first quartile (Q1) and third quartile (Q3). Dashed whiskers indicate the most extreme data points within 1.5*IQR where IQR is the interquartile range: Q3-Q1. The points represent outliers which are outside 1.5*IQR. Numbers in boxes represent sample sizes.
Figure 11. Pre- and post-test scores for males and females. Horizontal lines represent the median score, and box limits represent the first quartile (Q1) and third quartile (Q3). Dashed whiskers indicate the most extreme data points within 1.5*IQR where IQR is the interquartile range: Q3-Q1. The points represent outliers which are outside 1.5*IQR. Numbers in boxes represent sample sizes.
Table 11. Top five models for population demographics and quantitative skills analysis evaluating the factors that predict differences between pre- and post-test scores. The gray box highlights the top three models averaged together to make the best-fit model. The relative importance of each variable in the averaged model is included at the bottom of the table; higher numbers indicate more importance. The top four other models are listed in order from best to worst fit; lowest to highest Akaike Information Criterion (AIC) value. All models tested are included in Appendix J. df represents degrees of freedom for each model, delta AIC is the difference between the given model’s AIC score and the best model’s AIC score. Evidence ratios compare the best-fit model to the other top five models; these values indicate how much more likely the best-fit model is than the given model.

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>AIC</th>
<th>∆AIC</th>
<th>Akaike Weight (w_i)</th>
<th>Evidence Ratio (w_1/w_i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEx29 = score ~ test + transfer + p.math + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>15</td>
<td>786.44</td>
<td>0.00</td>
</tr>
<tr>
<td>MEx8 = score ~ test + year + transfer + p.math + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>16</td>
<td>786.92</td>
<td>0.48</td>
</tr>
<tr>
<td>MEx10 = score ~ test + transfer + gender + p.math + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>16</td>
<td>788.39</td>
<td>1.95</td>
</tr>
<tr>
<td>MEx1 = score ~ test + year + transfer + gender + p.math + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>17</td>
<td>788.83</td>
<td>2.39</td>
</tr>
<tr>
<td>MEx4 = score ~ test + year + transfer + p.math + Pmajor + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>19</td>
<td>791.36</td>
<td>4.92</td>
</tr>
</tbody>
</table>

Relative importance of variables in averaged model
p.math: 1.00
transfer: 1.00
Pmajor: 1.00
test: 1.00
year: 0.36
gender: 0.17
factors as above: quarter and lab section. Previous math, transfer history, major, and test were equally important. Year was slightly less important and gender was again the least important factor (Table 11).

As above, I based the following descriptions on visual comparisons from the graphs, not from statistical analyses, to interpret the best-fit model described above. In general, students who had completed any level of mathematics class improved the same amount from pre- to post- test (Fig. 7). It appears that students in the transfer history groups performed more similarly on the post-test as compared to the pre-test (Fig. 8). Students who did not transfer had larger gains, students who transferred from a university had negative gains, while students who transferred from a two-year college stayed approximately the same (Fig. 8). Evidently, students in the biology majors scored higher on the post-test than students in all the other groups of majors (Fig. 9). Freshmen performed best on the post-test, sophomores and juniors second best, and seniors worst, though sample sizes of freshmen and seniors were small and year was not as important a factor as the previously described factors (Fig. 10). Though the averaged best-fit model again included gender, it was the least important of all variables, and graphical representation of the data revealed very little difference between males and females from pre- to post- test except that they both improved slightly (Fig. 11).

Using graphical representations, many of the differences between pre- and post- tests for different groups of students were very small, but these variables still provided enough predictive power to be included in the best-fit model, which means they are important at some level.
Student opinions

Contingency table analyses of student responses to several survey questions revealed their attitudes towards mathematics in biology, and the experimental activities. The majority of students in both treatment groups answered, “yes” on both the pre- and post-tests to the question: “are quantitative skills important for biology?” (Table 12). However, approximately two thirds of students in both the treatment and control groups responded “no” on the pre-test when asked, “should Biology 205 include more math?” In the treatment group this number was roughly the same on the post-test, but in the control group, significantly more students responded, “yes” to this question on the post-test. On the post-test, significantly more students in the treatment group than the control group thought the activities helped improve their quantitative skills, though it was still less than half of the students in the treatment group (Table 12).
Table 12. Contingency table comparing student responses to opinion questions. I used G-tests of independence to make comparisons between the pre- and post- test responses for each treatment group for the first two questions, and between the treatment and control group for the third question because it was only on the post-test. There were two degrees of freedom for each analysis and I adjusted alpha to 0.01 to correct for multiple comparisons. Significant results are highlighted with the gray box.

<table>
<thead>
<tr>
<th></th>
<th>(1). Are quantitative skills important for biology?</th>
<th>(2). Should Biology 205 include more math?</th>
<th>(3). Do you think the activities helped improve your quantitative skills?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-test Post-test</td>
<td>Pre-test Post-test</td>
<td>Post-test</td>
</tr>
<tr>
<td><strong>TREATMENT</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>85 84</td>
<td>28 36</td>
<td>42</td>
</tr>
<tr>
<td>No</td>
<td>2 6</td>
<td>54 53</td>
<td>47</td>
</tr>
<tr>
<td>NA</td>
<td>4 1</td>
<td>9 2</td>
<td>2</td>
</tr>
<tr>
<td>G-statistic</td>
<td>4.03</td>
<td>5.83</td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td>0.1333</td>
<td>0.0542</td>
<td></td>
</tr>
<tr>
<td><strong>CONTROL</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>91 96</td>
<td>30 69</td>
<td>21</td>
</tr>
<tr>
<td>No</td>
<td>5 3</td>
<td>62 27</td>
<td>77</td>
</tr>
<tr>
<td>NA</td>
<td>4 1</td>
<td>8 4</td>
<td>2</td>
</tr>
<tr>
<td>G-statistic</td>
<td>2.57</td>
<td>31.29</td>
<td>14.04</td>
</tr>
<tr>
<td>p-value</td>
<td>0.2767</td>
<td>&lt;0.0001</td>
<td>0.0009</td>
</tr>
</tbody>
</table>
DISCUSSION

Improving quantitative skills

The best-fit model describing the assessment results from both quarters included an interaction between test, achievement, and construct. Graphical exploration showed that low-achieving students (achievement) improved from pre- to post-testing (test) at the questions that integrated quantitative skills with biology concepts (construct). I expected students in the treatment group to improve from pre- to post-testing, but the presence of this trend in one group of students (low-achieving students) still provided us with information about how the Biology 205 course affected students’ quantitative skills. Since the treatment and control groups both took the same Biology 205 course, it seems likely that low-achievers improved at applying their quantitative skills to biological problems because of the Biology 205 course. The trend implies that the Biology 205 course targeted this construct at a most basic level. Students in one achievement group (low-achieving) seemed to improve with the level of material addressed in Biology 205, but it was likely too easy for high-achieving students. This is a positive finding, even though it was not due to the treatment, because it implies that lower-achieving students improved at their ability to recognize quantitative concepts underlying biological problems.

Though the best-fit model for predicting whether or not students answered questions correctly did not include treatment, the second best-fit model did include an interaction between treatment and test (Table 8). Also, students in the treatment group appeared to improve from pre- to post-testing. As explained below, we saw this trend even though the best-fit model did not include an interaction with treatment.

When assigning Biology 205 lab sections to the treatment and control groups, I haphazardly selected sections from each professor to assign to each group. I assumed that
students with better and worse quantitative skills would be randomly distributed amongst the different lab sections, and therefore students in the control and treatment groups would perform similarly on the pre-test. There were no differences between the control and treatment groups on the pre-test in the winter quarter, but in the spring students in the treatment group scored significantly higher on the pre-test than those in the control group. The sample size of the spring quarter treatment group was also smaller than all other groups, which may have contributed to this difference. If students in the treatment group in the spring quarter had generally better quantitative skills than all other students in the study, the activities directed to average Biology 205 students would likely not make a difference. The difference between pre-test scores in the spring quarter is one possible reason that treatment was not included in the overall best-fit model.

The averaged best-fit model describing the assessment results from only the winter quarter included three contributing models. The first contributing model included an interaction between test and achievement, the second contributing model included an interaction between test, treatment, and achievement, and the third contributing model included an interaction between test, achievement and construct. Graphical exploration of the second contributing model revealed that low-achieving students in both treatment groups appeared to improve from pre- to post-testing while high-achieving students either did worse or approximately the same. Though it is difficult to see any differences between the treatment groups, this interaction was included in one of the contributing models, which means the treatment gave some explanatory power. This result gives some evidence that the treatment might improve students’ quantitative skills, but it made too small of a difference or we had too small of a sample size to detect any improvement.

Another possible reason treatment was not included in the overall best-fit model is that I did not give students points based on their performance on the assessment or activities, but based
on completion of them. As a result, some students may have put less effort into the study materials than they would have if the materials had counted towards their grades. The students who put in less effort should have been randomly distributed amongst the treatment and control groups. Students in the control group who put less effort into the activities as compared to those who put more effort into the activities should have all performed similarly on the post-assessment because the activities were not designed to improve their assessment scores. However, students in the treatment group who put less effort into their activities as compared to those who put more effort into their activities would likely not have improved on the post-assessment relative to the pre-assessment because those activities were designed to improve their assessment scores. Lack of student effort, even randomly distributed, would likely have caused the treatment and control groups to perform similarly on the post-test.

**Ways to improve the treatment activities**

I examined the assessment results for the specific quantitative skill areas (exponents, logs, graphs, and fractions), and students’ responses to the activities to infer areas for improvement. Based on patterns emerging from this study, I have several recommendations for improvement to the treatment that should be evaluated in the future.

Students in both the treatment and control groups did relatively well on the exponent questions, but those in the treatment group did not improve on the post-test relative to those in the control group. Students in the treatment group also scored very high on the activity about exponential growth and decay (98% average). These results imply that the treatment activity may have been too easy to teach students new material, and even if it had, the assessment questions may have been too easy to detect changes. Students did struggle in this area on the preliminary
assessment, so their ease with the activity was likely not because they knew everything about exponents, but because the activity did not sufficiently address and improve upon their existing knowledge. In the future, I recommend altering this activity (and perhaps the assessment questions) to be more challenging.

Students in the control group scored higher on the post-test questions about logarithms than they did on the pre-test, while those in the treatment group did not improve. However, students in the control group did score lower on the pre-test than students in the treatment group, so the post-test scores of both groups were similar. Students in the control group may have improved due to random chance, increased skill from the Biology 205 course, or increased effort after reading about the importance of quantitative skills for biology. Students in the treatment group scored poorly on the activity about logarithms (62% average). These results imply that the activity may have been too difficult, and was not very effective.

The activity about logarithms was also the least interactive of all the activities. Students worked through a very short MathBench lesson, reviewed basic rules about logarithms online (like an online textbook), and then worked through some practice problems. They did poorly on these practice problems, which implies that they did not retain the material after reviewing it passively. Research has indicated that interactive online material is more effective for teaching than online books (Paay & O’Brien 2000; Moreno et al. 2001; Zhang 2005). A future activity should include more interactive practice with the logarithm rules as part of the activity.

Students in both the control and treatment groups generally scored higher on the post-test questions about graphing than those pre-test questions. It seems likely that the Biology 205 course improved student competence in this area. Another possible explanation is that the treatment activity and essays about the importance of quantitative skills in biology both
improved scores equally effectively. Compared to the activities about exponents and logarithms, the activity about graphing appeared to be more appropriately matched to students’ incoming skill-set (81% average). These scores imply that the activity itself was more appropriately challenging than the other two activities.

Students in both treatment groups did well on the fractions questions, but students in the treatment group did not improve on the post-test relative to those in the control group for this set of questions. There was no specific treatment activity targeted to improve this skill, but most of student mistakes were related to simple mathematical errors and following directions, which I hoped to address with increased time spent using quantitative skills. In the future, I suggest adding an activity targeting the use of fractions, ratios, proportions, and percentages within a biological context to improve students’ comfort with this skill, and reduce their number of simple mathematics errors.

**Ways to improve the treatment design**

There are several differences between the implementation of MathBench biology modules at UMD (Thompson et al. 2010) and this study that may have resulted in the modules’ decreased efficacy for this university’s students. I will use a comparison of the two studies to make recommendations for improvements to the treatment design that might make it more effective at Western Washington in the future.

Thompson et al. (2010) incorporated nine of the 37 MathBench modules into UMD’s equivalent course to this university’s Biology 205 (BSCI 105 Principles of Biology 1: Cell and Molecular Biology). Western is on a quarter-schedule so all courses are approximately one third shorter than semester-long courses, and I only selected portions of three of the MathBench
modules to include in Biology 205. At UMD, they better integrated MathBench with the rest of
the course; students conducted weekly lab exercises complementary to particular MathBench
lessons and took follow-up quizzes (Thompson et al. 2010). In our study, the MathBench
activities were separate from both the lecture and lab portions of the course. Previous research
has indicated that when students practice concepts in multiple contexts, for example on the
computer and in the lab, they learn more effectively (Richter & Timm 2005). Giving students
feedback about their understanding has also been shown to help students learn (Richter & Timm
2005). I did not give students feedback about their performance on the activities because time
and resources were too limited. I encourage using undergraduate teaching assistants to help grade
the treatment activities in the future.

In Thompson and colleagues’ (2010) study, the lab exercises and quizzes following a
MathBench activity were graded and counted towards a student’s overall grade in the course. As
discussed above, I gave study participants points for completing the assessments and activities,
but not for their performance. I graded this way to avoid giving students an incentive to cheat,
and to avoid potentially giving students bad grades due to problems with the materials under
study. Now that the assessment and activities have been studied, can be improved, and can be
overseen to prevent cheating, I recommend that professors give students grades based on
performance rather than completion. Grades are commonly used in most educational systems and
all educational levels to evaluate student learning and encourage students to put effort into
learning class material. Studies also show that when students take regular quizzes covering the
subject-material in a course, they learn it better (Pennebaker et al. 2013). I recommend giving
regular quizzes based on the material in the treatment activities and grades based on correctness
to try to increase the effectiveness of these activities.
At UMD, students who took the course with MathBench were more comfortable using quantitative skills to solve biological problems and were very positive (83% of responses) about the modules helping improve their quantitative skills (Thompson et al. 2010). Fewer than half of the students who completed MathBench modules at this northwest university thought their quantitative skills were improved by the activities. MathBench modules may not have been effective in our study because we did not use as many of them or integrate them as well into the rest of the course. One goal of this study was to evaluate if a small portion of supplementary material would improve students’ basic quantitative skills. Based on a large body of evidence from the literature, it appears that a more integrative approach would likely be more effective (Metz 2008; Depelteau et al. 2010; Duncan et al 2010; Miller & Walston 2010; Thompson et al. 2010; Colon-Berlinger & Burrowes 2011).

Contingency table analyses of student responses to the opinion questions revealed that a significant number of students in the control group changed their answer from “no” on the pre-test to “yes” on the post-test in response to the question: “should Biology 205 include more math?” Like the treatment group, students in the control group did three activities, except their activities were reading essays and answering follow-up questions about the relationship between mathematics and biology. One essay was about why quantitative skills are important for biological research, the second essay was an excerpt from BIO2010 recommending that institutions improve the quantitative curriculum of undergraduate biology courses, and the third essay was an example of the successful addition of quantitative material to a biology course (Appendix H). After reading these essays, students in the control group seemed to better understand the importance and effectiveness of adding quantitative components to a biology undergraduate course, and thought that doing so in Biology 205 would be a good idea. In the
future I recommend first giving students a version of these control activities, and then revised
treatment activities targeted to improve their quantitative skills. If students first understand the
value of the targeted quantitative skills activities, they will likely be more receptive to the
activities and put more effort into learning the material. Studies show that when students are
guided to engage in meta-cognition (awareness of and participation in the learning process) they
learn better (Donovan et al. 2000).

**An additional skill to include: statistics**

Though the faculty did not identify this skill as one of the most critical for undergraduate
biology students at Western, biology research increasingly demands statistically literate students,
and universities are recognizing that their undergraduates severely lack statistics skills (Peterson
2000; Metz 2008; Depelteau et al. 2010; Speth et al. 2010; Colon-Berlinger & Burrowes 2011).
Many universities have conducted studies incorporating statistics into introductory-level biology
courses, and have had positive results (Peterson 2000; Casem 2005; Metz 2008; Depelteau et al.
2010; Speth et al. 2010; Colon-Berlinger & Burrowes 2011). Colon-Berlinger & Burrowes
(2011) added a statistics pre-requisite to the core biology courses at the University of Puerto
Rico, and modified labs in two courses to include statistical analyses of data the students
collected in hopes of making the biological applications of statistics clearer. Students’ statistics
skills improved and they better understood why those skills are important tools for biology
researchers.

Ronsheim et al. (2009) revised the existing introductory-level biology course at Vasser
College to include inquiry-based lab modules that required statistics. These modules required
students to develop testable hypotheses, collect data, analyze those data, and present the results
(Ronsheim et al. 2009). The revamped curriculum effectively prepared students for more advanced courses within the major, and increased the major’s popularity at the school (Ronsheim et al. 2009). Metz (2008) and Depelteau et al. (2010) similarly incorporated statistics into introductory-level biology courses with positive results. I recommend adding a statistics activity that includes one or two MathBench modules, and incorporating statistics components into at least some Biology 205 lab write-ups.

**Population demographics and quantitative skills analysis**

I hesitate to draw strong conclusions from the results of the population demographics analyses because the large number of factors investigated resulted in a large number of models tested (60 for each question), and because there were small differences between pre- and post-test scores for most groups. Nevertheless, we have identified some factors important for predicting (1) quantitative skills of Biology 205 students at this university and (2) how they changed from the pre- to post-assessment.

The two best-fit models for the population demographics analyses contained the same suite of factors, but these factors differed in importance between the two analyses. The factors included were previous math, transfer, major, gender and year. I expected students who had taken higher college-level mathematics classes to perform better on the assessments. That this pattern was generally true indicated that this university’s mathematics courses prepared students to use quantitative skills in biology courses and/or students who had completed higher college-level mathematics courses were generally more comfortable using quantitative skills. Students who had completed the calculus series through Math 125: calculus II performed better on both assessments than students who had only completed lower-level courses, even though the
assessments did not include any topics advanced enough to require calculus II. It seems most likely that this factor was an important predictor of assessment score because students who had taken more mathematics classes before taking Biology 205 were generally more comfortable using quantitative skills. Above Math 125, the more advanced mathematics classes did not appear to improve assessment scores. Students who had completed Math 240: statistics did not do any better on the assessment. Though Math 240 is a higher-level class, it has lower-level mathematics pre-requisites than all the other math courses besides Math 114 (Math 112 is the prerequisite for both courses). The assessment also did not address statistics so I did not expect students in this course to perform better on the assessment. The skills tested by the assessment were basic enough that if students had not learned them by calculus II, higher-level mathematics courses did not appear to help. This result is further evidence for the need to target basic quantitative skills in the introductory-level biology courses.

I expected students who had transferred to Western from two-year colleges to perform worse on both assessments than students who had not transferred to Western or who had transferred from other four-year universities because other studies have found that students who transferred from two-year colleges performed worse in four-year universities than non-transfers or students who transferred from other four-year universities (Laband & Hanby 2003; Asarta et al. 2013). Asarta et al. (2013) found that in an intermediate-level economics class students who transferred from two-year colleges did significantly worse than their fellow students, which included some who had transferred from other four-year universities. To my knowledge nobody has published studies about the academic success of students in biology classes comparing these groups, but Western Washington University’s Biology Department has unpublished data on the subject. Students who transferred from two-year colleges with GPAs in the 2.7-3.4 range for
prerequisite courses did not perform very well in the 300-level biology courses compared to non-transfer students with this past GPA range (Acevedo-Gutiérrez 2004). However, another previous study showed that a higher percentage of students who transferred from a two-year college in California graduated than students who did not transfer (Jones & Lee 1992). In my study I was surprised to find that transfer students from two-year colleges and transfer students from four-year universities performed better on the assessment than non-transfer students, though the groups of transfer students were much smaller than those who had not transferred. These results indicate that regional two-year colleges offer high-quality pre-requisite classes to Biology 205. Though I did not survey where transfer students were from, 90% of Western students are Washington State residents (Admissions 2012).

I expected students in the natural and physical sciences besides biology to perform better on the assessments because education in these fields has historically been more quantitative than biology education. However, students in the biology majors improved most from pre- to post-testing. Half of the assessment questions tested a student’s ability to use quantitative skills to solve biological problems or vice versa, so it follows that biology majors would be more familiar with the biological aspects of these questions, and be most affected by the biology course. While improvement of students within the major is positive, students in a lot of other majors take the introductory-level biology series so it would be better to target the courses to a slightly wider audience, while still teaching the essential biology concepts.

Both averaged best-fit models also included year, though this variable was much less important than the others included in the models. I expected students in later years of school to perform better on the assessment(s), but on the pre-test sophomores and juniors performed worse while freshmen and seniors performed better. It is possible that these groups of students scored
higher because students who could take Biology 205 as freshmen (meaning they had completed required pre-requisites) were generally high-achieving, and seniors had taken more college-level courses. Seniors actually performed worse on the post-test, which may have been because they had more academic and non-academic responsibilities towards the end of the quarter. The sample sizes for both the freshmen and seniors were relatively small and this factor did not contribute very strongly to either averaged model, so the disparate sample sizes may have driven the trends. The averaged best-fit models for both population demographic analyses included gender, but it was the least important of all included variables, and I cannot see any differences between the groups with graphical representation of the data.

**Study contributions**

My study has contributed to this northwest university because even though the treatment did not appear to improve students’ basic quantitative skills or ability to transfer them to biology, we began searching for ways to do this. It was also valuable to learn that the simple intervention we attempted was not as effective as we had hoped it would be. This study was the first Discipline Based Education Research (DBER) project, an emerging field of research, conducted by a graduate student in this university’s Biology Department in at least two decades (NRC 2012). It will hopefully encourage future collaborative work between scientists and science educators at this university and other institutions.

My study also has the potential to contribute to the science education field because I used generalized linear mixed-effect models (GLMMs) to analyze my data. These robust, powerful, and flexible tests are becoming increasingly popular in some fields of biology research (Bolker et al. 2008). To my knowledge nobody has yet used them for biology education research, though
they are likely also appropriate for these data (Freeman per. comm.). GLMMs were the most appropriate tool for answering my research questions, and therefore my project has the potential to become a model for future science education researchers to use these analyses.

Though our intervention did not improve students’ quantitative skills, it did not impair them either, and previous studies have shown that adding quantitative concepts to biology courses did not harm biological learning. At the University of Puget Sound (UPS), Madlung et al. (2011) added a quantitative learning module to two biology courses to evaluate the potential negative effects of this intervention, predicting that adding quantitative components to the course would prevent students from retaining the biology material because there would be too much information and/or the quantitative components would cause anxiety. Results indicated no reduction in biology comprehension and improvement in mathematics, which supports adding this kind of curriculum module to undergraduate biology courses. Though our treatment did not improve students’ basic quantitative skills or their ability to apply them to biology, we have learned a lot that will hopefully be used at this university and elsewhere for continued research. We hope that future studies continue working to improve the basic quantitative skills of undergraduate biology students to better prepare these students for continued education, the research field, and to become better educators for the next generation of biologists.

**Future work**

I recommend that graduate students and faculty continue working together on this project with some improvements to the treatment that will likely make it more effective. Because students in the control group thought that Biology 205 should include more math after they

---

4 Scott Freeman University of Washington Biology Department, Box 351800 Seattle, WA 98195, June 2012.
completed the control activities, I recommend assigning students readings similar to these control activities before administrering the treatment activities. Perhaps they could read the papers in the first couple weeks of the quarter or they could read them in Biology 204. If they are more open to the treatment activities, it is likely that they will benefit more from them. I also recommend incorporating more of the MathBench biology modules into the Biology 205 course; perhaps one per week instead of one every other week as I did. Alternatively, the modules could be incorporated into all three introductory-level courses to better space them throughout one academic year. This design might also allow the activities to provide more continuity between the three courses. I also recommend that if faculty time and resources permit, they incorporate skills from the MathBench modules into lab exercises so that students have the opportunity to practice the skills in multiple contexts. Faculty could also or alternatively administer graded quizzes to follow-up the MathBench activities because quizzes improve student learning and grades provide incentive to put more effort into the material (Pennebaker et al. 2013).

I recommend that future studies also revise two of the treatment activities to make the activity about exponents more difficult and the activity about logarithms more interactive. I also recommend including a MathBench statistics module as one treatment activity because understanding basic statistics is important in many biological fields. It would also be relatively simple to add statistical components to existing labs and there are quite a few resources available from other institutions that have done so. It would be valuable for a future study to create an assessment to evaluate students’ learning of the biology material, much like Madlung et al. (2011) to evaluate if the added mathematics material detracts from student learning of the biology material. Transfer is a person’s ability to apply knowledge or skills to multiple different contexts, and this process is often difficult for students (Jacobson & Archodidou 2000; Newman
2012). A future study could also more directly study students’ abilities to transfer quantitative concepts to biological contexts.

Most researchers wish to increase their sample size and cannot due to various constraints. With this study system it would be easy to increase the sample size by conducting the study over a longer period of time (one or two academic years) and/or by including the other introductory-series courses in the study. Finally, if future studies collect data about the population demographics of introductory-level biology students, we will continue gathering valuable information about this population of students and important predictors of their quantitative skills.
REFERENCES


http://admissions.wwu.edu/contact/at_a_glance.html (accessed 17 November 2013).


APPENDIX A

Faculty Quantitative Skills Survey

Below is a list of skills, from a previous faculty survey, identified as important basic quantitative skills for biology undergraduates. From this list, we are trying to identify the three most critical basic skills for undergraduate students in our core courses (skills that faculty feel our students lack).

Please rank them in order of most critical (1) to least critical (3) and include a list of core and breadth courses you teach.

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Skill</th>
</tr>
</thead>
<tbody>
<tr>
<td>_______</td>
<td>Using the metric system</td>
</tr>
<tr>
<td>_______</td>
<td>Using scientific exponents, including scientific notation</td>
</tr>
<tr>
<td>_______</td>
<td>Reading and analyzing data, including graphical presentation and interpretation</td>
</tr>
<tr>
<td>_______</td>
<td>Basic numeracy/arithmetic, order of operations, measured numbers</td>
</tr>
<tr>
<td>_______</td>
<td>Basic algebra</td>
</tr>
<tr>
<td>_______</td>
<td>Enough calculus to know that dN/dt is instantaneous rate of change</td>
</tr>
<tr>
<td>_______</td>
<td>Sense of the size of numbers</td>
</tr>
<tr>
<td>_______</td>
<td>Using dimensional analysis</td>
</tr>
<tr>
<td>_______</td>
<td>Using and manipulating fractions/proportions/ratios/percentages</td>
</tr>
<tr>
<td>_______</td>
<td>Using and manipulating logarithms</td>
</tr>
<tr>
<td>_______</td>
<td>Other?</td>
</tr>
</tbody>
</table>

Your name:

Core/Breadth classes you teach:

Thanks you for your input. Please return this survey electronically to:
dethiel@students.wwu.edu
Please read the directions CAREFULLY, and work independently. Each question will have instructions for whether or not you should use a calculator.

1. You may use a calculator for this question. There has been a recent outbreak of SARS (Severe Acute Respiratory Syndrome) in Singapore. As of today, 30 cases have been reported. If the virus has an infection rate of 4% per day, meaning the number of infected individuals increases by 4% each day, how many individuals will have the disease in two weeks? Round to the nearest individual. *52 120 47 198 I don’t know how to approach this problem

2. You may use a calculator for this question. You put $500 into a savings account with an annual interest rate of 1.6%. How much money will you have in 20 years assuming you don’t add any money to the account? Round to the nearest dollar. *$687 $800 $660 $2123 I don’t know how to approach this problem

3. Please do not use a calculator for this question. A researcher studying rats models their population size with the equation N=Ie^{0.2t} where I=initial population size, and N=population size after time (t) in weeks. If Population A’s initial size is 5 times larger than population B’s, how will their populations differ after the same amount of time? *Population A will be 5 times larger than population B Population A will be e^5 times larger than population B Population A will grow exponentially while population B grows linearly Population A’s growth rate will be 5 times larger: e^t, while population B’s growth rate will stay the same: e^{0.2t} I don’t know how to approach this problem

4. Please do not use a calculator for this question. \( x = a5^b \). What will happen to \( x \) if you multiply \( a \) by 3? *\( x \) will be three times larger \( x \) will increase by 3 \( x \) will be \( 5^3 \) times larger \( x \) will equal \( a5^{3b} \) I don’t know how to approach this problem
5. Please do not use a calculator for this question. A researcher is studying the growth rate of mice. She discovers that in the first three months of life body length $= k\log(\text{weight})$ where length is in centimeters, $k$ is a known constant, and weight is in grams. Which of the following is true?
* If you multiply mouse weight by 10, body length increases by $k$ cm
If you multiply mouse weight by 10, body length is multiplied by AT LEAST 10 cm
If you multiply mouse weight by 10, body length is multiplied by EXACTLY 10 cm
At very high mouse weights, there is no increase in body length with increasing mouse weight
I don’t know how to approach this problem

6. Please do not use a calculator for this question. Expand the expression: $a\log_5(5^b)$
* $a + a\log_5b$
> $5(a\log_5b)$
$5(a\log_5b)$
When $b$ is very large, $a\log_5b = a\log_5b$
I don’t know how to approach this problem

7. Please do not use a calculator for this question. A student working in the Johnson-Williams lab is studying the effects of a new antibiotic. She discovers that after one week $\log_{10}/\log_3 27$ of her bacteria have died. She is getting ready to give these results to her adviser when one of her lab-mates looks over and says, “Hey, you can simplify that fraction”. What fraction of bacteria died from the antibiotic?
* $1/3$
$1/17$
$1/\log_3 27$
$\log_{10}/\log_3$
I don’t know how to approach this problem

8. Please do not use a calculator for this question. Calculate $\log_{100}/\log_{10}$.
* $2$
$10$
$1$
$\log_{10}$
I don’t know how to approach this problem

9. Please do not use a calculator for this question. In the following scenario, what would be the best way to graph the experimental results? A biologist is interested in the effect of elevation on plant growth. She collects native blackberry plants from 10 different altitudes along Mt. Baker highway, starting in Bellingham and increasing by 200 ft in elevation between each site. Back in the lab, she measures dry plant weight.
* Scatterplot of “plant weight” on the y-axis and “elevation” on the x-axis with a best-fit line through the points
Line graph of “plant weight” on the y-axis and “elevation” on the x-axis
Bar graph comparing “plant weight” from the 5 lowest sites to “plant weight” in the 5 highest sites
Bar graph of “plant weight” for each elevation
I don’t know how to approach this problem

10. Please do not use a calculator for this question. In the following scenario, what would be the best way to graph these data? You work for the Washington State Department of Transportation, and you’re interested in the effect of car color on number of traffic accidents. You pull up the records of traffic accidents in 2010 and copy down the number of red, silver, black, and blue cars in accidents.
Bar graph of “number of accidents” for each “car color”
Bar graph comparing “number of accidents” for the light-colored cars to “number of accidents” for the dark-colored cars
Scatterplot of “number of accidents” vs. car color with a best-fit line through the points
Line graph of “number of accidents” vs. “car color”
I don’t know how to approach this problem

11. Please do not use a calculator for this question. The graph below expresses the relationship between price per pound of apples, and the number of days the apples ripened. Which of the following statements is true?

*The slope of the graph is 5 cents/pound per day, meaning that apples increase in value by 5 cents/pound for every day they are allowed to ripen
The y-intercept of the graph is 20 cents/pound, meaning that apples increase in value by 20 cents/pound for every day they are allowed to ripen
The y-intercept of the graph is 20 cents/pound, meaning that the average cost of apples is 20 cents/pound
The slope of the graph is 0.5 cents/pound per day, meaning that apples increase in value by 0.5 cents/pound for every day they are allowed to ripen
I don’t know how to approach this problem

12. Please do not use a calculator for this question. What is the slope of the line MN?
*6/7
1
7/6
13. You may use a calculator for this question. Ranger Roger has three days to count the 1200 Douglas fir trees in Arroyo Park so he needs to count one third of them per day. He counted this number on the first day, but due to lightning storms only counted 60% as many on the second day. What simplified fraction of the total number of fir trees has he counted after the first two days?

*8/15
640
about 1/2
2/5

I don’t know how to approach this problem

14. You may use a calculator for this question. What is 75% of 280 plus 1/3 of 630? Divide by 900 and simplify.

*7/15
21/45
6/10
49/120

I don’t know how to approach this problem

15. You may use a calculator for this question. In a herd of 970 zebras, about three-fifths of two percent of all animals are injured crossing a stream. How many are injured? Round to the nearest whole animal.

*12
I don’t know how to approach this problem

16. You may use a calculator for this question. What is 5% of 1/4 of 540? **ROUND** to the nearest whole number.
*7
6.75
67.5
68
I don’t know how to approach this problem
APPENDIX C

Basic Quantitative Skills Survey
Winter & Spring pre- test

Pre-test
1. Write the last four digits of your student ID, Birth date, and Birth month (as on consent form). For example, 84930703 where:
   8493=last 4 digits of ID
   07=birth date
   03=birth month

Post-test
1. What’s your full name? This will be used to anonymously code your survey results with your assessment results.

Pre & Post-tests
2. What year are you in school?
   Freshmen
   Sophomore
   Junior
   Senior
   Post-baccalaureate

3. What’s your major or intended major?
   General Biology
   Biology with Ecology, Evolution & Organismal emphasis
   Biology with Cell emphasis
   Biology with Marine emphasis
   Biology with Molecular and Cell emphasis
   Biology with Teaching emphasis
   Chemistry
   Biochemistry
   Kinesiology
   Environmental Science/Studies
   Behavioral Neuroscience
   Psychology
   Community Health
   Anthropology
   English
   Other

4. Did you transfer to Western?
   No
   Yes, from a Community College
   Yes, from a Four-Year University
Yes, from a Community College and Four-year University

5. What is the highest-level math class you have completed? If completed at another school, select the WWU equivalent based on the course title.
   112 Algebra
   114 Precalculus I
   115 Precalculus II
   118 Accelerated Precalculus
   124 Calculus I
   125 Calculus II
   135 Calculus II Honors
   204 Elementary Linear Algebra
   224 Multivariable Calculus & Geometry I
   226 Limits and Infinite Series
   240 Introduction to Statistics
   300 level or higher

6. Are you currently enrolled in a math course? If so, what?
   None
   112 Algebra
   114 Precalculus I
   115 Precalculus II
   118 Accelerated Precalculus
   124 Calculus I
   125 Calculus II
   135 Calculus II Honors
   204 Elementary Linear Algebra
   224 Multivariable Calculus & Geometry I
   226 Limits & Infinite Series
   240 Introduction to Statistics
   300 level or higher

7. What is your gender?
   Male
   Female
   I don’t want to answer this question

8. What is your ethnicity?
   Caucasian
   Hispanic or Latino
   Black or African American
   American Indian/Alaska Native
   Asian/Asian-American
   Hawaiian/Pacific Islander
   I don’t want to answer this question
9. Do you think math is important for biology, and why or why not?

10. Do you think biology courses should include more math concepts? If so, which ones?

Post-Test

11. Do you think the activities of this past quarter improved your scores on this assessment?

12. What did you think of the activities this quarter? Did you like them? Ideas for improvement?
Basic Quantitative Skills Survey
Winter post-test

1. What’s your full name? This will be used to anonymously code your survey results with your assessment results.

2. What year are you in school?
   Freshmen
   Sophomore
   Junior
   Senior
   Post-baccalaureate

3. What’s your major or intended major?
   General Biology
   Biology with Ecology, Evolution & Organismal emphasis
   Biology with Cell emphasis
   Biology with Marine emphasis
   Biology with Molecular and Cell emphasis
   Biology with Teaching emphasis
   Chemistry
   Biochemistry
   Kinesiology
   Environmental Science/Studies
   Behavioral Neuroscience
   Psychology
   Community Health
   Anthropology
   English
   Other

4. Did you transfer to Western?
   No
   Yes, from a Community College
   Yes, from a Four-Year University
   Yes, from a Community College and Four-year University

5. What is the highest-level math class you have completed? If completed at another school, select the WWU equivalent based on the course title.
   112 Algebra
   114 Precalculus I
   115 Precalculus II
   118 Accelerated Precalculus
   124 Calculus I
   125 Calculus II
   135 Calculus II Honors
6. Are you currently enrolled in a math course? If so, what?
None
112 Algebra
114 Precalculus I
115 Precalculus II
118 Accelerated Precalculus
124 Calculus I
125 Calculus II
135 Calculus II Honors
204 Elementary Linear Algebra
224 Multivariable Calculus & Geometry I
226 Limits & Infinite Series
240 Introduction to Statistics
300 level or higher

7. What is your gender?
Male
Female
I don’t want to answer this question

8. What is your ethnicity?
Caucasian
Hispanic or Latino
Black or African American
American Indian/Alaska Native
Asian/Asian-American
Hawaiian/Pacific Islander
I don’t want to answer this question

9. Do you think math is important for biology, and why or why not?

10. Do you think biology courses should include more math concepts? If so, which ones?

11. Do you think the activities of this past quarter improved your scores on this assessment?

12. What did you think of the activities this quarter? Did you like them? Ideas for improvement?
APPENDIX D

Undergraduate Participant Consent Form
Improving basic quantitative skills in biology undergraduate students

Researcher: Lilly Dethier, M.S. candidate, Western Washington University
Contact Information: 425-830-2903 dethiel@students.wwu.edu

Adviser: Deborah Donovan, Professor, Western Washington University
Contact Information: 360-650-7251 donovan@fire.biol.wwu.edu

Researcher’s statement
We are asking you to participate in a research study. This consent form will give you information about the study so you can decide whether or not you wish to participate. Please read it carefully and ask any questions you may have about the purpose of the study, what we would ask you to do, possible risks/benefits, your rights as a volunteer, and anything else that may be unclear. Then please decide whether or not you wish to participate in the study, and we will give you a copy of this form for your records. This process is called “informed consent”

PURPOSE AND BENEFITS
The purpose of this study is to improve the basic quantitative skills of biology undergraduates at Western Washington University. We would like to obtain information about BIO 205 students’ understanding of three critical basic quantitative skills: (1) using and manipulating fractions, proportions, ratios, and percentages; (2) reading and analyzing data, including graphical presentation and interpretation; and (3) using and manipulating logarithms and exponents. We also wish to evaluate the effectiveness of an activity targeted to improve understanding of these skills. If you choose to participate, you will share with the researcher your responses to a pre- and post- instruction assessment. You may also receive the targeted course activity, and share your responses to this activity. You may benefit directly from this study with an improved understanding of these basic quantitative skills. You may not benefit directly, but we hope the results of this project will help biology faculty modify their course curricula to help students better understand basic quantitative aspects of biology. Improving the quantitative skills of undergraduate biology students would help better prepare them for research, graduate school, professional school, and other related careers.

PROCEDURES
Fall 2012
If you choose to participate in this study, you will be helping us to develop the assessment. You will share with the researcher your responses to the assessment, which will help to determine if the assessment is reliable and valid. If you participate in this portion of the study, you will receive some candy as a small compensation for your time. You may also be asked to participate in a “talk-aloud” interview, which will take approximately 30 minutes. This entails meeting with a researcher after taking the assessment to explain your thought processes while answering each of the questions. The researcher will make an audio recording of this interview. Talk-aloud interviews help us to determine if the assessment questions truly evaluate the skills we wish to
evaluate. If you participate in the “talk-aloud” interview, you will receive a small gift card as compensation for your participation.

**Winter or Spring 2013**
If you choose to participate in this study, you will be helping us to evaluate the effectiveness of a course activity. You will share with the researcher your responses to the pre- and post-instructional assessments. You will receive a course activity, and share with the researchers your responses to this activity. We may also ask you to share your perception of your math skills and the activity’s effectiveness. You will receive course points for the pre- and post-assessment and activity based on completion. If you choose not to participate in the study, you will still receive points for completion, but we will not include your responses in the study.

Signing this form means you are willing to participate, but not obligated to participate. You will always be given reasonable notice and have the option of refusing to participate at any time.

**POTENTIAL RISKS, STRESSES, OR DISCOMFORT**
You may feel nervous or uneasy when asked about your understanding of quantitative concepts in biology.

**OTHER INFORMATION**
Participating in this study is voluntary. Whether or not you choose to participate in this study will not affect your grade or standing in this course or program. You can decide at any time to stop participating in the study. Information about you will remain confidential. We will code research records to ensure your anonymity. The link between the code and your name will be kept in a secure location, separate from the research information, and only the research team will have access to that information. We will keep the link between the records and your name until 2020 and then destroy the link. If we publish results from this study, we will not use your name.

Liliane E Dethier
Printed name of researcher

Liliane E Dethier__
Signature of Researcher

11/13/2012__
Date

Subject’s statement
I am at least 18 years of age. I have initialed the items below for which I give my permission

________ I give permission for my responses to the assessments of my understanding of basic quantitative skills to be included in future research reports. I understand my name will not be used.

________ I give permission for the information obtained from audio recordings of interviews about the assessment, and why I chose the responses I did, to be included in future research reports. I understand my name will not be used.

________ I give permission for my responses to the course activity to be included in future research reports. I understand my name will not be used.
This study has been explained to me, and I volunteer to take part in this research. I have had a chance to ask questions. If I have questions regarding my rights as a research participant, I can contact Janai Symons, Human Protections Administrator (HPA) at 360-650-3082 or janai.symons@wwu.edu. I have access to a copy of this consent form, and am at least 18 years old.

<table>
<thead>
<tr>
<th>Printed name of Subject</th>
<th>Signature of Subject</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

☐ ☐ ☐ ☐ ☐ ☐ ☐ 
Last 4 digits of student ID Birth date Birth month
Ex: 07 03 (for March 7)

The above information will be used to anonymously identify your survey responses so they can be saved or discarded according to your choices above. **Please fill this out even if you do not initial any choices**
Please read the directions CAREFULLY, and work independently. Each question will have instructions for whether or not you should use a calculator.

1. You may use a calculator for this question. There has been a recent outbreak of SARS (Severe Acute Respiratory Syndrome) in Singapore. As of today, 30 cases have been reported. If the virus has an infection rate of 4% per day, meaning the number of infected individuals increases by 4% each day, how many individuals will have the disease in two weeks? Round to the nearest individual.
   *52
   120
   47
   198

2. You may use a calculator for this question. You put $500 into a savings account with an annual interest rate of 1.6%. How much money will you have in 20 years assuming you don’t add any money to the account? Round to the nearest dollar.
   *$687.00
   $800.00
   $660.00
   $2,123.00

3. Do not use a calculator for this question. A researcher studying rats models their population size with the equation \( N = I e^{0.2t} \) where \( I \) = initial population size, and \( N \) = population size after time \( t \) in weeks. If Population A’s initial size is 5 times larger than population B’s, how will their populations differ after the same amount of time?
   *Population A will be 5 times larger than population B
   Population A will be \( e^5 \) times larger than population B
   Population A will grow exponentially while population B grows linearly
   Population A’s growth rate will be 5 times larger: \( e^t \), while population B’s growth rate will stay the same: \( e^{0.2t} \)

4. Do not use a calculator for this question. \( x = a 5^b \). What will happen to \( x \) if you multiply \( a \) by 3?
   *\( x \) will be three times larger
   \( x \) will increase by 3
   \( x \) will be \( 5^3 \) times larger
   \( x \) will equal \( a 5^{3b} \)

5. Do not use a calculator for this question. A researcher is studying the growth rate of mice. She discovers that in the first three months of life body length = \( k \log \) (weight) where length is in centimeters, \( k \) is a known constant, and weight is in grams. Which of the following is true?
   *If you multiply mouse weight by 10, body length increases by \( k \) cm
If you multiply mouse weight by 10, body length is multiplied by AT LEAST 10cm
If you multiply mouse weight by 10, body length is multiplied by EXACTLY 10cm
At very high mouse weights, there is no increase in body length with increasing mouse weight

**UPLOAD LOGS WITH SUBSCRIPT AS PICTURES**

6. Do not use a calculator for this question. Expand the expression: $a \log_5(5b)$

\[
\begin{align*}
& a + a \log_5 b \\
& > 5(a \log_5 b) \\
& 5(a \log_5 b)
\end{align*}
\]

When $b$ is very large, $a \log_5 b = a \log_5 b$

**UPLOAD LOGS WITH SUBSCRIPT AS PICTURES**

7. Do not use a calculator for this question. A student working in the Johnson-Williams lab is studying the effects of a new antibiotic. She discovers that after one week the following number of her bacteria have died. She is getting ready to give these results to her adviser when one of her lab-mates looks over and says, “Hey, you can simplify that fraction”. What fraction of bacteria died from the antibiotic? $\frac{\log_{10}}{\log_{27}}$

\[
\begin{align*}
& 1/3 \\
& 1/17 \\
& 1/\log_{27} \\
& \log_{10}/\log_{3}
\end{align*}
\]

8. Do not use a calculator for this question. Calculate $\frac{\log_{100}}{\log_{10}}$.

\[
\begin{align*}
& 2 \\
& 10 \\
& 1 \\
& \log_{10}
\end{align*}
\]

9. Do not use a calculator for this question. In the following scenario, what would be the best way to graph the experimental results? A biologist is interested in the effect of elevation on plant growth. She collects native blackberry plants from 10 different altitudes along Mt. Baker highway, starting in Bellingham and increasing by 200 ft in elevation between each site. Back in the lab, she measures dry plant weight.

*Scatterplot of “plant weight” on the y-axis and “elevation” on the x-axis with a best-fit line
Line graph of “plant weight” on the y-axis and “elevation” on the x-axis
Bar graph comparing “plant weight” from the 5 lowest sites to “plant weight” in the 5 highest sites
Bar graph of “plant weight” for each elevation

10. Do not use a calculator for this question. In the following scenario, what would be the best way to graph these data? You work for the Washington State Department of Transportation, and you’re interested in the effect of car color on number of traffic accidents. You pull up the records of traffic accidents in 2010 and copy down the number of red, silver, black, and blue cars in accidents.

Bar graph of “number of accidents” for each “car color”
Bar graph comparing “number of accidents” for the light-colored cars to “number of accidents” for the dark-colored cars
Scatterplot of “number of accidents” vs. car color with a best-fit line
Line graph of “number of accidents” vs. “car color”

11. Do not use a calculator for this question. The graph below expresses the relationship between price per pound of apples, and the number of days the apples ripened. Which of the following statements is true?

*The slope of the graph is 5 cents/pound per day, meaning that apples increase in value by 5 cents/pound for every day they are allowed to ripen
The y-intercept of the graph is 20 cents/pound, meaning that apples increase in value by 20 cents/pound for every day they are allowed to ripen
The y-intercept of the graph is 20 cents/pound, meaning that the average cost of apples is 20 cents/pound
The slope of the graph is 0.5 cents/pound per day, meaning that apples increase in value by 0.5 cents/pound for every day they are allowed to ripen

12. Do not use a calculator for this question. What is the slope of the line MN?
*6/7
1
7/6
5/6
13. You may use a calculator for this question. Ranger Roger has three days to count the 1200 Douglas fir trees in Arroyo Park so he needs to count one third of them per day. He counted this number on the first day, but due to lightning storms only counted 60% as many on the second day. What simplified fraction of the total number of fir trees has he counted after the first two days?
*8/15
640
about 1/2
2/5

14. You may use a calculator for this question. What is 75% of 280 plus 1/3 of 630? Divide by 900 and simplify.
*7/15
21/45
6/10
49/120

15. You may use a calculator for this question. In a herd of 970 zebras, about three-fifths of two percent of all animals are injured crossing a stream. How many are injured? Round to the nearest whole animal.
*12
11.64
116
291
16. You may use a calculator for this question. What is 5% of 1/4 of 540? \textbf{ROUND} to the nearest whole number.

*7
6.75
67.5
68
APPENDIX F

Basic Quantitative Skills Survey
Spring 2013 post-test

1. What’s your full name? This will be used to anonymously code your survey results with your assessment results.

2. What’s your major or intended major?
   General Biology
   Biology with Ecology, Evolution & Organismal emphasis
   Biology with Cell emphasis
   Biology with Marine emphasis
   Biology with Molecular and Cell emphasis
   Biology with Teaching emphasis
   Chemistry
   Biochemistry
   Kinesiology
   Environmental Science/Studies
   Behavioral Neuroscience
   Psychology
   Community Health
   Anthropology
   English
   Other

3. Have the activities from this past quarter changed your answer to this question: “Do you think math is important for biology, and why or why not?” What’s your response now?

4. Have the activities from this past quarter changed your answer to this question: “Do you think biology courses should include more math concepts? If so, which ones?” What’s your response now?

5. Do you think the activities of this past quarter improved your scores on this assessment?

6. What did you think of the activities from this past quarter? Were they too short or long? Did you like them? Do you have any ideas for improving them?

7. On a scale of 1-5 (with 1 being the least and 5 being the most), how much effort did you put into the activities?
   1=very low effort
   2=low effort
   3=neutral effort
   4=some effort
   5=a lot of effort; I think I did the best I possibly could have
8. On a scale of 1-5 (with 1 being the least and 5 being the most), how much effort did you put into the post-assessment? 1=very low effort while 5=pretty high effort.
1=very low effort
2=low effort
3=neutral effort
4=some effort
5=a lot of effort; I think I did the best I possibly could have
APPENDIX G

Treatment Activities

Name __________________________
Date __________________________

Exponential Growth & Decay (35-50 minutes)
You may type or hand-write answers, but show your work. You may submit responses electronically or by paper. Submit electronically via email to dethiel@students.wwu.edu, and title the document with your last name and activity number & letter: “Dethier_IA”. Submit paper documents to your lab TA or my office: BI231.

(1) Work through Parts I & II of this activity⁵
http://mathbench.umd.edu/modules/popn-dynamics_exponential-growth/page01.htm

(2) Answer Practice Problems “13. Drug Dosage” and “15. Viral Growth”

---

⁵ MathBench Biology Modules from the University of Maryland
Name ________________________________
Date ________________________________

Logarithms (20-35 minutes)
You may type or hand-write answers, but show your work. You may submit responses electronically or by paper. Submit electronically via email to dethiel@students.wwu.edu, and title the document with your last name and activity number & letter: “Dethier_2A”. Submit paper documents to your lab TA or my office: BI231.

(1) Work through numbers 1-4 and 8-12 (skip “messy numbers” 5-7)⁶
http://mathbench.umd.edu/modules/measurement_logs/page01.htm


(3) Expand the following logarithms using the power, product, or quotient rules or a combination of rules.

1. \( \log_2(6a) = \) ______________________________________________________________

2. \( \log_5(19/2) = \) __________________________________________________________

3. \( \log_2[(1-x)/y]^3 = \) ______________________________________________________

4. \( \log_2(4a/5) = \) __________________________________________________________

Write the following logarithms as a single expression

5. \( 2\log_310 - \log_34 = \) __________________________________________________________

6. \( 4\log_5x - \log_5y + \log_5z = \) ________________________________________________

---

⁶ MathBench Biology Modules from the University of Maryland
⁷ From mathwords.com webmastered by Bruce Simmons
Graphing (30-45 minutes)
Winter quarter version

You may type or hand-write answers, but show your work. You may submit responses electronically or by paper. Submit electronically via email to dethiel@students.wwu.edu, and title the document with your last name and activity number & letter: “Dethier_3A”. Submit paper documents to your lab TA or my office: BI231.

(1) Work through numbers 1-3
http://mathbench.umd.edu/modules/visualization_graph/page01.htm

(2) Work through this entire activity
http://mathbench.umd.edu/modules/prob-stat_bargraph/page01.htm

(3) Make the following graphs.
1. Ethylene is a plant hormone that causes fruit to mature. The following table presents results from an experiment studying the effect of ethylene concentration on growth rate in Gala apples. Graph the results; be sure to choose an appropriate graph type and include all necessary labels.

<table>
<thead>
<tr>
<th>Amount of ethylene (ml/m²)</th>
<th>Days to maturity</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>15</td>
<td>13</td>
</tr>
<tr>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>25</td>
<td>9</td>
</tr>
<tr>
<td>30</td>
<td>8</td>
</tr>
<tr>
<td>35</td>
<td>7</td>
</tr>
</tbody>
</table>

2. A researcher is interested in growth of different cichlid fish species in Lake Malawi. She collects four individuals of each species and weighs them. After four weeks she weighs them again and calculates their change in weight. Graph her experimental results; be sure to choose an appropriate graph type and include all necessary labels. (hint: you can use either standard deviation or standard error for error bars)

<table>
<thead>
<tr>
<th>Change in Fish weight (g) in cichlid fish from Lake Malawi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
</tbody>
</table>

---

8 MathBench Biology Modules from the University of Maryland
9 MathBench Biology Modules from the University of Maryland
10 Worksheet from Gull Lake Community Schools
APPENDIX H

Control Activities

Name

Date

Mathematics and Biology (35-50 minutes)
You may type or hand-write answers, but show your work. You may submit responses electronically or by paper. Submit electronically via email to dethiel@students.wwu.edu, and title the document with your last name and activity number & letter: “Dethier_1B”. Submit paper documents to your lab TA or my office: BI231.

(1) Read the following article\textsuperscript{11}
http://www.plosbiology.org/article/info%3Adoi%2F10.1371%2Fjournal.pbio.0020439

(2) Follow-up Questions

The Past
1. Describe one of the examples of an innovation in mathematics that benefited biology.

The Present
2. How are the “pyramid” of “applied mathematics” and “rectangular table” of the “biological landscape” related (Cohen 2004) related? Why does the author use these images to illustrate his point?

The Future
3. Describe the three potential problems and solutions the author lists. Then devise specific methods of achieving one of those solutions (i.e. how would provide “opportunities for training in both biology and mathematics at undergraduate, graduate, and postdoctoral levels”?)

4. Describe one other potential problem you see in the growing collaboration between mathematics and biology.

\textsuperscript{11} Cohen JE (2004). Mathematics is biology's next microscope, only better; biology is mathematics' next physics, only better. PLOS biology. (accessed 12 September 2012).
(1) Read pages 1-10 of “Bio2010” (available in the “Documents” folder on Blackboard)\textsuperscript{12}

(2) Follow-up Questions:

1. Why must biology education become more quantitative?

2. How does this document recommend increasing the quantitative nature of undergraduate biology education? What is one way the Biology Department at Western could do this?

3. Describe one challenge preventing institutions from increasing the quantitative nature of their undergraduate biology curricula? What is one potential solution to this problem?

\textsuperscript{12} From Free Executive Summary of “BIO2010: Transforming Undergraduate Education for Future Research Biologists”
Mathematics in Biology Education (Example) 30-45 minutes
You may type or hand-write answers, but show your work. You may submit responses electronically or by paper. Submit electronically via email to dethiel@students.wwu.edu, and title the document with your last name and activity number & letter: “Dethier_3B”. Submit paper documents to your lab TA or my office: BI231.

(1) Read Speth et al. (2010) (available in the “Documents” folder on Blackboard)

(2) Answer the following questions.

Follow-up Questions
1. What is the definition of “quantitative literacy”? Please write this in your own words.

2. Describe the “termite activity” and the goals of this activity.

3. What do the authors conclude about changes in their students’ QL abilities?

4. Do you think Western’s Biology Department should try a similar type of change in our introductory biology curriculum? Why or why not?

---

APPENDIX I

Revised Activity 3A (Spring)

You may type or hand-write answers, but show your work. You may submit responses electronically to dethiel@students.wwu.edu, titling the document with your last name and activity number/letter (Dethier_3A) OR on paper to the box outside my office: BI231.

(1) Work through numbers 1-3
http://mathbench.umd.edu/modules/visualization_graph/page01.htm

(2) Work through this entire activity
http://mathbench.umd.edu/modules/prob-stat_bargraph/page01.htm

(3) Make the following graphs.

3. Ethylene is a plant hormone that causes fruit to mature. The following table presents results from an experiment studying the effect of ethylene concentration on growth rate in Gala apples. Graph the results in a way that best represents any trends. Include any necessary labels.

<table>
<thead>
<tr>
<th>Amount of ethylene (ml/m²)</th>
<th>Days to maturity</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>15</td>
<td>13</td>
</tr>
<tr>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>25</td>
<td>9</td>
</tr>
<tr>
<td>30</td>
<td>8</td>
</tr>
<tr>
<td>35</td>
<td>7</td>
</tr>
</tbody>
</table>

1. A researcher is interested in growth of different cichlid fish species in Lake Malawi. She collects four individuals of each species and weighs them. After four weeks she weighs them again and calculates their change in weight. Graph the results in a way that best represents any trends (hint: you can use either standard deviation or standard error). Include any necessary labels.

14 MathBench Biology Modules from the University of Maryland
15 MathBench Biology Modules from the University of Maryland
16 Worksheet from Gull Lake Community Schools
17 Worksheet from Gull Lake Community Schools
<table>
<thead>
<tr>
<th>Individual</th>
<th><em>M. auratus</em></th>
<th><em>M. microstoma</em></th>
<th><em>R. longiceps</em></th>
<th><em>C. moorei</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.23</td>
<td>0.44</td>
<td>0.34</td>
<td>0.58</td>
</tr>
<tr>
<td>2</td>
<td>0.14</td>
<td>0.56</td>
<td>0.28</td>
<td>0.73</td>
</tr>
<tr>
<td>3</td>
<td>0.37</td>
<td>0.42</td>
<td>0.32</td>
<td>0.86</td>
</tr>
<tr>
<td>4</td>
<td>0.28</td>
<td>0.68</td>
<td>0.47</td>
<td>0.68</td>
</tr>
</tbody>
</table>
APPENDIX J

Analyses—Complete List of Models
Table 13. All models for hypothesis testing analysis evaluating factors that predict whether or not a student answered a question correctly. The best model is highlighted with the gray box. df represents degrees of freedom for each model and ΔAIC is the difference between the given model’s AIC score and the best model’s AIC score. Models are listed in order from best to worst fit (lowest to highest AIC value).

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>AIC</th>
<th>ΔAIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>M14 = score ~ construct<em>achievement</em>test + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>10</td>
</tr>
<tr>
<td>M16 = score ~ treatment<em>test + construct</em>achievement*test + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>12</td>
</tr>
<tr>
<td>M10 = score ~ achievement*test + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>6</td>
</tr>
<tr>
<td>M13 = score ~ treatment<em>achievement</em>test + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>10</td>
</tr>
<tr>
<td>M15 = score ~ treatment<em>construct</em>achievement*test</td>
<td>18</td>
<td>7439.9</td>
<td>8.9</td>
</tr>
<tr>
<td>M8 = score ~ construct*achievement + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>6</td>
</tr>
<tr>
<td>M11 = score ~ treatment<em>construct</em>achievement + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>10</td>
</tr>
<tr>
<td>M3 = score ~ achievement + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>4</td>
</tr>
<tr>
<td>M6 = score ~ treatment*achievement + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>6</td>
</tr>
<tr>
<td>M1 = score ~ treatment + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>4</td>
</tr>
<tr>
<td>M4 = score ~ test + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>4</td>
</tr>
<tr>
<td>M7 = score ~ treatment*test + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>6</td>
</tr>
<tr>
<td>M5 = score ~ treatment*construct + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>6</td>
</tr>
<tr>
<td>MR = score ~ (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>3</td>
</tr>
<tr>
<td>M9 = score ~ construct*test + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>6</td>
</tr>
<tr>
<td>M2 = score ~ construct + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>4</td>
</tr>
<tr>
<td>M12 = score ~ treatment<em>construct</em>test + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>10</td>
</tr>
</tbody>
</table>
Table 14. All models for hypothesis testing analysis evaluating factors that predict whether or not a student answered a question correctly during the winter quarter. The gray box highlights the top three models averaged together to make the best-fit model. df represents degrees of freedom for each model and ∆AIC is the difference between the given model’s AIC score and the best model’s AIC score. Models are listed in order from best to worst fit (lowest to highest AIC value).

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>AIC</th>
<th>∆AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>M10 = score ~ achievement*test + (1</td>
<td>section)</td>
<td>5</td>
<td>4130.6</td>
</tr>
<tr>
<td>M13 = score ~ treatment<em>achievement</em>test + (1</td>
<td>section)</td>
<td>9</td>
<td>4131.1</td>
</tr>
<tr>
<td>M14 = score ~ construct<em>achievement</em>test + (1</td>
<td>section)</td>
<td>9</td>
<td>4132.0</td>
</tr>
<tr>
<td>M16 = score ~ treatment<em>test + construct</em>achievement*test + (1</td>
<td>section)</td>
<td>11</td>
<td>4133.0</td>
</tr>
<tr>
<td>M6 = score ~ treatment*achievement + (1</td>
<td>section)</td>
<td>5</td>
<td>4135.5</td>
</tr>
<tr>
<td>M8 = score ~ construct*achievement + (1</td>
<td>section)</td>
<td>5</td>
<td>4136.9</td>
</tr>
<tr>
<td>M3 = score ~ achievement + (1</td>
<td>section)</td>
<td>3</td>
<td>4137.3</td>
</tr>
<tr>
<td>M11 = score ~ treatment<em>construct</em>achievement + (1</td>
<td>section)</td>
<td>9</td>
<td>4137.8</td>
</tr>
<tr>
<td>M15 = score ~ treatment<em>construct</em>achievement*test + (1</td>
<td>section)</td>
<td>17</td>
<td>4139.0</td>
</tr>
<tr>
<td>M1 = score ~ treatment + (1</td>
<td>section)</td>
<td>3</td>
<td>4278.9</td>
</tr>
<tr>
<td>M7 = score ~ treatment*test + (1</td>
<td>section)</td>
<td>5</td>
<td>4280.4</td>
</tr>
<tr>
<td>M5 = score ~ treatment*construct + (1</td>
<td>section)</td>
<td>5</td>
<td>4281.2</td>
</tr>
<tr>
<td>M12 = score ~ treatment<em>construct</em>test + (1</td>
<td>section)</td>
<td>9</td>
<td>4285.5</td>
</tr>
<tr>
<td>M17 = score ~ construct<em>treatment</em>test + (1</td>
<td>section)</td>
<td>9</td>
<td>4285.5</td>
</tr>
<tr>
<td>MR = score ~ (1</td>
<td>section)</td>
<td>2</td>
<td>4286.4</td>
</tr>
<tr>
<td>M4 = score ~ test + (1</td>
<td>section)</td>
<td>3</td>
<td>4287.7</td>
</tr>
<tr>
<td>M2 = score ~ construct + (1</td>
<td>section)</td>
<td>3</td>
<td>4288.4</td>
</tr>
<tr>
<td>M9 = score ~ construct*test + (1</td>
<td>section)</td>
<td>5</td>
<td>4290.6</td>
</tr>
</tbody>
</table>
Table 15. All models for population demographics and quantitative skills analysis evaluating factors that predict pre-test score. The gray box highlights the four models averaged together to make the best-fit model. df represents degrees of freedom for each model and ∆AIC is the difference between the given model’s AIC score and the best model’s AIC score. Models are listed in order from best to worst fit (lowest to highest AIC value).

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>AIC</th>
<th>∆AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEx47 = score ~ transfer + p.math + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>11</td>
</tr>
<tr>
<td>MEx29 = score ~ transfer + p.math + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>14</td>
</tr>
<tr>
<td>MEx22 = score ~ year + transfer + p.math + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>12</td>
</tr>
<tr>
<td>MEx24 = score ~ transfer + gender + p.math + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>12</td>
</tr>
<tr>
<td>MEx8  = score ~ year + transfer + p.math + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>15</td>
</tr>
<tr>
<td>MEx10 = score ~ transfer + gender + p.math + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>15</td>
</tr>
<tr>
<td>MEx7  = score ~ year + transfer + gender + p.math + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>13</td>
</tr>
<tr>
<td>MEx1  = score ~ year + transfer + gender + p.math + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>16</td>
</tr>
<tr>
<td>MEx35  = score ~ transfer + p.math + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>14</td>
</tr>
<tr>
<td>MEx12  = score ~ year + transfer + p.math + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>15</td>
</tr>
<tr>
<td>MEx19  = score ~ transfer + p.math + Pmajor + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>17</td>
</tr>
<tr>
<td>MEx53  = score ~ p.math + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>12</td>
</tr>
<tr>
<td>MEx14  = score ~ transfer + gender + p.math + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>15</td>
</tr>
<tr>
<td>MEx59  = score ~ p.math + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>9</td>
</tr>
<tr>
<td>MEx2  = score ~ year + gender + p.math + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>16</td>
</tr>
<tr>
<td>MEx4  = score ~ year + transfer + p.math + Pmajor + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>18</td>
</tr>
<tr>
<td>MEx30  = score ~ gender + p.math + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>13</td>
</tr>
<tr>
<td>MEx6  = score ~ transfer + gender + p.math + Pmajor + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>18</td>
</tr>
<tr>
<td>MEx50  = score ~ gender + p.math + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>10</td>
</tr>
<tr>
<td>MEx16  = score ~ year + gender + Pmajor + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>10</td>
</tr>
<tr>
<td>MExF  = score ~ year + transfer + gender + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>19</td>
</tr>
<tr>
<td>MEx9  = score ~ year + gender + p.math + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>14</td>
</tr>
<tr>
<td>MEx23  = score ~ year + gender + p.math + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>11</td>
</tr>
<tr>
<td>Model</td>
<td>Formula</td>
<td>AIC</td>
<td>BIC</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>MEx40</td>
<td>score ~ p.math + Pmajor + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx54</td>
<td>score ~ year + p.math + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx20</td>
<td>score ~ gender + p.math + Pmajor + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx18</td>
<td>score ~ year + p.math + Pmajor + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx36</td>
<td>score ~ gender + p.math + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx34</td>
<td>score ~ year + p.math + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx5</td>
<td>score ~ year + gender + p.math + Pmajor + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx48</td>
<td>score ~ transfer + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx13</td>
<td>score ~ year + gender + p.math + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx25</td>
<td>score ~ year + transfer + gender + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx57</td>
<td>score ~ transfer + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx41</td>
<td>score ~ year + transfer + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx21</td>
<td>score ~ year + transfer + gender + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx46</td>
<td>score ~ transfer + gender + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx38</td>
<td>score ~ transfer + Pmajor + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx17</td>
<td>score ~ transfer + gender + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx15</td>
<td>score ~ year + transfer + Pmajor + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx3</td>
<td>score ~ year + transfer + gender + Pmajor + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx43</td>
<td>score ~ year + p.math + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx60</td>
<td>score ~ Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx31</td>
<td>score ~ year + transfer + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx49</td>
<td>score ~ transfer + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx11</td>
<td>score ~ year + transfer + gender + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx33</td>
<td>score ~ transfer + gender + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx44</td>
<td>score ~ year + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx26</td>
<td>score ~ year + gender + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx51</td>
<td>score ~ year + gender + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx55</td>
<td>score ~ Pmajor + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>Model</td>
<td>Formula</td>
<td>df</td>
<td>Mean Dev</td>
</tr>
<tr>
<td>--------</td>
<td>--------------------------------------------------------------------------</td>
<td>----</td>
<td>----------</td>
</tr>
<tr>
<td>MEx39</td>
<td>score ~ gender + Pmajor + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MExR</td>
<td>score ~ (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx58</td>
<td>score ~ gender + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx37</td>
<td>score ~ year + Pmajor + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx56</td>
<td>score ~ year + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx42</td>
<td>score ~ year + gender + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx61</td>
<td>score ~ ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx52</td>
<td>score ~ gender + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx45</td>
<td>score ~ year + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx32</td>
<td>score ~ year + gender + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx28</td>
<td>score ~ year + p.math + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
</tbody>
</table>
Table 16. All models for population demographics and quantitative skills analysis evaluating factors that predict changes between the pre- and post-test scores. The gray box highlights the two models averaged together to make the best-fit model. df represents degrees of freedom for each model and ΔAIC is the difference between the given model’s AIC score and the best model’s AIC score. Models are listed in order from best to worst fit (lowest to highest AIC value).

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>AIC</th>
<th>ΔAIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEx29 = score ~ test + transfer + p.math + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>15</td>
</tr>
<tr>
<td>MEx8 = score ~ test + year + transfer + p.math + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>16</td>
</tr>
<tr>
<td>MEx10 = score ~ test + transfer + gender + p.math + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>16</td>
</tr>
<tr>
<td>MEx1 = score ~ test + year + transfer + gender + p.math + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>17</td>
</tr>
<tr>
<td>MEx4 = score ~ test + year + transfer + p.math + Pmajor + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>19</td>
</tr>
<tr>
<td>MEx6 = score ~ test + transfer + gender + p.math + Pmajor + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>19</td>
</tr>
<tr>
<td>MExF = score ~ test + year + transfer + gender + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>20</td>
</tr>
<tr>
<td>MEx53 = score ~ test + p.math + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>13</td>
</tr>
<tr>
<td>MEx30 = score ~ test + gender + p.math + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>14</td>
</tr>
<tr>
<td>MEx28 = score ~ test + year + p.math + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>14</td>
</tr>
<tr>
<td>MEx9 = score ~ test + year + gender + p.math + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>15</td>
</tr>
<tr>
<td>MEx40 = score ~ test + p.math + Pmajor + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>16</td>
</tr>
<tr>
<td>MEx22 = score ~ test + year + transfer + p.math + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>13</td>
</tr>
<tr>
<td>MEx20 = score ~ test + gender + p.math + Pmajor + ethnicity (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>17</td>
</tr>
<tr>
<td>MEx18 = score ~ test + year + p.math + Pmajor + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>17</td>
</tr>
<tr>
<td>MEx7 = score ~ test + year + transfer + gender + p.math + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>14</td>
</tr>
<tr>
<td>MEx47 = score ~ test + transfer + p.math + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>12</td>
</tr>
<tr>
<td>MEx12 = score ~ test + year + transfer + p.math + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>16</td>
</tr>
<tr>
<td>MEx19 = score ~ test + transfer + p.math + Pmajor + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>18</td>
</tr>
<tr>
<td>MEx5 = score ~ test + year + gender + p.math + Pmajor + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>18</td>
</tr>
<tr>
<td>MEx24 = score ~ test + transfer + gender + p.math + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>13</td>
</tr>
<tr>
<td>MEx2 = score ~ test + year + transfer + gender + p.math + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>17</td>
</tr>
<tr>
<td>MEx35 = score ~ test + transfer + p.math + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
<td>15</td>
</tr>
<tr>
<td>Model</td>
<td>Formula</td>
<td>N</td>
<td>Log Likelihood</td>
</tr>
<tr>
<td>---------</td>
<td>------------------------------------------------------------------------</td>
<td>----</td>
<td>----------------</td>
</tr>
<tr>
<td>MEx14</td>
<td>score ~ test + transfer + gender + p.math + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx25</td>
<td>score ~ test + year + transfer + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx48</td>
<td>score ~ test + transfer + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx27</td>
<td>score ~ test + transfer + gender + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx59</td>
<td>score ~ test + p.math + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx15</td>
<td>score ~ test + year + transfer + Pmajor + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx3</td>
<td>score ~ test + year + transfer + gender + Pmajor + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx50</td>
<td>score ~ test + gender + p.math + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx38</td>
<td>score ~ test + transfer + Pmajor + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx54</td>
<td>score ~ test + p.math + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx17</td>
<td>score ~ test + transfer + gender + Pmajor + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx23</td>
<td>score ~ test + year + gender + p.math + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx34</td>
<td>score ~ test + year + p.math + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx36</td>
<td>score ~ test + gender + p.math + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx13</td>
<td>score ~ test + year + gender + p.math + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx60</td>
<td>score ~ test + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx44</td>
<td>score ~ test + year + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx51</td>
<td>score ~ test + year + gender + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx26</td>
<td>score ~ test + year + gender + Pmajor + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx55</td>
<td>score ~ test + Pmajor + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx39</td>
<td>score ~ test + gender + Pmajor + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx37</td>
<td>score ~ test + year + Pmajor + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx16</td>
<td>score ~ test + year + gender + Pmajor + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx41</td>
<td>score ~ test + year + transfer + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx21</td>
<td>score ~ test + year + transfer + gender + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx31</td>
<td>score ~ test + year + transfer + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx11</td>
<td>score ~ test + year + transfer + gender + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx57</td>
<td>score ~ test + transfer + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx46</td>
<td>score ~ test + transfer + gender + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>Model</td>
<td>Formula</td>
<td>df</td>
<td>AIC</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------------------------------------------------------</td>
<td>----</td>
<td>---------</td>
</tr>
<tr>
<td>MEx49</td>
<td>score ~ test + transfer + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx33</td>
<td>score ~ test + transfer + gender + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx43</td>
<td>score ~ test + year + p.math + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx62</td>
<td>score ~ test + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MExR</td>
<td>score ~ (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx58</td>
<td>score ~ test + gender + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx42</td>
<td>score ~ test + year + gender + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx56</td>
<td>score ~ test + year + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx61</td>
<td>score ~ test + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx45</td>
<td>score ~ test + year + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx52</td>
<td>score ~ test + gender + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
<tr>
<td>MEx32</td>
<td>score ~ test + year + gender + ethnicity + (1</td>
<td>section) + (1</td>
<td>quarter)</td>
</tr>
</tbody>
</table>