THE EFFECTS OF *MILLER & LEVINE BIOLOGY* (2010) ON STUDENT PERFORMANCE

Final Report

October 1, 2010

Submitted by: Rebecca M. Eddy, Ph.D.
H. Todd Ruitman, M.A., Michelle Sloper, & Nancy Hankel
and special acknowledgement to Xiaoxia Newton, Ph.D.

COBBLESTONE
APPLIED RESEARCH & EVALUATION, INC.

COBBLESTONE APPLIED RESEARCH AND EVALUATION, INC.
2120 Foothill Blvd., Ste. 202 | La Verne, CA 91750 | (800) 971-3891

Acknowledgements

We would like to thank all of the participating teachers and students who generously provided their time and effort to participating in all aspects of this study. We appreciate all of the work that you do as current and future scientists, and your diligent work to make the study a success. Special thanks to all site liaisons that helped to coordinate all study activities. We also appreciate everyone from Pearson Education including Mary Hill, Nick Raducanu, Kathy Immel, Terri Clark, Deborah Murray, Judy Porter, Laurie Morgan, Ann Vilcheck, Anne Thomasmeyer and especially Marcy Baughman. We would also like to thank John Hattie for his help with technical assistance and statistical analyses. Finally, thanks to our families for their ongoing support.

With gratitude,

Rebecca Eddy
Todd Ruitman
Michelle Sloper
Nancy Hankel

Table of Contents

Tables

Figures

Executive Summary

Cobblestone Applied Research and Evaluation, Inc. was hired by Pearson Education to conduct an efficacy study of the *Miller & Levine Biology* (2010) curriculum in the 2009-10 school year. The primary purpose of the study was to determine if students using the program would increase their knowledge of biology concepts throughout the year and outperform other students using a competitor biology program. We also investigated the extent to which teachers adhered to the *Understanding by Design* strategies contained within the program. Finally, we investigated usage and satisfaction for all program components. This report describes all study activities and provides results related to the research questions.

Study Description, Design, and Measures

Teachers and their corresponding classes were randomly assigned to either the treatment condition (using the *Miller & Levine Biology* program) or the control condition (using the existing biology program at their school). The study required that treatment teachers and their students use a minimum of Units 1-5 of the *Miller & Levine Biology* (2010) curriculum (including ancillary and online program components) in classrooms during the 2009-10 school year. Treatment study teachers were free to use additional chapters/units that aligned to their state standards as well. The *Miller & Levine Biology* (2010) program included two versions: Mainstream (written at the tenth grade level) and Foundation (geared toward below grade-level readers).

The study was designed to assess implementation of the curriculum in classrooms, answer research questions related to student achievement and attitudes, and to assess product satisfaction from teachers and students. **Implementation measures** were collected to assess the extent to which students and teachers implemented their respective biology programs in their classrooms. **Outcome measures** were administered as pre-test and post-test instruments and assessed the impact on student attitides and acheivement.

Study Sample

Twenty four teachers across six schools in five states (Oregon, North Carolina, Indiana, Washington, Oklahoma) from a combination of suburban and rural areas taught using either the *Miller & Levine Biology* (2010) program (treatment) or their existing biology program (control) in their classrooms during the study. Data were analyzed for 1,974 participating students in 79 separate class groups. The study sample was primarily Caucasian students, and included students from various levels of Biology courses, including lower-level non-college preparatory classes, college-prepartory classes, and honors classes. Teachers taught Biology for 13.1 years, on average, and more than half (54%) possessed a Master's level degree.

Outcome Measures	
Biology Core Content Assessment	Developed using released science questions from various states' Department of Education websites and included 30 multiple choice questions addressing concepts covering the five core units including: the nature of life; ecology; cells; genetics; and evolution.
Stanford Acheivement Test 9 (SAT9)	Published instrument including general high school science assessment that included 40 multiple choice science items.
Student attitude survey	Included questions related to students' intrinsic motivation and personal relevance, self efficacy and assessment anxiety, self determination, career motivation, grade motivation and perceptions of teacher efficacy. An additional section on the posttest asked students to rate their satisfaction with elements of the *Miller & Levine Biology* (2010) program.
Implementation Measures	
Online Teacher Logs	Completed by all participating teachers weekly to report the content covered and specific program components used in their classrooms.
Classroom Observations	Observed by researchers, all teachers and their students participated including two times for treatment classrooms (fall and spring) and one time for control classrooms (fall or spring).
Teacher Interviews/ Focus Groups	Completed at the end of the study, most teachers participated in individual interviews or focus groups to discuss the program implementation over the duration of the school year.

Program Implementation

We systematically tracked components of program utilization by teachers and observed teachers and their students using the program during the study. Implementation ratings (low, medium, high) were established for every participating teacher based on information reported in their weekly logs in comparison to established implementation guidelines.

Most Frequently Used Program Components

Classroom Activities	Online Activities: (biology.com)	Teacher Resources	Lab Activities
Section Assessments	Art in Motion	Editable Worksheets	Quick Labs
Chapter Assessments	Interactive Art	Lesson PowerPoint Presentations	Skills Lab
"Building Vocabulary"	Art Review	Exam *View* Assessment Suite	Lab Skills #3 Activity
"In Your Notebook"	Tutor Tubes	Teacher's Resource CD-Rom	Inquiry Into Scientific Thinking
Visual Analogy (Mainstream Only)	Data Analysis	Transparencies	Real-World Lab

Study Results

Research Question 1:

Are teachers able to successfully integrate pedagogical elements of the Understanding by Design *model using the* Miller & Levine Biology *(2010) curriculum?*

Answer: The implementation of the *Understanding by Design* elements varied across teachers and versions of the textbook used. On average, teachers used 44% of *Big Idea* sections; this differed for users of the Mainstream and Foundation books in which teachers using the Mainstream textbook used this feature more consistently.

Teachers found both the *Chapter Mystery* and *Big Idea* components to be an effective means of connecting main ideas and engaging students. Teachers felt that students enjoyed the *Big Idea* and that it was useful in focusing student attention on the core concepts. Teachers most often utilized Chapter Mysteries as a means to introducing the new chapter and felt they were successful for that purpose.

"I liked the *Big Idea* because it always comes back to that one thing. We always introduce the new chapter and the new unit with the big question and then periodically come back to that. I felt [that] was a real strength."

– Teacher using *Miller & Levine Biology* (2010)

Research Question 2:

How does student achievement differ for those using the Miller & Levine *(2010) program compared with those using another Biology program?*

Answer: Overall, there were no differences in terms of student achievement between those students using the *Miller & Levine Biology* (2010) program and those using a competitor biology program; however, when data were broken down by implementation level, those using the *Miller & Levine Biology* (2010) program in the high implementation group significantly outperformed control group students using a competitor program on the SAT9 science exam. Therefore, these data suggest that when properly implemented as intended by the publisher, students using the *Miller & Levine Biology* (2010) program should outperform students using competitor biology programs on tests of science achievement.

Research Question 3:

How do students with different characteristics (e.g., English learners, various ethnicities) using the Miller & Levine Biology *(2010) program perform on student-related outcomes?*

Answer: While there were different student characteristics that were shown to be related to on the SAT9 and Biology Core Content Assessment, students the *Miller & Levine Biology* (2010) program showed significant growth from pretest to posttest on both assessments when certain demographic groups were analyzed. Specifically, Latino and African American students had significant gains from pretest to posttest with Latino students showing larger gains when compared to Caucasian students from pretest to posttest on both outcome measures. Results also showed students speaking a non-English primary language showed significant gains from pretest to posttest on the SAT9.

Research Question 4:

What is the relationship among students' attitudes towards science and science achievement?

Answer: Findings suggest that the more positive certain student attitudes are towards science the better students will perform on science assessments. The HLM analyses

showed intrinsic motivation and personal relevance, self efficacy and assessment anxiety, and grade motivation as significant predictors of student achievement. Specifically, all three of these constructs were significant predictors of student scores on the Biology Core Content Assessment while only the self efficacy and assessment anxiety construct was a significant predictor on the SAT9.

Research Question 5:

How do students using the Miller & Levine Biology *(2010) program perform from pretesting to posttesting on assessments related to attitudes about science and achievement in science?*

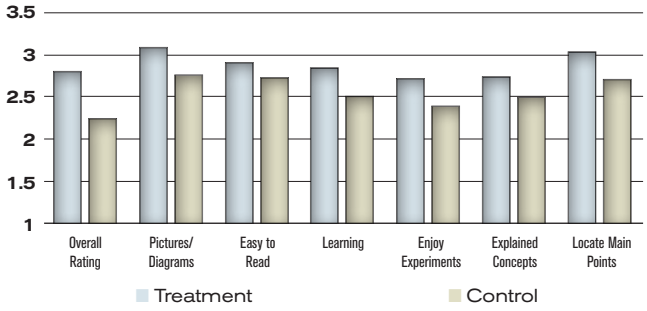
Answer: Overall, the students using the *Miller & Levine Biology* (2010) program showed significant decreases from pretest to posttest on their attitudes and motivation toward science as measured by the student survey; however, results were virtually identical to the control groups' pretest and posttest ratings. These results were not consistent with the students using the Foundation textbook who showed no statistical difference between their pretest and posttest ratings. For the achievement measures, students in the treatment group showed highly significant gains from pretest to posttest on both the SAT9 and Biology Core Content Assessment. These results were the same overall and regardless of textbook (i.e., Mainstream or Foundation) used in the classroom.

Product Satisfaction

Product satisfaction was assessed using input and feedback from multiple sources (student survey items, teacher interviews, and teacher focus groups) regarding program use and satisfaction in participating classrooms.

Overall, student and teacher users of the *Miller & Levine Biology* (2010) program were satisfied with the curriculum program. Students using the *Miller & Levine Biology* (2010) program rated their textbook significantly better than control textbooks in all categories. The top strengths of this curriculum, as reported by students and teachers, were organization of materials, alignment and relevance of supplemental elements, and finally the pictures, graphics, and figures created to visually represent important information. Students using the Foundation series text rated their material more enjoyable than those using the Mainstream version. Combined ratings suggest that both Mainstream and Foundation users reported that they enjoyed building vocabulary sections, or highlighted words, and the pictures of scientific concepts the most. On average, no particular component was aggregated as being *disliked*, however both Foundation and Mainstream users rated "Analyzing Data" as their least enjoyable (Mean = 2.41 out of 4). Treatment teachers reported that differentiated instruction elements of the text were helpful and provided useful ideas and suggestions. Teachers in higher-level biology (e.g., honors) classrooms would have liked to see more higher-order thinking in the biology materials, including more in-depth review questions and scenarios.

Comparison of Student Textbook Ratings: Treatment vs. Control



Online Resources

Teachers overwhelmingly agreed that the online resources provided with the *Miller & Levine Biology* (2010) program were one of the strongest components.

Favorite	Least Favorite
Aesthetics: Students found the diagrams, pictures, visual analogies, and images to be a favorite element of the textbook. "The pictures and diagrams…they made it easy to visualize and understand the process of things." ***Ease of Comprehension & Use:*** Students felt that the book was well organized, making important information easy to access. "…the book was organized in a nice way so that main points were easy to find, summarize, and understand." ***Key Concepts & Vocabulary Terms:*** Many students liked the presentation of the Key Concepts and vocabulary terms. The highlighting of the terms and the provided definitions were user-friendly.	***Chapter Assessments:*** Students found chapter assessments to be lengthy and repetitive. ***Some Areas Confusing:*** Students reported that the book was confusing in some areas; specifically some diagrams were unclear and hard to understand. ***Weight & Size:*** Students found the book to be large and bulky. For those with access to the internet from their homes, students appreciated the ability to access the textbook online so that they did not have to carry their textbook.

Students also rated online components high (2.82 out of 4). Several teachers reported being very satisfied with the ability to edit worksheets with the online resources. Teachers also reported liking the ability to access the textbook online, although they were often unable to assign homework online because not all students had internet access in the home.

Lab Activities/Lab Manual

The lab activities included in the *Miller & Levine Biology* (2010) program were rated highly by students and teachers. Students felt these activities supplemented their understanding of course material and listed these activities as one of their favorite components of the program.

Supplemental Materials

Some teachers found the amount of material and available ancillary materials overwhelming, but most reported satisfaction in lesson planning and in writing exams using components such as *ExamView*, PowerPoint presentations, and the lesson planner. Most teachers felt the *Untamed Science* video series did a good job of supplementing the student experience; students rated the video series highly as a component of the curriculum. Least favorite ancillary materials identified by teachers and students were the student workbooks.

Study Conclusions

Teachers and students alike enjoyed using the *Miller & Levine Biology* (2010) program. The program is rightly considered a vast resource for teachers to use with their biology classes, and provides opportunities for students to access difficult concepts and allows teachers efficient planning of lessons while using high-quality materials. While use of the program did not have an impact on students' attitudes towards science, there was an advantage for some students using the *Miller & Levine Biology* (2010) program in the classroom. While teacher and student factors had the highest impact on student scores, there is some evidence that for those students in classrooms where the program components were used most faithfully and most often, student scores on one measure of science achievement were superior to control students.

Section One: Efficacy Study Background, Study Purpose, and Program Description

Efficacy Study Background

Although current reports from the National Association of Education indicate recent, slight gains in science achievement for U.S. students, international comparisons in science achievement still rank the U. S. behind several nations in science competency (Forgione, 1998; Gonzales, Williams, Jocelyn, Roey, Kastberg, and Brenwald, 2008; Snyder & Dillow, 2010). Even though recent increases in science achievement are promising, education advocates across the nation are still concerned about improving science curriculum as a means to increasing scientific literacy for America's students.

With the introduction of federal mandates for standards-based education in No Child Left Behind, education reform has shifted its primary concern to accountability. This has led curriculum developers to place increased emphasis on the evaluation and assessment of their curricular programs to achieve greater alignment with federal and state standards (Marx & Harris, 2006). According to the National Science Teacher Association (www.nsta.org), assessment serves an important function in the development of programs aimed at scientific literacy. One main component of this assessment is textbook and curricular programs. Education researchers agree that science textbooks serve as the cornerstone of science instruction and should therefore be a main target of reform and improvement efforts (Ball & Cohen, 1996).

Science curricular materials have focused on meeting state and national standards, targeting multiple students learning levels, as well as engaging students substantially in the scientific thinking process and acquisition of knowledge of specific science concepts. The *Miller & Levine Biology* (2010) program is designed to address all of these priorities with their most recent revision which includes wide coverage of state and national standards, specific versions with varying reading levels, and engaging new features and activities. However, it remains an empirical question that students using this particular program would necessarily outperform those students using a competitor program. It is, therefore, imperative that research be conducted to examine the extent to which the *Miller & Levine Biology* (2010) program impacts students' attitudes towards and achievement in science. Given the requirements of U.S. Department of Education's *What Works Clearinghouse*[1], research designs must include the use of experimental controls (among other features), usually referred to as Randomized Controlled Trials (RCTs), which are considered efficacy studies.

Study Purpose

An efficacy study of the *Miller & Levine Biology* (2010) program was conducted at six school sites across five states (Indiana, North Carolina, Oklahoma, Oregon, Washington) during the 2009-10 school year by Cobblestone Applied Research & Evaluation, Inc. During the study, teachers' implementation of the Biology curriculum as well as a diverse set of student outcomes was explored. These data provide insight into how the *Miller & Levine Biology* (2010) curriculum may affect students' achievement in science during high school. This study focused on systematically tracking curriculum implementation, measuring students' achievement in science, and investigating the relationship between these elements with an assessment of the students' attitudes towards/interest in science as well as other related constructs, and product satisfaction of the *Miller & Levine Biology* (2010) program. The main purpose for conducting the efficacy study was to answer the following research questions:

Research Question 1:
Are teachers able to successfully integrate pedagogical elements of the Understanding by Design *model using the* Miller & Levine Biology *(2010) curriculum?*

Research Question 2:
How does student achievement differ for those using the Miller & Levine Biology *(2010) program compared with those using another biology program?*

Research Question 3:
How do students with different characteristics (e.g., English learners, various ethnicities) using the Miller & Levine Biology *(2010) program perform on student-related outcomes?*

Research Question 4:
What is the relationship among students' attitudes towards science and science achievement?

Research Question 5:
How do students using the Miller & Levine Biology *(2010) program perform from pretesting to posttesting on assessments related to attitudes about science and achievement in science?*

[1]Detailed information regarding the What Works Clearinghouse can be accessed at www.w-w-c.org.

Table 1. Curriculum Components of the *Miller & Levine Biology* (2010) Program

Understanding by Design	*Big Idea* UbD Teacher Edition Margin Notes
Understanding Through Inquiry Resources	Labs *Quick Labs* *Chapter Mystery*
Differentiated Instruction Resources	Teacher Edition Margin Notes Multiple Texts (Foundation & Mainstream)
Online Resources: www.biology.com	Online Student Explorations Digital Inquiry Activities Interactive Visuals
Electronic Resources	*Untamed Science* Video Series PowerPoint Presentations

Program Description

The *Miller & Levine Biology* (2010) program provides a comprehensive approach to teaching biology. Teachers implementing this program have the opportunity to utilize a wide variety of curricular elements to tailor instruction to their students. The program includes numerous features designed to engage students' interest in science and comprehension of scientific content knowledge including colorful, carefully designed pictures, graphics, and Visual Analogies; Key Questions and highlighted vocabulary terms; and special features such as Careers in Biology. The program has been adapted to include two main versions of the student edition textbook: Mainstream and Foundation. A number of elements have also been included with both of these versions to support a diverse base of student and teacher users. Table 1 summarizes the main program components in the Mainstream textbook.

- The **Mainstream** version is the complete textbook to be utilized by the vast majority of student users. This version is written at the tenth grade reading level and can be used for regular-track, college preparatory and advanced students. This version of the text includes some advanced features not found in the Foundation version, such as *Biology & Society*, and *Careers in Biology*.
- The **Foundation** series is an adapted version of the Mainstream textbook intended to address the learning needs of struggling students through the inclusion of distinct features designed to target lower-level readers and provide more specific support. This version is written one to two grade levels below the Mainstream version, has fewer pages, reduced quantity of text, more comprehensible graphics, and vocabulary support.

Each chapter within the *Miller & Levine Biology* (2010) program begins with a *Big Idea* question, which ties together important concepts throughout the unit. This idea is used to answer questions and apply the information to the end-of-chapter review.

***Unit 3* Big Idea:**
Cellular Basis of Life, Homeostasis

Question:

How are cell structures adapted to their functions?

Each chapter also includes a *Chapter Mystery*, where students read an interesting scenario an attempt to solve a mystery as it unfolds throughout the chapter as "Mystery Clues". The *Chapter Mystery* also connects the *Big Idea* with concepts from the unit. Both the *Big Idea* as well as the *Chapter Mystery* were designed to engage students' interest in science, to make the concepts relevant to them, and to introduce an avenue to reinforce prior concepts throughout a unit. These components, in combination with other program components, were designed to address the *Understanding by Design* (UbD) philosophy (Wiggins & McTighe, 1998). In addition to these distinct features, there are numerous online and electronic resources available designed to further engage students in substantial scientific knowledge. Teacher resources include a lesson planner, editable worksheets, PowerPoint presentations, *ExamView* software, and electronic features include teacher edition and student edition texts, *Untamed Science* videos and an array of online resources. Examples of ancillary materials include Study Workbooks (A and B) and Lab Manuals (A and B).

Adapted from* Chapter Mystery *(Chapter 7)

Michelle was a healthy 25-year-old running in her first marathon…Michelle made sure to drink water at every opportunity. Gradually she began to feel weak and confused. Michelle was rushed to the hospital, where she was gripped by a seizure and went into a coma…

Why did treating Michelle with water make her condition worse?

Background, Study Purpose and Program Description Summary

An efficacy study of the *Miller & Levine Biology* (2010) program was conducted during the 2009-10 school year. The program includes numerous features designed to engage students' interest in science and comprehension of scientific content knowledge including colorful, carefully designed pictures, graphics, and Visual Analogies; Chapter Mysteries and Mystery Clues; Key Questions and highlighted vocabulary terms; and special features such as Careers in Biology and Biology & Society. Teacher resources include a lesson planner, editable worksheets, PowerPoint presentations, *ExamView* software, and electronic features include teacher edition and student edition texts, *Untamed Science* videos and an array of online resources. Finally, ancillary materials include Study Workbooks (A and B) and Lab Manuals (A and B). The *Miller & Levine Biology* (2010) program included two versions: Mainstream (written at the tenth grade level) and Foundation (geared toward below grade-level readers). The study was designed to assess implementation of the curriculum in classrooms, answer research questions related to student achievement and attitudes, and to assess product satisfaction from teachers and students.

Section Two: Study Design, Setting, and Sample

Study Design

The *Miller & Levine Biology* (2010) efficacy study was designed as a Randomized Controlled Trial (RCT) in which teachers (and their corresponding class periods) were randomly assigned to either the treatment group, using the *Miller & Levine Biology* (2010) program or a control group (using the existing biology program at their schools). Teachers and their students used their respective biology programs in their Biology classes for the duration of the 2009-10 school year. An experimental design (specifically an RCT) was selected, as this design is well-regarded as the strongest in terms of internal validity (appropriately assigning cause to a particular treatment) while having the highest probability for ruling out alternative explanations of cause (Shadish, Cook, & Campbell, 2002). In addition to collecting information related to program outcomes (e.g., student achievement data) we also collected information related to program implementation, given that varying levels of implementation can have differential impacts on related outcomes (Sechrest, et al., 1979). The study design is also considered a cluster-type design in which a cluster (class period of students) is nested within one teacher, hence allowing analyses to be conducted on multiple levels to more specifically identify potential treatment effects and moderator variables.

Site Selection

Cobblestone actively recruited sites to participate during the spring and summer of 2009. Initially, Pearson Education provided references to schools and districts to Cobblestone researchers who were interested in participating in the study. In addition, Cobblestone researchers identified potential sites throughout the United States by selecting specific criteria from districts listed in the National Center for Education Statistics (http://nces.ed.gov/ccd/schoolsearch/ and http://nces.ed.gov/ccd/districtsearch/). Several hundred school districts were contacted through phone and email. Recruitment was focused on schools with at least 500 enrolled students and a minimum of four teachers at each school site with multiple sections of Biology classes. Most schools solicited for participation were unable to participate in the study. The most common reasons provided for declining participating included lack of interest or resources at schools to participate in an experimental study and current use of the 2010 or earlier copyrights of the *Miller & Levine Biology* program. It is important to note that schools with diverse student ethnicity and lower-socio economic status individuals were targeted specifically for inclusion in the study to determine the impact of the program in a variety of settings. Ultimately, a majority of districts that had the most diverse group of students declined to participate in the study. This was not unexpected, as the most diverse districts tend to be concentrated in urban areas where students typically have high mobility, district research protocols are particularly stringent, and numerous competing district initiatives does not allow participation in a research study to be a priority.

Of the schools (and/or districts) that met the inclusion criteria, securing their participation occurred through initial contact with district science or curriculum directors, after which point school principals, science chairpersons or teachers were contacted directly. Schools were required to complete applications to participate in the study. All participating teachers, site liaisons, district personnel, and Cobblestone researchers signed a Memorandum of Understanding to formally secure the school's participation. Passive parent consent forms were distributed to all participating students, and students also signed active assent forms to secure participation. All teachers also signed active consent forms. Through the combined efforts of Cobblestone and Pearson, the final sample included six schools across five states (i.e., Indiana, North Carolina, Oklahoma, Oregon, and Washington).

Site Demographic Characteristics

The majority of the participating sites were located in suburban areas, servicing roughly 1,000-2,000 students in grades 9 – 12. Students from all schools were primarily of Caucasian descent; however several schools had large percentages of students from other ethnicities. Participating schools had 16% to 40 % of students receiving free or reduced lunch and median annual household income was between $37,000 - $54,000. Table 2 provides complete information about each site.

Student Participants

There were a total of 1,974 students who participated in the study including 1,126 treatment group students and 848 control group students. Students were contained within 79 classroom clusters. Table 3 summarizes the demographic characteristics of participating students, including

Table 2. School Level Demographic Characteristics for Participating Sites

State		Indiana	North Carolina	Oklahoma	Oregon		Washington
School Site		Site 1	Site 2	Site 3	Site 4	Site 5	Site 6
Location*		Suburban	Rural	Suburban	Suburban	Suburban	Rural
School Size*		2000	1400	1100	1100	1500	1300
Ethnicity*	% Caucasian	75%	54%	58%	75%	75%	89%
	% Hispanic/ Latino	2%	21%	7%	7%	4%	4%
	% African American	15%	24%	21%	3%	3%	1%
	% Other Ethnicity	8%	1%	14%	15%	18%	6%
Economic Measure*	% Free & Reduced Lunch	40%	40%	38%	24%	16%	19%
Community Measure**	% Age 25+ With College Degree	21%	17%	42%	48%	48%	13%
	Median Household Income	$45,000	$39,000	$37,000	$45,000	$45,000	$54,000

** Information obtained from each state's department of education or district websites;*

***US Census 2000.*

gender, primary language, and ethnicity derived from the student surveys. There were approximately equal numbers of male and female students at all sites, and overall 48% of treatment students were male, while 53% control students were male. Consistent with the ethnic distributions within the six communities included in this study, they were fairly equivalent across the treatment and control groups. Students were primarily of Caucasian descent and spoke English as their primary language.

Pre-test surveys also asked students to provide information regarding the education level for both their parents. Students most often reported that they did not know their parents' education level. Those students who did report their mother and father's education levels, the

Table 3. Demographic Characteristics of Student Participants

		Site 1 IN	Site 2 NC	Site 3 OK	Site 4 OR	Site 5 OR	Site 6 WA
Gender	Male	52%	52%	52%	47%	48%	48%
	Female	47%	48%	46%	51%	52%	51%
Ethnicity	Caucasian	48%	45%	35%	66%	73%	67%
	African American	13%	17%	19%	3%	1%	1%
	Native American	1%	1%	3%	1%	-	2%
	Latino/Hispanic	3%	21%	9%	8%	3%	5%
	Asian	1%	1%	1%	3%	5%	-
	Other Ethnicity	13%	5%	9%	9%	10%	12%
Primary Language	English	77%	79%	70%	82%	90%	82%
	Other	1%	11%	4%	6%	3%	4%

Table 4. Summary of Teacher Characteristics

	Highest Degree Attained				Teaching Experience	
	Bachelor of Arts/ Science	Teaching Credential/ Certificate	Master of Arts/ Science	Doctorate/ Professional Degree	Number of years teaching (average)	Number of years teaching Biology (average)
Treatment	4 (27%)	0	9 (60%)	1 (7%)	14.4 years	14.7 years
Control	4 (40%)	1 (10%)	5 (50%)	0	11.7 years	11.5 years
Overall	8 (31%)	1 (4%)	14 (54%)	1 (4%)	13.0 years	13.1 years

largest number of students reported that their mother had completed some college and that their father had completed high school, and parent education levels were also fairly equivalent between treatment and control groups. See Appendix A for a complete table of parent education levels reported by student participants.

Teacher Participants

There were a total of 24 teachers who participated in the study, 14 treatment teachers implementing the *Miller & Levine Biology* (2010) program and 10 control teachers. Teachers' education levels ranged from Bachelor's degrees to Doctoral degrees. Overall, more than half of participating teachers obtained a Master's degree. On average, teachers had approximately 13 years of teaching experience. This mean was approximately equal to the number of years teaching specifically biology. Table 4 summarizes teacher characteristics. Education levels were comparable for treatment and control teachers, however, treatment teachers taught for more years, on average.

Design, Setting, and Sample Summary

Twenty four teachers across six schools in five states (Oregon, North Carolina, Indiana, Washington, Oklahoma) from a combination of suburban and rural areas taught using either the *Miller & Levine Biology* (2010) program (treatment) or their existing biology program (control) in their classrooms during this efficacy study. Data were analyzed for 1,974 participating students in 79 separate class groups. The study sample was primarily Caucasian students, and included students from various levels of Biology courses, including lower-level non-college preparatory classes, college-preparatory classes, and honors classes. Teachers taught Biology for 13.1 years, on average, and more than half (54%) possessed a Master's level degree.

Section Three: Description of Study Procedures and Measures

Teacher Compensation

Compensation for participation in the study was a $300 cash stipend for control teachers, treatment teachers, and study liaisons. In addition, all teachers in the treatment group were provided with a teacher's edition textbook, training on the *Miller & Levine Biology* (2010) program, access to online resources, and all available ancillary materials. Each participating school received enough student edition textbooks for the number of participating students in the treatment group prior to the start of the 2009-2010 school year, with a balance of textbooks for the number of participating control group students delivered at the end of the study as additional compensation.

Training Activities

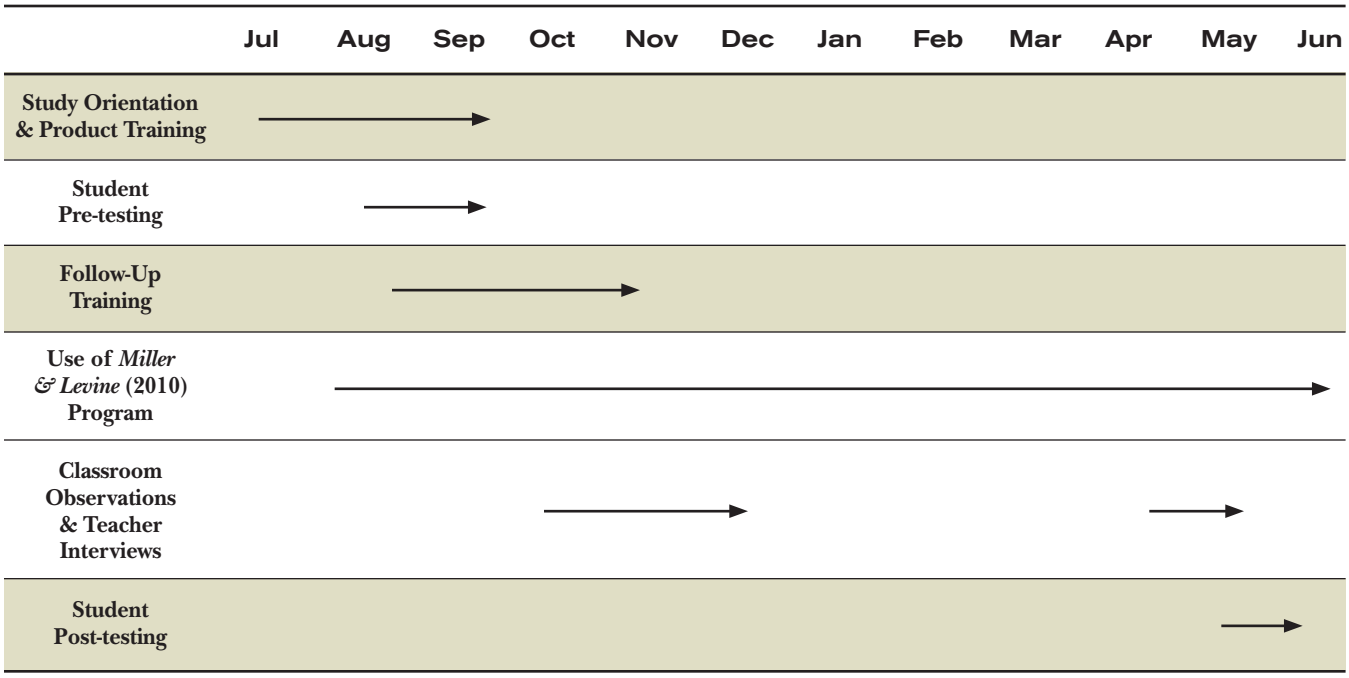
Treatment teachers and their students used the *Miller & Levine Biology* (2010) program over the course of the 2009-2010 school year. Though schools varied in their start and end dates, the sequence of study activities was similar across sites. A detailed description of teacher training is documented in more detail in Appendix B; a summary of study activities and corresponding dates can be found in Table 5.

Data Collection Measures: Curriculum Implementation

Implementation measures were developed to monitor and assess the activities in participating classrooms throughout the year. Implementation measures included weekly teacher logs, classroom observations, and teacher interviews/ focus groups. Teachers also communicated informally with the researchers via email, phone, open-ended sections of the teacher logs, and during informal interviews conducted during the fall and spring classroom observations. Treatment teachers were continuously encouraged to provide feedback about the *Miller & Levine Biology* (2010) program throughout the study. All teachers participated in focus group interviews during spring observations.

Teacher Implementation Logs. Each week teachers were required to submit online logs that described activities in their classrooms including the curricular content covered, as well as which components of the curriculum package were utilized (e.g. the website, workbooks, DVDs, etc.) Logs also provided a place for feedback regarding any problems or issues with the study, materials, or other relevant communication. Separate logs were created for control and treatment teachers. Treatment teacher logs

Table 5. Schedule of Study Activities



highlighted the usage of the *Miller & Levine Biology* (2010) program while the components of the control teacher logs collected information about sections covered in the control textbooks. Participating teachers submitted online logs starting the second week of August 2009 through June 2010.

The primary goals in developing an online teacher log system were:

- Capture as accurately as possible both the *Miller & Levine Biology* (2010 content covered in classes (e.g., Unit, Chapter *Big Idea*, *Chapter Mystery*), as well as supplemental material utilized by treatment teachers.
- Allow teachers to report any activities or events whether at school or within the classroom that might have impacted their teaching or student learning.
- Reduce the strain on teachers by making the process user-friendly and efficient.
- Collect data in a way that was meaningful to researchers and could be reported back easily.

The logs were developed using SurveyMonkey survey software and were presented to teachers via an email link that directed them to the appropriate log. In addition, the online log process also allowed researchers to remain in constant communication with participating teachers so that issues such as dates for training, observations, and test administration could be planned effectively.

Classroom Observations. To validate and supplement the information contained in the teacher implementation logs, Cobblestone researchers (and Pearson Academic Research representatives, when possible) conducted classroom observations lasting the entire class period for every participating treatment classroom. Teachers also provided informal feedback directly before or after observation sessions. Participating treatment classrooms were visited twice during the course of the study, once during fall 2009 and again in spring 2010; while control classrooms were visited once.

Observation Protocol. To establish consistency of observations across sites, an observation protocol was established to help guide researchers in data collection. This protocol required that observers collect information related to areas such as the physical environment, instructional time, sequencing of classroom activities, use of materials, student engagement, and teacher practice. Two protocols were developed: one for classrooms implementing the *Miller & Levine Biology* (2010) and one for control classrooms. Most information in these separate forms was consistent, however, the protocol for treatment classrooms attempted to capture specific elements of the *Miller & Levine Biology* (2010) program. For example, the treatment protocol allowed researchers to check off which curriculum elements were used in the classroom and in which resource these elements could be found.

Teacher Interviews. As part of the debriefing process, a teacher interview protocol was developed for all participating (treatment and control) teachers. Questions were similar for both groups with the exception of product satisfaction questions specifically geared toward the *Miller & Levine Biology* (2010) program for treatment teachers. In April and May 2010, all teachers were interviewed by phone or in person by a member of the research team, following the established protocol. Interviews with treatment teachers lasted approximately 40 minutes, while interviews with control teachers lasted 20 minutes. The teacher interview protocols can be found in Appendix C.

Data Collection Measures: Outcomes

Participation in the study required students to complete three measures at pretest and posttest. The student outcome measures in this study were the Stanford Achievement Test 9 (SAT-9), the Biology Core Content Assessment, and a student attitude survey. The goal of the SAT-9 and the Biology Core Content assessments was to obtain objective measures of student achievement to compare across schools in multiple states. These instruments were intended to measure the impact of the *Miller & Levine Biology* (2010) curriculum in comparison to the control curriculum. The following includes a description of outcome measures used in the current study.

Stanford Achievement Test 9 (SAT9). The SAT9 is a standards-based, nationally recognized science assessment that measures student learning in high school science. The SAT9 was selected as an objective measure of achievement because of its strong psychometric properties. The assessment is norm-based and grade-equivalent scores can be computed from this assessment. The science section of the assessment was used for the study and included 40 multiple-choice questions. The assessment also aligns with the content recommendations from Science for All Americans, Benchmarks for Science Literacy, and the National Science Education Standards.

Biology Core Content Assessment. Given that there were no specific Biology assessments available, an assessment was created from a pool of life science items released from state education departments. To create the final 30-item Biology assessment, five subtests were first created that focused on five core Biology content areas (6 questions for each core area). These areas included Nature of Life, Ecology, Cells, Genetics, and Evolution—which also corresponds to the first five chapters of the *Miller & Levine Biology* (2010) program, and widely considered core topics in most high school biology classes. An assessment specific to the Cells unit (Unit 3) was created for the Biology Pilot Study in spring 2009, and included 23 questions. Data

from the Cells assessment related to reliability scales was established during the pilot study, and researchers selected six items with high reliabilities.

In addition to the Cells section of the assessments, all pilot sites administered additional posttest assessments related to one additional core area in biology (e.g., four additional subtests). The assessments were completed by 526 students taking a Biology course in grades 9-12. It is important to note that each student did not complete all five subtests. Instead, each student was asked to complete only one subtest. This process prevents students from experiencing test fatigue and can also provide reliable data that can be used for the Item Response Theory (IRT) procedure (Baker, 2001). IRT is a statistical method that allows for the development of educational assessments by evaluating each item on the assessment rather than examining test scores. Thus, reliability data are provided for each item individually. The IRT procedure was used to examine which questions on each subtest were the most reliable. The six most reliable questions for each subtest were chosen to be a part of the final assessment. The result was a 30-item assessment that had high reliability. The following Table 6 provides a summary of the number of questions from each area and from which states released items were drawn.

Table 6. Biology Core Content Assessment Content Areas and States

Biology Core Content Area	States and Number of Questions for Released Test Items
Nature of Life	Arizona – 1 Louisiana – 1 Maryland – 1 New York – 2 Oregon – 1 Arizona – 2
Ecology	North Carolina – 1 New York – 1 Oregon – 1 Pennsylvania – 1 Arizona – 1
Cells	California – 1 Illinois – 1 Mississippi – 2 Oregon – 1
Genetics	Arizona – 1 Maryland – 2 North Carolina – 1 Oregon – 2
Evolution	Arizona – 1 New York – 2 North Carolina – 3

Student Survey. Student surveys were administered as both a pretest and posttest to assess attitude change over the duration of the study. The survey included the *Science Motivation Questionnaire* (Glynn & Koballa, 2006), which is a published, reliable scale of student attitudes towards science, and can be customized to address attitudes towards biology specifically. This survey consisted of five subscales as established by the survey creators. Additional questions on the student survey assessed perceptions of teacher influence and product satisfaction. All motivation questions were answered on a 5-point scale, ranging from *never* to *always*.

A factor analysis was conducted by the Cobblestone research team on posttest responses to assess the reliability of underlying constructs. Our results supported the subscales identified by the survey creators. A more specific description of the subscales is provided next. Please see Appendix D for a list of all biology motivation questions used on the student survey with associated reliabilities obtained for our sample.

- *Intrinsic motivation and personal relevance:* According to Glynn and Koballa (2006), intrinsic motivation and personal relevance explained 30% of the total variance in students' responses to the questions assessing motivation and learning. They thus believed this to be the most important factor in student learning. This subscale included 10 questions such as "Understanding science gives me a sense of accomplishment," and "I think about how the science I learn will be helpful to me." This subscale had a high reliability of .903
- *Self-efficacy and assessment anxiety:* Nine questions were used to measure self-efficacy and assessment anxiety, the second-most important factors in measuring motivation according to the authors. Similar to the factor analysis conducted by Glynn and Koballa (2006), self-efficacy and assessment anxiety questions were interdependent, and thus categorized as comprising one subscale. Examples of questions included: "I am confident I will do well on the science labs and projects" and "I become anxious when it is time to take a science test." Reliability for this subscale was .790.
- *Self-determination:* This subscale included four questions which were labeled as being the next most important in motivation to learn science. One question, for example, asked students to indicate how often they "put enough effort into learning science." Another question asked them to respond to the following statement: "If I am having trouble learning science, I try to figure out why." The reliability for this subscale was .743.

- *Career motivation:* Two questions were used to measure extrinsic motivation (i.e., career motivation). As suggested by Glynn and Koballa (2006), results from this subscale should be interpreted with caution because of the small number of questions. These questions include "I think about how learning science can help my career" and "I think about how learning science can help me get a good job." Although there were only two questions on this subscale, it had high reliability of .845.
- *Grade motivation:* Although the original grade motivation subscale included five questions, our factor analysis suggested that one question did not fit in with this scale. This question assessed self-determinism, which the authors suggest "should be revised to make [it] fit better with the other items in this [subscale]" (p. 137). Thus, this subscale was comprised of four questions assessing grade motivation. Sample questions include "Earning a good science grade is important to me" and "I think about how my science grade will affect my overall grade point average." Reliability for this subscale was .709.
- *Perceptions of teacher efficacy.* Four questions were added to the posttest student survey to assess students' perceptions of teacher influence. For example, students were asked to respond to the following statements: "My science teacher explains concepts clearly" and "My teacher inspires me to do my best in science." Reliability for this subscale was .851.

In addition to questions about motivation and teacher efficacy, the posttest also included questions regarding product satisfaction for the individual components of the *Miller & Levine Biology* (2010) program. In this section, students rated their level of satisfaction with components of the curriculum by responding on a scale from 1 (strongly disagree) to 4 (strongly agree).

Study Procedures and Measures Summary

The efficacy study was designed to assess implementation of the curriculum in classrooms, answer research questions related to student achievement and attitudes, and to assess product satisfaction from teachers and students. **Implementation measures** were collected to assess the extent to which students and teachers implemented their respective biology programs in their classrooms. **Outcome measures** were administered as pretest and posttest instruments and assessed the impact on student attitudes and achievement.

Outcome Measures	
Biology Core Content Assessment	Developed using released science questions from various states' Department of Education websites and included 30 multiple choice questions addressing concepts covering the five core units including: the nature of life; ecology; cells; genetics; and evolution.
Stanford Achievement Test 9 (SAT9)	Published instrument including general high school science assessment that included 40 multiple choice science items.
Student attitude survey	Included questions related to students' intrinsic motivation and personal relevance, self efficacy and assessment anxiety, self determination, career motivation, grade motivation and perceptions of teacher efficacy. An additional section on the posttest asked students to rate their satisfaction with elements of the *Miller & Levine Biology* (2010) program.
Implementation Measures	
Online Teacher Logs	Completed by all participating teachers weekly to report the content covered and specific program components used in their classrooms.
Classroom Observations	Observed by researchers, all teachers and their students participated including two times for treatment classrooms (fall and spring) and one time for control classrooms (fall or spring).
Teacher Interviews/ Focus Groups	Completed at the end of the study, most teachers participated in individual interviews or focus groups to discuss the program implementation over the duration of the school year.

Section Four: Assessment of Curriculum Implementation

Implementation is a key factor in a curriculum study because it is possible for implementation of a particular program to vary across sites and teachers. To interpret student outcomes appropriately, it was important to measure implementation within treatment and control classrooms. This study tracked program implementation from the initial training through the final assessment. Through the classroom observations, formal and informal teacher interviews, and online teacher logs, we were able to examine the depth and breadth of the content covered as well as the quality of implementation. The following section provides an analysis of the implementation of the treatment curriculum (specifically focusing on the breadth of coverage and fidelity to implementation guidelines) and implementation in control classrooms. We also address the efficacy study's first research question related to implementing the *Understanding by Design* model. Comparisons between treatment and control curricula are also reviewed.

Treatment Curriculum Implementation

Teachers were required to adhere to specific implementation guidelines requiring the integration of specific components of the *Miller & Levine Biology* (2010) program into their classes. All treatment teachers were given a copy of the implementation guidelines (Appendix E) prior to the start of their school year, so that they would be aware of the program components that were crucial to the curriculum. These guidelines were reviewed with all treatment teachers during study training sessions. The purpose of the implementation guidelines was to ensure that treatment teachers would fully implement the *Miller & Levine Biology* (2010) curriculum as intended by the developers. Guidelines were developed by the *Biology* product development team and Cobblestone researchers and include program components placed in three categories: required, strongly recommended, and optional.

Based on the established implementation guidelines, we tracked the extent to which treatment teachers followed these guidelines throughout the year. Data on teacher level of adherence was retrieved from weekly implementation logs provided by teachers for the entire school year. Teachers were assigned ratings of high, medium and low for overall implementation (see Table 7 below). For level of adherence to implementation, a "High" rating corresponded to the teachers who covered the majority of elements in the *Miller & Levine Biology* (2010) program as required in the implementation guidelines (e.g., total textbook sections covered, number of Chapter Mysteries covered, number of Big Ideas covered). A "Low" rating indicated that a teacher did not fulfill many of the curriculum component requirements when compared to the other treatment teachers. It is noteworthy that some teachers covered sections in the textbook that were in addition to the first five (required) units. We also counted coverage of these additional sections to create the implementation ratings. These ratings were used in later analyses to compare level of implementation to student performance in these classrooms.

Coverage of *Miller & Levine Biology* (2010) Program

In weekly teacher logs, treatment teachers reported which content sections were covered that week in the *Miller & Levine Biology* (2010) textbook and which supplemental materials were utilized in their classrooms. Teachers were allowed to select which version of the textbook to utilize for each individual classroom. As indicated in the implementation guidelines, teachers were required to cover units one through five in the textbooks in addition to any specified in their state standards.

Table 7 provides a complete breakdown of the percent of these required units completed by teachers using both Mainstream and Foundation versions of the textbook, as well as the percent of completed *Understanding by Design* elements. Of the 14 teachers utilizing the *Miller & Levine Biology* (2010) program, 1 teacher taught only the Foundation curriculum, 10 teachers taught only the Mainstream curriculum, 2 teachers used both levels in separate class periods, and 1 teacher (teacher 10 in Table 7) used the Foundation curriculum for the first semester and the Mainstream curriculum in the second semester with the same students. While only Units 1 through 5 are report in Table 7, some teachers covered other units and chapters of the Miller & Levine program to satisfy specific state standards and personal teaching preferences.

Online logs also allowed teachers to indicate all components of the curriculum and supplemental materials that were utilized each week. Based on these weekly teacher logs, the research team determined the frequency of use of the main elements of the *Miller & Levine Biology* (2010) program for all teachers implementing the curriculum. The boxes to the right represent the top five most frequently utilized elements of online elements, classroom activities, lab activities, and teacher resources.

Table 7. Percent of Teacher Implementation for Required Program Components

Teacher	Program	Unit 1	Unit 2	Unit 3	Unit 4	Unit 5	*Chapter Mystery* (19 total)	*Big Idea* (19 total)	Overall Implementation Level**
1	Mainstream	100%	-	86%	39%	57%	63%	84%	Low
2	Mainstream	86%	44%	100%	78%	67%	53%	63%	High
3	Foundation	100%	94%	93%	61%	31%	26%	26%	Med
4	Mainstream	100%	94%	57%	61%	29%	37%	63%	Med
5	Mainstream	100%	94%	86%	50%	50%	–	21%	Med
6	Mainstream	100%	50%	79%	83%	14%	11%	16%	Low
7	Mainstream	100%	97%	100%	100%	79%	47%	79%	High
8	Mainstream	71%	81%	43%	67%	36%	42%	32%	High
9	Mainstream	43%	56%	64%	44%	-	11%	11%	Low
10	Both	29%	69%	86%	39%	29%	37%	5%	High
11(1)*	Mainstream	-	44%	50%	61%	79%	–	–	Low
11(2)*	Mainstream	57%	69%	100%	89%	100%	21%	26%	Med
12(1)*	Foundation	71%	44%	64%	83%	57%	–	–	Med
12(2)*	Mainstream	71%	44%	64%	94%	57%	–	–	Med
13	Foundation	71%	25%	100%	28%	79%	63%	84%	Low
13	Mainstream	71%	50%	93%	56%	86%	37%	16%	Med
14	Mainstream	100%	36%	43%	11%	-	79%	84%	Low

** Teachers 11 and 12 taught semester-length biology classes instead of the standard full year schedule;*

***Overall implementation level was computed by combining the total number of sections covered with other required elements of the program implementation guidelines*

Classroom Activities:
- Section Assessments
- Chapter Assessments
- "Building Vocabulary"
- "In Your Notebook"
- Visual Analogy (Mainstream Only)

Online Activities: (biology.com)
- Art in Motion
- Interactive Art
- Art Review
- Tutor Tubes
- Data Analysis

Teacher Resources:
- Editable Worksheets
- Lesson PowerPoint Presentations
- *ExamView* Assessment Suite
- Teacher's Resource CD-Rom
- Transparencies

Lab Activities:
- *Quick Labs*
- Skills Lab
- Lab Skills #3 Activity
- Inquiry Into Scientific Thinking
- Real-World Lab

Other Implementation Findings:

- Teachers reported using both phschool.com and millerandlevine.com approximately equal times throughout the semester.
- The *Untamed Science* video series was most often implemented in classrooms using the provided DVD's as opposed to the online videos.
- Teachers reported greater use of *Lab Manual A* than *Lab Manual B*.
- A high frequency of use was reported for the online versions of both the teacher and student edition textbooks in the classroom.

Implementation of *Understanding by Design* Model

Research Question 1:

Are teachers able to successfully integrate pedagogical elements of the Understanding by Design *model using the* Miller & Levine Biology *(2010) program?*

The efficacy study aimed to determine if teachers were able to implement *Understanding by Design* elements, a new addition to the *Miller & Levine Biology* (2010) program, as intended by the program developers. As mentioned previously, *Understanding by Design* principles included "*Chapter Mystery*" and "*Big Idea*" components found in each chapter of the textbook. Table 7 above provides a percentage value of "*Chapter Mystery*" and "*Big Idea*" sections that were covered by each teacher based on a total of nineteen possible chapters.

As Table 7 suggests, the implementation of these *Understanding by Design* elements varied greatly across teachers and versions of the textbook utilized. On average, teachers utilized 41% of "*Chapter Mystery*" sections and 44% of "*Big Idea*" sections. The percentage of use of these items varied only slightly between teachers using primarily Mainstream and those using Foundation. The product satisfaction section (Section Six) provides more detailed information regarding teacher and student use and satisfaction with these elements.

Coverage of Non-*Miller & Levine Biology* (2010) Materials

In weekly teacher logs, we also asked teachers to report non-*Miller & Levine Biology* materials that teachers incorporated into their lessons throughout the school year. This could be considered one element of "infidelity" to the program wherein the activities present in the treatment classrooms are very similar to activities in control classrooms (Cordary & Pion, 2006). While the vast majority of instruction in the treatment classrooms came from the *Miller & Levine Biology* (2010) program, some teachers supplemented occasional lessons with non-program materials. The box below shows the most common materials that teachers used to supplement the *Miller & Levine Biology* (2010) program.

Most Commonly Used* Non-Miller & Levine Biology *(2010) Materials:

- Supplemental videos
- Activities & projects (e.g. posters, flip books, coloring diagrams, flash cards)
- Worksheets & handouts
- Labs & experiments
- Exam review (specifically for state testing)

Table 8. Control teacher coverage of five core content areas in control textbooks

Unit Content	T1	T2	T3	T4	T5	T6	T7	T8	T9
Nature of Life	100%	100%	100%	–	–	50%	–	50%	53%
Ecology	100%	56%	100%	100%	–	13%	75%	75%	100%
Cells	100%	–	100%	77%	69%	83%	100%	50%	53%
Genetics	100%	58%	100%	92%	67%	45%	100%	82%	26%
Evolution	100%	33%	100%	78%	22%	67%	100%	67%	24%
Other Units Covered				Invertebrates Vertebrates	Diversity of Life: Invertebrates	Diversity of Life: Plants	Plants		Human Biology

Control Curricula Implementation

Control teachers also tracked coverage of the content material reviewed in class each week via online teacher logs. Based on the information that control teachers reported in these documents, it was evident that content coverage varied greatly across regions and teachers. Table 8 provides a summary of control teacher coverage of concepts in control textbooks. The five content sections addressed (Nature of Life, Ecology, Cells, Genetics, and Evolution) were chosen specifically because they correspond to the primary units (Units 1-5) in the *Miller & Levine Biology* (2010) program as well as the Biology Core Content Assessment. The table also provides information on the other units covered by control teachers.

Classroom Observations

Researchers conducted observations in participating classrooms one or two times during the course of the study. The first set of observations was scheduled for one to two months after implementation began, and the second set of observations was scheduled during the final weeks of the school year. During the observations, researchers documented classroom activities carefully and completed an observation protocol form. Observation protocol forms prompted the research team to gather information about the students in the classroom, instructional variables, teaching materials, teacher variables, and student engagement. Brief summaries of the fall and spring observations at each school site can be found in Appendix F.

Observation Summary. Researchers were able to see many different elements of the *Miller & Levine Biology* (2010) program used during classroom observations. Students were generally engaged in classroom lessons, although this varied by teacher. Although there were a variety of ancillary materials available for use, researchers most often observed students taking notes based on PowerPoint lectures, completing workbook pages, or viewing *Untamed Science* videos or other electronic components. Teachers' ability to integrate many online components was directly related to technology available in their classrooms, and most teachers expressed a willingness to use even more online or electronic components with students, but were limited by their own facilities. However, most teachers reported that the ease of using ancillary materials such as the *ExamView* software, lesson planner and electronic version of the teacher edition text made their course planning easier. There were obvious differences between those classrooms using Mainstream versions of the text versus Foundation versions of the text, in which students in Foundation classes were more prone to off-task behaviors and slower rates of work completion. Teachers generally targeted instruction to the level of students that corresponded to the class level. For those teachers that used both versions of the text, a few complained that the lack of alignment in terms (e.g., "Check Understanding" in the Foundation series was equivalent to section assessments in the Mainstream series). Additional teacher feedback regarding both types of text can be found in Section Six of this report.

Comparing Classroom Environments across Treatment and Control Groups

Classroom observations provided the research team with the opportunity to assess aspects of classroom environment, including classroom management and rapport between teachers and students, in participating classrooms. Classroom observations took place after curriculum implementation had taken place for a few months (October and November 2009) and a spring observation during the final months of the school year (April and May 2010).

Classroom Environment. During classroom observations, teachers were rated by members of the research team and received a rating on two dimensions, teacher/student rapport and classroom management, using a scale from "1" (lowest) to "5" (highest).

Rapport is an indication of the quality of the teacher-student relationship. This relationship has been found to have a profound impact on student achievement (Hattie, 2009).

Qualities of High Student-Teacher Rapport (Hattie, 2009):

- Non-directivity
- Empathy
- Warmth
- Encouragement of Higher Order Learning
- Encouraging Learning
- Adapting to differences
- Genuineness
- Learning-Center beliefs

Classroom Management is a measure of how well the teacher established, controlled, and maintained the learning environment of the classroom with respect to fostering the best possible student behavior through clear expectations.

Qualities of Good Classroom Management:

- Students are engaged and on task
- Teacher responds quickly and effectively to classroom disruptions
- Students provided with clear expectations of their behavior
- Teacher uses a positive and respectful tone in classroom

After each classroom observation, members of the research team discussed ratings provided for each teacher. After carefully reviewing the scoring rubric, most teacher rating scores remained within a 1-point difference and ultimately established inter-rater agreement above 90%. Scores were compiled per teacher and the average score became the associated variable for that teacher in subsequent analysis. Mean teacher ratings of rapport and classroom management were included in our quantitative analyses (Section Five) to investigate the extent to which these variables might be associated with student achievement.

Implementation Summary

To establish construct validity of our implementation fidelity measures, we assessed teachers in a variety of ways including self-reported online teacher logs of coverage of problems, interviews, and classroom observations. The level and quality of implementation varied throughout the study in both conditions, although there did not appear to be an overwhelming advantage for students in either condition in terms of potential quality of the learning environment. In comparison to established implementation guidelines, treatment teachers adhered to most, but not all, implementation guidelines throughout the study. Their adherence appears to be somewhat based on their likability (or dislike) for particular elements of *Miller & Levine Biology* (2010) program as well as their comfort level in using the new curriculum. Treatment teachers used most of the *Miller & Levine Biology* (2010) curriculum in classrooms, however, also supplemented with other non-*Miller & Levine Biology* materials to tailor the learning environment in their classrooms based on standardized test preparation, differing student ability levels, and the incorporation of materials they had found previously effective.

Section Five: Results Related to Students' Attitudes and Achievement in Biology

In this section, we answer the major research questions involving student outcomes in achievement and attitudes. Each research question addressed in this section is listed and followed by a detailed account of the results obtained from the achievement and attitudes data obtained by the outcome measures (i.e., SAT9, Biology Core Content Assessment, and student survey).

Analysis of Outcome Measures

Given that we randomly assigned teachers to the treatment and control conditions, and students were nested within different classrooms (i.e., non-random assignment of students into different classrooms), we used hierarchical linear modeling (HLM) to examine differences in achievement between the treatment and control groups, taking into account various key student and teacher characteristics. HLM models were particularly appropriate for analyzing data of this kind (i.e., students within different classrooms) because they simultaneously examined the effect of student background variables (e.g., ethnicity) and teacher/instructional characteristics (e.g., rapport with students) on students' biology achievement. For a complete discussion of the rationale and theory underlying HLM models, please see Raudenbush and Bryk (2002).

Appendix G describes the HLM statistical model (i.e., random intercept model in STATA) and includes a list of variables and their operational definitions associated with student background characteristics and teacher/classroom/school characteristics that were used in the HLM models. These variables fell into the following four categories: (1) key student demographic background characteristics (e.g., gender, ethnicity); (2) proxy measure of prior biology achievement (i.e., pretest scores); (3) affective measures related to motivation and self-efficacy; and (4) teachers' rapport with students, classroom management, and years of teaching experience.

Research Question 2:

How does student achievement differ for those using the Miller & Levine Biology *(2010) program compared with those using another biology program?*

Student achievement was measured using the SAT9 and Biology Core Content Assessment. The results from the SAT9 were converted into scaled scores for the analyses which can further be interpreted by conversion into grade level equivalents. The Biology Core Content Assessment results are reports by their raw score (i.e., norms were not available for this assessment). The following sections address each achievement measure individually.

It is important to note that there were classrooms using the *Miller & Levine Biology* Foundation series excluded from the HLM analyses for both the SAT9 and Biology Core Content Assessment. These classrooms were excluded from the overall HLM analyses because the control group did not have corresponding and/or equivalent classrooms of this type (e.g., lower level and special needs students). Specifically, there were four class sections (71 students) from two teachers and two schools that were removed from the HLM analysis. These students are included in all further analyses for research questions three through five where appropriate.

Stanford Achievement Test 9

As shown in Table 9, controlling for various student and teacher variables, we found a marginally significant treatment effect for the high implementation treatment group. Students in treatment classrooms characterized as high implementers outperformed students in the control group on the SAT9 assessment (see coefficient associated with "Treatment Implementation-high" in Table 9). No significant differences were observed between the treatment and control groups where the implementation was low or medium on the SAT9 assessment.

Note on statistical significance:

If results are statistically significant when comparing groups (e.g., treatment vs. control; various ethnic groups) this should be interpreted to mean that there is reasonable evidence that a true difference exists between the groups in the study sample (i.e., that the difference is not the result of chance alone). If the results were not statistically significant this should be interpreted to mean that any difference between the groups is a result of chance, thus the groups are performing comparably. Lack of statistical significance should not be interpreted to mean that the treatment (program) had no effect on the performance of the study sample.

There were additional student and teacher characteristics that were significantly associated with predicting the SAT9 posttest scaled scores. When interpreting the results of the HLM analysis, it is important to realize that each variable is reported on after controlling for all other characteristics in the HLM model. In other words, the results of the variables are reported after considering all other characteristics as equal. Using all of the available data gathered on this sample of students, this HLM model was the best fit to the outcomes measured. Essentially, this HLM model could be used to predict a student's score after identifying the student and teacher characteristics that are contained therein. With regard to student characteristics, the following covariates that were significantly associated with students' SAT9 scale scores:

- SAT9 pretest score: higher pretest scores predicted higher posttest scores
- Ethnicity
  - o Latino: predicted a lower score than the reference group (Caucasian)
  - o African American: predicted a lower score than the reference group (Caucasian)
- Language other than English as primary language: predicted higher scores than reference group (English as primary language)
- Mother's education: higher level of education by a students' mother predicted higher scores
- Students' self-efficacy: a higher rating on the self-efficacy construct of the student survey predicted a high score

Table 9. HLM Results for SAT9 Scaled Scores (*n* = 1,315)

Fixed Effect	Coefficient	Standard Error	Approx. T-Ratio	p-value
SAT9 Pretest	0.65	0.03	20.80	< 0.01
Treatment Implementation-low	4.85	3.58	1.35	0.18
Treatment Implementation-medium	-1.24	4.78	-0.26	0.80
Treatment Implementation-high	7.89	4.59	1.72	0.09
Male	2.60	1.71	1.52	0.13
Latino	-7.45	4.05	-1.84	0.07
African American	-13.85	3.32	-4.17	< 0.01
Other ethnicity	-2.58	2.33	-1.11	0.27
English is not primary language	11.41	5.57	2.05	0.04
Mother's education	1.34	0.68	1.97	0.05
Grade 10 indicator	-3.19	2.06	-1.54	0.12
Junior and senior indicator	1.43	4.90	0.29	0.77
Rapport	6.69	2.47	2.71	0.01
Teacher classroom management	4.28	2.30	1.86	0.06
Student motivation	0.34	1.27	0.27	0.79
Student self-efficacy	3.76	1.38	2.73	0.01
Student motivation for grades	0.28	1.32	0.21	0.83
Teacher years of teaching experience	0.44	0.18	2.43	0.02
Intercept	179.77	22.55	7.97	< 0.01

In addition, we found the following teacher level variables that were significantly related to students' SAT9 scores:

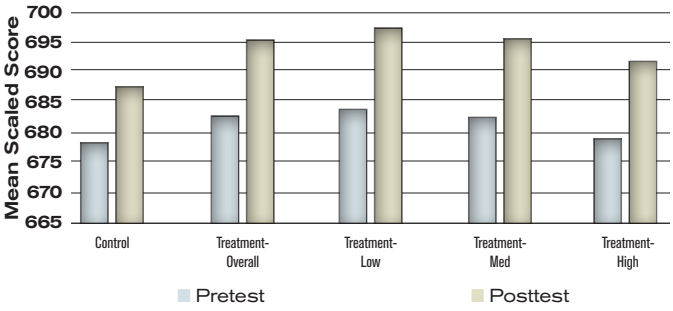
- Teachers' rapport with students (continuous scale; 1 = low, 5 = high): higher rapport between students and teacher predicted higher student scores
- Classroom management (continuous scale; 1 = low, 5 = high): better classroom management predicted higher student scores
- Years of teaching experience: more teaching experience by classroom teachers predicted higher student scores

Although the other variables in the HLM model are not significant, they are included because they provided the best fit for the SAT9 data (i.e., they are theoretically meaningful and provide more precision in the overall prediction of the SAT9 scaled score).

To further explain and explore the results of the HLM analysis, Figure 1 shows the pretest and posttest scaled scores of the SAT9 assessment disaggregated by control and treatment implementation level (i.e., low, medium, high). While each of these groups increased their scores from pretest to posttest, the students in the control group increased an average of about 8 points compared to the treatment group's average increase of about 12.5 points (see Table 10). While this rate of change for the treatment group versus the control group is statistically significant when ignoring other teacher and student variables, $F(1, 1313) = 9.967$, $p < .01$, the more appropriate HLM analysis shows treatment and control differences only between the control group and high implementation treatment group.

It is important to understand the context of these data, specifically counterintuitive findings such as the non-English speakers predicting higher SAT9 scores. For a more in-depth discussion regarding the interpretation of the HLM analyses and accompanying examples, please see Appendix H. Based on these data, the biggest predictor of the posttest SAT9 scaled score is the pretest score. This is followed by teacher/classroom variables (e.g., teacher rapport with students, classroom management) and student attitudes (e.g., self-efficacy) and several other student characteristics (e.g., study condition, ethnicity).

Figure 1. Pretest and Posttest SAT9 Scaled Score: Control versus Treatment Implementation Levels



Biology Core Content Assessment

A similar analysis was conducted for the Biology Core Content Assessment. With respect to this assessment, we found no significant treatment effect between the treatment and control groups after controlling for various student and teacher level variables (see Table 11).

Among the various student and teacher characteristics, we found the following covariates were significantly associated with students' Biology Core Content Assessment scores:

Table 10. Scaled Scores and Grade Equivalents for SAT9 Assessment

	Pretest Average Percentile*	Posttest Average Percentile*	Pretest Grade Equivalent	Posttest Grade Equivalent**	Pretest Scaled Score	Posttest Scaled Score	Difference in Scaled Score
Control (*n* = 593)	56.5	59.9	10.6	12.3	678.52	686.66	8.14
Treatment-Overall (*n* = 722)	60.9	70.3	11.1	PHS	682.30	695.13	12.83
Treatment-Low (n = 282)	59.8	69.8	11.5	PHS	683.62	696.81	13.19
Treatment-Med (*n* = 282)	65.0	73.3	11.3	PHS	682.61	695.38	12.77
Treatment-High (*n* = 158)	55.6	66.0	10.6	PHS	679.37	691.68	12.31

** The percentile score is calculated by students' grade level. This reported percentile score is the weighted average of the 9th grade and 10th grade students' percentile scores.*

See Appendix L for the percentile scores by grade level.

*** PHS = post high school.*

- Biology Core Content Assessment pretest score: higher pretest scores predicted higher posttest scores
- Ethnicity
  - o Latino: predicted a higher score than the reference group (Caucasian)
  - o African American: predicted a lower score than the reference group (Caucasian)
- Grade 10 students: predicted lower scores than the reference group (9th grade students)
- Students' intrinsic motivation: a higher rating on the intrinsic motivation and personal relevance construct of the student survey predicted a high score
- Students' self-efficacy: a higher rating on the self-efficacy construct of the student survey predicted a high score
- Students' motivation for grades: a higher rating on the grade motivation construct of the student survey predicted a high score

In terms of teacher level variables, the following were significantly related to students' Biology Core Content Assessment scores:

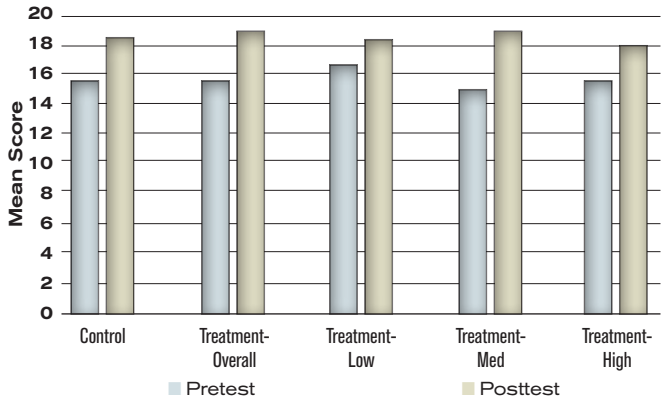
- Teachers' rapport with students (continuous scale; 1 = low, 5 = high): classroom rated as higher rapport between students and teacher predicted higher student scores
- Classroom management (continuous scale; 1 = low, 5 = high): classrooms rated as better in classroom management predicted higher student scores

Figure 2 shows the mean scores from the control and treatment groups from pretest to posttest on the Biology

Table 11. HLM Results for Biology Core Content Assessment Scores (*n* = 1,356)

Fixed Effect	Coefficient	Standard Error	Approx. T-Ratio	p-value
Biology Core Pretest	0.68	0.03	19.84	< 0.00
Treatment Implementation-low	-0.45	0.71	-0.63	0.53
Treatment Implementation-medium	-1.49	0.95	-1.57	0.12
Treatment Implementation-high	0.02	0.92	0.02	0.98
Male	-0.50	0.33	-1.49	0.14
Latino	1.30	0.78	1.68	0.09
African American	-1.65	0.63	-2.62	0.01
Other ethnicity	-0.47	0.45	-1.04	0.30
English is not primary language	0.24	1.05	0.23	0.82
Mother's education	0.07	0.13	0.54	0.59
Grade 10 indicator	-0.83	0.41	-2.06	0.04
Junior and senior indicator	-1.28	0.91	-1.40	0.16
Rapport	1.36	0.49	2.76	0.01
Teacher classroom management	1.11	0.45	2.46	0.01
Student motivation	0.65	0.25	2.65	0.01
Student self-efficacy	0.96	0.27	3.51	< 0.00
Student motivation for grades	0.50	0.26	1.92	0.05
Teacher years of teaching experience	0.05	0.04	1.33	0.18
Intercept	-8.26	2.69	-3.07	< 0.00

Figure 2. Pretest and Posttest Biology Core Content Assessment: Control versus Treatment



Core Content Assessment. The overall control and treatment pretest to posttest difference was almost identical (see Table 12). The medium implementation treatment group showed the largest gain (3.93 points) among the groups represented in Figure 2 and Table 12.

Table 12. Biology Core Content Assessment Pretest and Posttest Scores

Biology Core Content Assessment			
	Pretest	**Posttest**	**Difference**
Control (*n* = 577)	15.49	18.24	2.75
Treatment-Overall (*n* = 779)	15.58	18.55	2.97
Treatment-Low (*n* = 290)	16.53	18.38	1.85
Treatment-Med (*n* = 333)	14.87	18.98	4.11
Treatment-High (*n* = 156)	15.34	17.97	2.63

Research Question 3:

How do students with different characteristics (e.g., English learners, various ethnicities) using the Miller & Levine Biology *(2010) program perform on student-related outcomes?*

Research question three assessed how students with different demographic characteristics in the treatment group performed on the SAT9 and the Biology Core Content Assessment. The HLM analysis showed that ethnicity and primary language were significant predictors for the SAT9 scaled score (see Table 9 above). The HLM analysis for the Biology Core Content Assessment also showed just ethnicity was a significant predictor for overall score (see Table 11 above). Our analyses found no other significant student characteristics that could be used to predict the posttest scores on either achievement measure. The following sections address these student characteristics (i.e., ethnicity and primary language) by providing descriptive information. It is important to note that reporting the data this way ignores the relationships between variables that were found to exist in the HLM analyses.

Ethnicity

Specifically for the SAT9, results showed that overall (i.e., treatment and control) Latino and African American students performed significantly lower than Caucasian students (see Table 9 above) after controlling for the other student and teacher characteristics. Figure 3 shows the treatment group's pretest and posttest scaled scores on the SAT9. Despite the significantly lower scores according to the HLM analysis, Latino students in the treatment group showed a larger overall change from pretest to posttest than Caucasian students (see Table 13). In addition, African American students in the treatment group also showed an increase from pretest to posttest (see Table 13).

Figure 3. Pretest and Posttest Treatment SAT9 Scaled Scores by Ethnicity

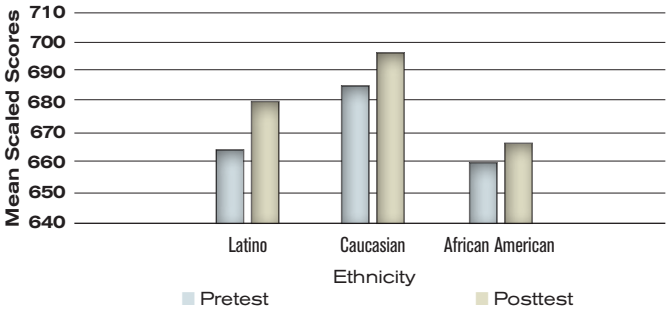


Table 13. Treatment Pretest and Posttest SAT9 Scaled Scores by Ethnicity

SAT9 Scaled Scores				
	Pretest	**Posttest**	**Difference**	**% increase**
Latino (*n* = 60)	663.25	679.93	16.68	2.5%
Caucasian (*n* = 503)	685.63	698.19	12.56	1.8%
African American (*n* = 61)	660.00	665.79	5.79	0.9%

For the Biology Core Content Assessment HLM analyses, Latino students had a marginally significant result in predicting the posttest score as being *greater* than Caucasian students all other teacher and student characteristics being equal. While this finding may seem contradictory to the descriptive information seen in Table 14, it is important to remember that the HLM analysis is controlling for the other variables in the model. The results also showed African American students performed significantly worse than Caucasian students overall (see Table 11 above). Figure 4 shows the treatment group's mean scores for the Biology Core Content Assessment divided into these ethnic groups. Again we see that Latino students had a greater difference between pretest and posttest than Caucasian and African American students (see Table 14).

Figure 4. Pretest and Posttest Treatment Biology Core Content Assessment Scores by Ethnicity

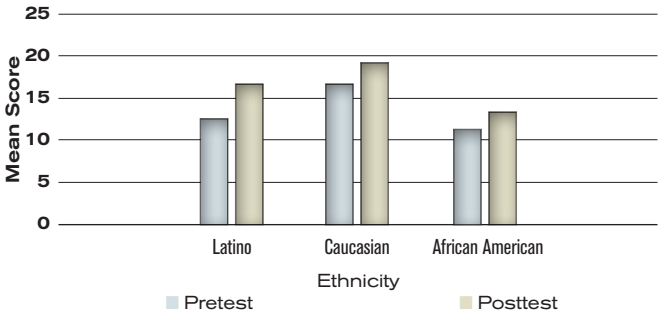


Table 14. Treatment Pretest Biology Core Content Assessment Scores by Ethnicity

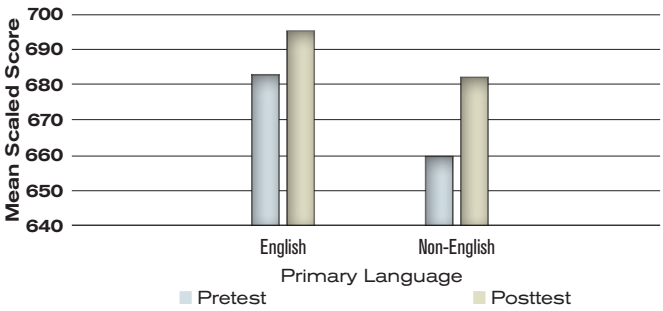
Biology Core Content Assessment				
	Pretest	Posttest	Difference	% increase
Latino (*n* = 83)	12.49	16.23	3.74	29.9%
Caucasian (*n* = 515)	16.21	19.28	3.07	18.9%
African American (*n* = 75)	11.07	13.16	2.09	18.9%

Primary Language

The SAT9 HLM analysis showed that students' primary spoken language was a significant predictor such that, after controlling for the other student and teacher characteristics, students speaking a non-English primary language predicted higher scores on the SAT9 than English speakers. Figure 5 shows the treatment group's pretest and posttest scores for these two groups on the SAT9.

Table 15 also shows the descriptive results of students speaking English and non-English primary languages. The non-English group shows a larger rate of change than

Figure 5. Pretest and Posttest Treatment SAT9 Scaled Scores by Primary Language



the English speaking students from pretest to posttest in the treatment group. Despite the larger increase in scores, non-English speaking students did not score as high on the SAT9 as English speaking students.

Table 15. Treatment Pretest and Posttest SAT9 Scores by Primary Language

SAT9 Scaled Scores				
	Pretest	Posttest	Difference	% increase
English (*n* = 695)	682.69	694.73	12.04	1.8%
Non-English (*n* = 30)	659.53	682.1	22.57	3.4%

Research Question 4:

What is the relationship among students' attitudes towards science and science achievement?

Research question four examines the relationship between student attitudes and achievement in biology class. Specifically, the student survey measured the overall constructs of intrinsic motivation and personal relevance, self efficacy and assessment anxiety, self determination, career motivation, grade motivation, and teacher satisfaction (see Section Three of this report for a broader explanation of the student survey components). Of these constructs, the HLM model fit best with intrinsic motivation and personal relevance, self efficacy and assessment anxiety, and grade motivation. The HLM analysis showed that the student's scores on all three of these constructs were significant predictors of the Biology Core Content Assessment while only self efficacy and assessment anxiety was a significant predictor on the SAT9 (see Tables 9 and 11). This finding suggests that the more positive certain student attitudes are towards science the better students will perform on science assessments.

Research Question 5:

How do students using the Miller & Levine Biology *(2010) program perform from pretesting to posttesting on assessments related to attitudes about science and achievement in science?*

Research question five examines the specific results of only those students using the *Miller & Levine Biology* (2010) program from pretest to posttest for the major outcome variable of the student survey and achievement measures (i.e., SAT9 and Biology Core Content Assessment).

Treatment Student Survey Results

The following figures display the pretest to posttest results of the constructs obtained from the student survey. The results of the t tests performed on each construct are displayed in Table 16. Figure 6 shows the pretest and posttest scores for student survey constructs for all the participating treatment and control students (i.e., students using the Mainstream and Foundation textbooks). Overall, students' attitudes decreased from pretest to posttest. This trend was consistent when just looking at Mainstream students; however, the students using a Foundation textbook ($n = 148$) showed no significant changes from pretest to posttest on any of the survey constructs. These results indicate that students using the *Miller & Levine Biology* (2010) program did not have improved attitudes about science over the course of the year, but in fact had more negative attitudes about science and their own skills over the course of the year.

Treatment Student Achievement Measures

Those students using the *Miller & Levine Biology* (2010) program significantly increased their achievement

Figure 6. Pretest and Posttest Scores for Treatment Students

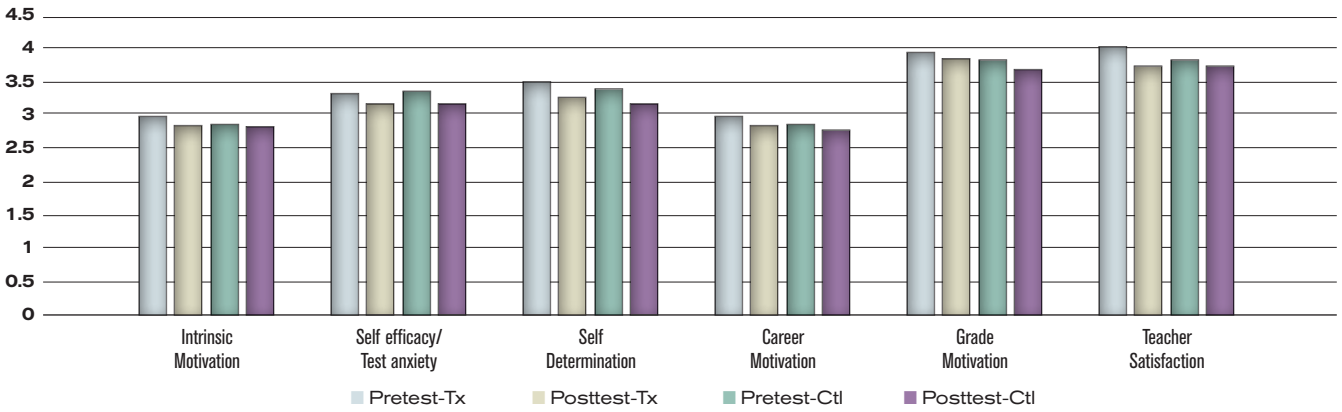


Table 16. Pretest and Posttest Student Survey Scores for Treatment Group Only

Composite Factors	Pre-test Mean	Post-test Mean	*SD*	*df*	Difference	*t*
Intrinsic Motivation and Personal Relevance (10 items, alpha = .903)	2.97	2.89	0.69	823	-0.08	3.44*
Self-Efficacy and Assessment Anxiety (9 items, alpha = .790)	3.25	3.17	0.64	823	-0.08	3.56*
Self-Determination (4 items, alpha = .743)	3.47	3.19	0.73	823	-0.28	10.94*
Career Motivation (2 items, alpha = .845)	2.97	2.83	1.14	820	-0.14	3.60*
Grade Motivation (5 items, alpha = .709)	3.94	3.83	0.77	823	-0.11	4.07*
Perception of Teacher Efficacy (4 items, alpha = .851)	3.96	3.70	1.02	794	-0.26	7.01*

Note. Scale ranged from 1 to 5 (1 = Never, 5 = Always). All relationships are significant for this table

** $p < .001$*

scores from pretest to posttest on the SAT9 and the Biology Core Content Assessment. Furthermore, the results are similar when we examined the students using the Mainstream and Foundation textbooks individually. The following figures show the overall and textbook-specific results for each of the achievement measures. Specifically, the overall results of the SAT9 are discussed followed by the results of the Biology Core Content Assessment. The overall results from the t tests are show in Table 17.

Figure 7. Treatment Students Pretest and Posttest Science Achievement Results (SAT9)

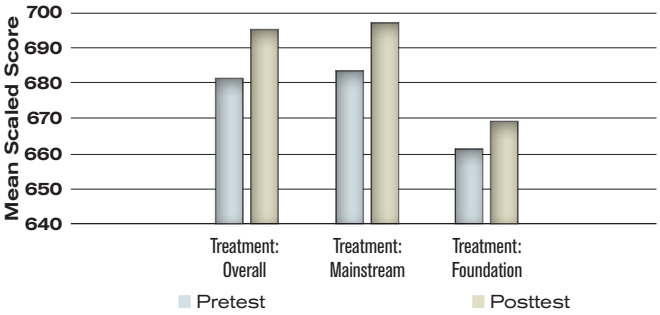
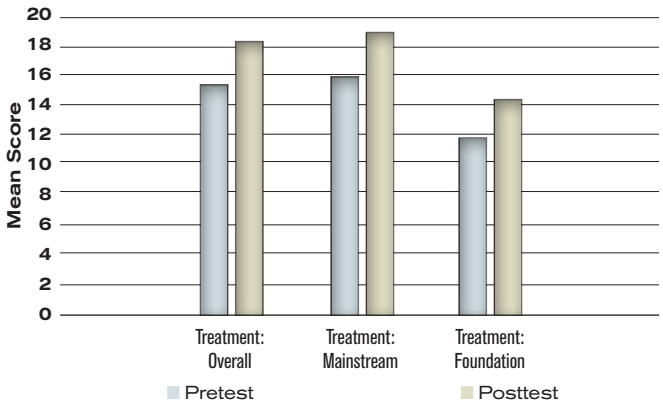


Figure 7 shows the pretest and posttest scores of the SAT9 overall results with textbook specific results. The increases in scores from pretest to posttest are highly significant for each analysis (see Table 17). Figure 8 shows the pretest and posttest scores of the Biology Core Content Assessment overall results with textbook specific results. Again, the increases in scores from pretest to posttest are highly significant for each analysis. Table 17 shows results of the paired t tests that were used to analyze the achievement measures.

Figure 8. Treatment Students Pretest and Posttest Science Achievement Results (Biology Core Content Assessment)



Attrition

We conducted attrition analyses based on a comparison of the student sample obtained from pretesting to posttesting. A full description of the study attrition and differential attrition (comparing treatment and control groups) can be found in Appendix I. Attrition analyses suggested some minor differences between treatment and control in terms of demographic characteristics that did not appear to affect the results obtained in Section Five. The analyses of the SAT9 and Biology Core Content Assessment showed no differences between control and treatment for the attrition students, $t(271) = .90$, *ns* and $t(270) = 1.25$, *ns*, respectively. This result combined with the results shown in Appendix I gives us confidence that the results of attrition are not a threat to the overall results of Section Five.

Table 17. Treatment Students Pretest and Posttest Achievement Scores

	Pretest Mean	Posttest Mean	Difference	*t*	*df*	*SD*
SAT9						
Overall	681.17	693.62	12.45	13.10**	769	26.36
Mainstream	683.48	696.52	13.04	12.70**	687	26.93
Foundation	661.26	669.36	8.10	3.20*	68	21.05
Biology Core Content Assessment						
Overall	15.34	18.28	2.94	14.72**	825	5.74
Mainstream	15.97	18.91	2.94	13.09**	700	5.95
Foundation	11.55	14.22	2.67	6.63**	109	4.22

** p < .01; ** p < .001*

Summary of Major Findings

Research Question 1: The implementation of the *Understanding by Design* elements varied across teachers and versions of the textbook used. On average, teachers used 41% of "*Chapter Mystery*" sections and 44% of "*Big Idea*" sections; this differed for users of the Mainstream and Foundation books in which teachers using the Mainstream textbook used these features more consistently.

Teachers found both the "*Chapter Mystery*" and "*Big Idea*" components to be an effective means of connecting main ideas and engaging students. Teachers felt that students enjoyed the "*Big Idea*" and that it was useful in focusing student attention on the core concepts. Teachers most often utilized "Chapter Mysteries" as a means to introducing the new chapter and felt they were successful for that purpose.

Research Question 2: Differences between the students using the *Miller & Levine Biology* (2010) program and students using a different program were evident on the SAT9 test and only among classrooms where the teacher was classified as a high implementer of the program. However, no differences were found between the control and treatment groups for the Biology Core Content Assessment. These results suggest that implementation of the program was important such that students receiving more of the Miller & Levine program have more positive results related to achievement than students using other biology programs. Also, because the SAT9 science test had more questions related to science in general and less biology specific questions than the Biology Core Content Assessment, the results may indicate that the Miller & Levine program may provide students with a broader science education than other programs.

Research Question 3: While there were a number of student characteristics that were shown to be related to control and treatment scores on the SAT9 (i.e., ethnicity, primary spoken language) and Biology Core Content Assessment (i.e., ethnicity), students from these demographic groups using the *Miller & Levine Biology* (2010) program showed significant growth from pretest to posttest on both assessments. Specifically, Latino and African American students had significant gains from pretest to posttest with Latino students showing larger gains than Caucasian students from pretest to posttest on both measures. Results also showed students speaking a non-English primary language showed significant gains from pretest to posttest.

Research Question 4: Findings suggest that the more positive certain student attitudes are towards science the better students will perform on science assessments. The HLM analyses showed intrinsic motivation and personal relevance, self efficacy and assessment anxiety, and grade motivation as significant predictors of student achievement. Specifically, all three of these constructs were significant predictors of the Biology Core Content Assessment while only self efficacy and assessment anxiety was a significant predictor on the SAT9.

Research Question 5: Overall, the students using the Miller & Levine program showed significant decreases from pretest to posttest on their attitudes and motivation toward science as measured by the student survey; however, results were virtually identical to the control groups' pretest and posttest ratings. These results were not consistent to the students using the Foundation textbook who showed no statistical difference between their pretest and posttest ratings. For the achievement measures, student in the treatment group showed highly significant gains from pretest to posttest on both the SAT9 and Biology Core Content Assessment. These results were the same overall and regardless of textbook (i.e., Mainstream or Foundation) used in the classroom.

Section Six: Product Satisfaction

Product satisfaction of the *Miller & Levine Biology* (2010) program was assessed using input and feedback from multiple sources regarding program use and satisfaction in participating classrooms. Data sources include closed and open-ended survey items on the student post-test survey, teacher interviews, and teacher focus groups. The information provided in this section is summarized from these sources and assembled according to elements of the *Miller & Levine Biology* (2010) program.

Student ratings identified in this section were provided on the student survey. Students were asked to rate how much they like or dislike each part of the *Miller & Levine Biology* (2010) program on a scale from 1 = Strongly Dislike to 4 = Strongly Like, which also included an option for them to report "N/A" or "Did not use." Appendix J provides a summary of student ratings related to product satisfaction, separated by Foundation and Mainstream users, as well as combined ratings for all students. In addition, we have provided a summary of product satisfaction specifically related to the Foundation series in Appendix K.

Student-Edition Textbook

We asked both treatment and control students to rate the degree to which they agreed to certain statements about their biology textbook, including specific features of the text. Figure 9 shows student ratings of the biology texts. Students using the *Miller & Levine Biology* (2010) program liked their text significantly more than control students in all categories ($p < .01$). The only items in which control students provided a higher rating were the following items: "I had to read sections of my Biology textbook multiple times before I understood the concepts" and "My Biology textbook was boring to read". Obviously, higher ratings on these questions were perceived as more negative views of the text. Given these data and the significantly higher ratings provided for all other aspects of the text, it is clear that students liked the *Miller & Levine Biology* (2010) program more than competitor programs.

Full Text of Biology Textbook Satisfaction Statements from Figure 9

Overall Rating: "I like my Biology textbook"

Pictures/Diagrams: "The pictures and diagrams in my Biology textbook helped me to understand the concepts."

Easy to Read: "My Biology textbook was easy to read."

Learning: "I have learned a great deal from my Biology textbook."

Enjoy experiments: "I enjoyed doing experiments and labs from my Biology textbook/ lab manual."

Explained Concepts: "My Biology textbook explained difficult concepts in a way that I could understand."

Locate Main Points: "In my Biology textbook it was easy to locate the main points of each paragraph."

Read Multiple Times: "I had to read sections of my Biology textbook multiple times before I understood the concepts."

Closed and open-ended student comments revealed several components of the student edition textbook that were well-liked by student participants (See Figure 10). Overall, students using the Foundation version of the text rated their material more enjoyable than those using the Mainstream version. Combined ratings suggest that both Mainstream and Foundation users reported that they enjoyed Building Vocabulary sections, or highlighted words, and the pictures of scientific concepts the most. On

Figure 9. Mean Ratings of Biology Textbook Features: Treatment vs. Control

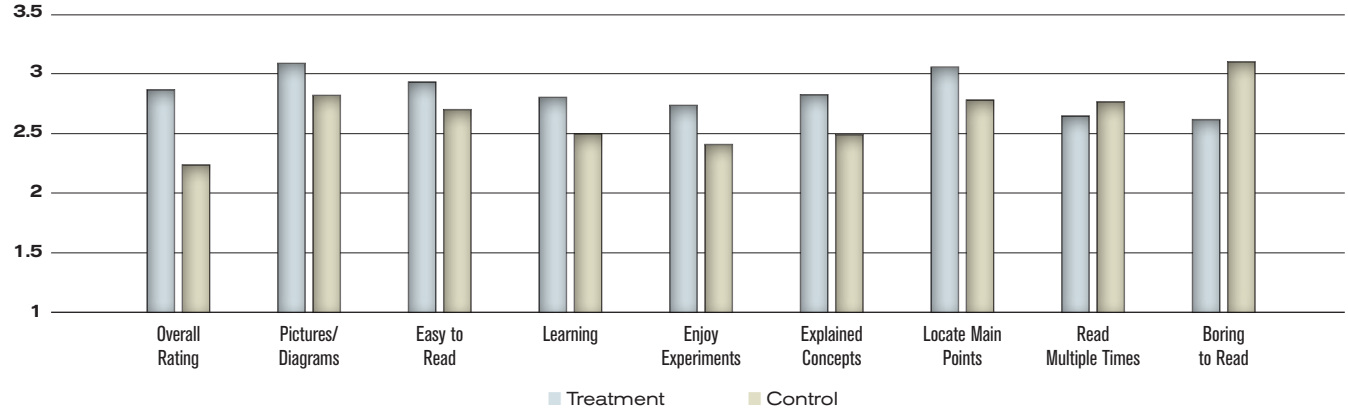


Figure 10. Student Favorite and Least Favorite Elements of the *Miller & Levine Biology* (2010) Program

Favorite	Least Favorite
Aesthetics: Students found the diagrams, pictures, visual analogies, and images to be a favorite element of the textbook. "The pictures and diagrams…they made it easy to visualize and understand the process of things." ***Ease of Comprehension & Use:*** Students felt that the book was well organized, making important information easy to access. "…the book was organized in a nice way so that main points were easy to find, summarize, and understand." ***Key Concepts & Vocabulary Terms:*** Many students liked the presentation of the Key Concepts and vocabulary terms. The highlighting of the terms and the provided definitions were user-friendly.	***Chapter Assessments:*** Students found chapter assessments to be lengthy and repetitive. ***Some Areas Confusing:*** Students reported that the book was confusing in some areas; specifically some diagrams were unclear and hard to understand. ***Weight & Size:*** Students found the book to be large and bulky. For those with access to the internet from their homes, students appreciated the ability to access the textbook online so that they did not have to carry their textbook.

average, no particular component was aggregated as being *disliked*, however both Foundation and Mainstream users rated "Analyzing Data" as their least enjoyable (Mean = 2.41 out of 4).

Understanding by Design

Understanding by Design (UbD) components included Big Ideas and Chapter Mysteries. Reactions to UbD elements were mixed among teacher and student users. Student participants listed Chapter Mysteries as one of their least favorite elements (Mean student rating = 2.64 out of 4) ; however several teachers found both the *Chapter Mystery* and *Big Idea* components to be an effective means of connecting main ideas and engaging students.

Teachers felt that students enjoyed the "*Big Idea*" and that it was useful in focusing student attention on the core concepts. One treatment teacher suggested that it might be improved by carrying through more of the chapter. Students gave the "*Big Idea*" a mean rating of 2.62 out of 4, indicating that it was not the most strongly liked aspect of the program.

"I liked the big question idea because it always comes back to that one thing. We always introduce the new chapter and the new unit with the big question and then periodically come back to that. I felt [that] was a real strength."

–Teacher using *Miller & Levine Biology* program

"I think it [the *Big Idea*] focused them in on what [is] the significance or what are we getting from this. I think it makes the teacher's job easier overall."

–Teacher using *Miller & Levine Biology* program

Teachers providing feedback on* Chapter Mystery *suggested the following improvements to this component:

- answers to the mysteries should not be in the back of the chapter
- some mysteries need to be enhanced for higher-level students
- hints or prompts could be provided to bridge the chapter information with the mysteries
- more structured information should be provided for students to access to solve mysteries

Differentiated Instruction

Overall, treatment teachers reported that differentiated instruction elements of the text were helpful and provided useful ideas and suggestions. Several teachers using control curriculum packages felt their textbooks lacked variety and opportunities for reaching different students. Teachers in higher-level biology (e.g., honors) classrooms would have liked to see more higher-order thinking in the biology materials, including more in-depth review questions and scenarios.

Chapter Mystery

Teachers most often utilized the *Chapter Mystery* as a means to introducing the new chapter and felt they were successful for that purpose. Specifically, teachers identified these chapter mysteries as most well-liked:

- "Death by Salt Water" (Chp. 27)
- "Such Varied Honeycreepers" (Chp. 16)

Teachers had mixed reactions to using different student edition versions of the textbook, Mainstream or Foundation. Despite the lower level elements of the Foundation version of the textbook, some teachers still found the Foundation textbook to be too advanced in terms of reading level. Another teacher thought the Foundation text was lacking content for school curriculum, resulting in difficulty in testing for students using the Foundation text. The main complaint from teachers utilizing both versions of the textbook was that they disliked how the elements of each were labeled differently.

Online Resources

Teachers overwhelmingly agreed that the online resources provided with the *Miller & Levine Biology* (2010) program were one of the strongest components of the program. Students also rated online components higher than most of the other components (2.82 out of 4).

Teachers felt that the online elements were useful during classroom activities and as a replacement for transparencies. Several teachers reported being very satisfied with the ability to edit worksheets with the online resources. Teachers also reported liking the ability to access the textbook online, although they were often unable to assign homework online because not all students had internet access in the home.

"I think that the [interactive tutorials online] is one thing that's really strong about the Miller & Levine program. My students are always more engaged when technology is involved. I can tell them go online and do this visual analogy, that's a lot more powerful."

–Teacher using *Miller & Levine Biology* program

Teacher online favorites

Tutor Tubes
InterActive Art
Art in Motion
Real World Inquiry

"There's lots of engaging activities in [the lab manual]. It's unfortunate that it's structured to where it was always at the end of the unit, so you'd almost have to work through the whole chapter to come to an understanding of the work. With as much material as we had to cover, it was not realistic to always be able to cover every section of the chapter."

–Teacher using *Miller & Levine Biology* program

Lab Activities & Lab Manual

The lab activities included in the *Miller & Levine Biology* (2010) program were rated highly by students and teachers. Students felt these activities supplemented their understanding of course material and listed these activities as one of their favorite components of the program in open-ended survey items.

Labs Teachers and Students Enjoyed Most

"Lab Equipment & Safety"(Lab Skills 1, pg. 1)
"Regeneration of Planaria" (Unit 3 Cells, pg. 61)
"Using DNA to Solve Crimes" (Unit 4 Genetics, pg. 91)
"Amino Acid Sequences: Indicators of Evolution" (Unit 5 Evolution, pg. 97)
"Extracting DNA" (Unit 4 Genetics, pg. 73)

Overall, teachers were able to adapt many labs to be useful in the classroom. Teachers did find some lab activities to be very demanding in terms of time and required materials, including even *Quick Labs*.

Teachers commented that they would have liked to have been provided with multiple labs per chapter for better selection between alternatives and that these labs should appear throughout the chapter, not just at the end. Control teachers felt unable to use the majority of labs in their current curricula and were forced to supplement with labs from other sources.

In terms of lab manuals, several teachers reported not utilizing the lab manual and that these manuals needed an upgrade. Student ratings supported this sentiment; lab manuals received low ratings on the student survey (Mean student rating = 2.63 out of 4).

"For going beyond the application of the synthesis level of learning, the workbooks just didn't do it for me."

–Teacher using *Miller & Levine Biology* program

Supplemental Materials

PowerPoints. The majority of treatment teachers utilized the provided PowerPoint presentations for each chapter and several reported satisfaction with this component. Teachers liked the ability to edit and revise presentations to be better-suited to their classrooms, as well as the ability to pull diagrams and pictures from the slides to supplement slides they had created. Teachers did report that the presentations were often too lengthy.

Workbooks. Both teachers and students were dissatisfied with the student workbooks. Students rated workbooks as their second least enjoyable component of the program

and suggested that improvements might include making the book more user-friendly, specifically by making workbook pages perforated for easy removal. Treatment teachers agreed that the workbooks did not work well in their classrooms. Teachers found the workbooks unclear, difficult to grade, as well as lacking in practice worksheets and objective activities.

ExamView Assessment Suite Software. Teachers overwhelmingly reported that the *ExamView* software was a valuable tool for creating tests and quizzes. Teachers found the program easy to use and liked the ability to easily create multiples versions of their test. Teachers did express a desire in having a better variety of question formats, specifically more objective format questions.

> *"I just use the Miller & Levine data base [*ExamView*] for the questions and my students have never scored so well on tests."*
>
> **–Teacher using *Miller & Levine Biology* program**

Untamed Science Video Series. The video series did a good job of supplementing the student experience according to student and teacher feedback. Students rated the video series highly as a component of the curriculum ($M = 2.82$ out of four). Teachers felt that while some of the video seemed slightly "corny" or "cheesy" students seemed to be engaged by the short length of the clips and interesting topics.

> *"The* Untamed Science *was very good for connecting. It did a good job of making the knowledge real and concrete."*
>
> **–Teacher using *Miller & Levine Biology* program**
>
> *"I think the* Untamed Science *videos are horrible and my students mostly think they're too juvenile. I feel that the Untamed videos provided with this series of materials is completely inadequate."*
>
> **–Teacher using *Miller & Levine Biology* program**

Overall Feedback from Teachers Implementing the Miller & Levine Biology *program*

"I really like the book. I love the pictures. I love the diagrams. I love the way stuff is phrased in the book, it hones in on the concept. I really like the prepping part of the book, the standardized test prep at the back, those are good."

"There are not that many weaknesses with the book. I like the set up. I like the way it's designed."

"I think this is the best text book we've had. I think that they're making things interesting by adding the chapter mysteries; I think that's a really good way to start. The activities in the lab manuals; some of those are really good. It's well designed."

"Knowing what actual science looks like and what it means to do science is very interesting– especially in the genetics chapter – some of the 'why this matters' kind of thing..."

"I love the book. I think it is very user friendly. There's just some much information and so much things that you can really tailor to what you need for the class, you kind of pick and choose. I've had parents who have advanced degrees in science make comments about how great it is."

Product Satisfaction Summary

Overall, student and teacher users of the *Miller & Levine Biology* (2010) program were satisfied the curriculum program. The top strengths of this curriculum, as reported by students and teachers, were organization of materials, alignment and relevance of supplemental elements, and finally the pictures, graphics, and figures created to visually represent important information. Organization and information layout were two main complaints of teachers in the control condition. Control teachers also found information in their texts to be incorrect and out of date, an aspect that was well-addressed by the *Miller & Levine Biology* (2010) program.

Some teachers did find the amount of material overwhelming and as well as the amount of ancillary materials that accompanied the text. Others asked that the program include more games and activities teachers could use to supplement the classroom experience.

Section Seven: Discussion

The purpose of the current study was to understand how student attitudes and achievement in biology would be impacted based on use of the *Miller & Levine Biology* (2010) program versus a similar, competitor program. The study included complete tracking of product use and satisfaction with the *Miller & Levine Biology* (2010) program as well as a range of other implementation and outcome measures. The following is a brief discussion of key themes from the efficacy study as well as study limitations.

Efficacy Study Key Findings

Complete Curriculum

Concerns about improving science curriculum as a means to increasing scientific literacy for America's students can be most effectively addressed through creating a comprehensive, complete curriculum. It is clear that teachers and students perceived the *Miller & Levine Biology* (2010) program as a complete curriculum. Although we received occasional comments indicating a desire for more lab exercises or additional *Untamed Science* videos, the program was widely perceived as being a necessary and sufficient program in which to educate students in the content of biology. This is contrasted with everything that we observed in control classrooms. Control teachers normally compiled a variety of resources to pull together lessons for their classes. Although the publication date of all control curricula was older than a 2010 copyright, the most common sentiment expressed toward control curricula was dissatisfaction with both the core text as well as most ancillary components. While control teachers spent time compiling lessons from a variety of online and other sources, those using the *Miller & Levine Biology* (2010) program spent that time selecting from a variety of published program resources. While most of the treatment teachers reported spending extra time planning lessons using the new program, actual use of the new materials (e.g., lesson planner and *ExamView*) actually made planning more efficient. So, it is evident that the wide variety of materials available (and in multiple formats) was perceived as a huge improvement over previous curricular options. Further, it is apparent that teachers and students using the *Miller & Levine Biology* (2010) program perceived it as a complete biology curriculum, and not just a collection of core materials with non-significant ancillary materials.

Students and Teachers Liked the *Miller & Levine Biology* (2010) Program

As previously mentioned, most control teachers expressed the opinion that they disliked their textbook and chose to use material outside of the textbook to teach core concepts. In fact, control teachers were much less likely to draw the source of course content from textbooks. We observed them using a variety of materials, including websites, teacher-created transparencies and handouts. This is contrasted with the frequent usage of the *Miller & Levine Biology* (2010) textbook and core ancillary materials that we observed in classrooms (and subsequently reported in logs). Not only did teachers and students use their books more frequently, but they also liked the materials when they used them. Students were especially likely to report how features such as the art/graphics, key questions, and highlighted vocabulary terms made learning concepts easier. Generally speaking, we observed more consistency across classrooms for those using the *Miller & Levine Biology* (2010) program. This is encouraging in that while teachers can be skilled in a variety of areas, the responsibility of designing one's own biology curriculum exceeds what school and district administrators should expect from a professional teacher. The burden of addressing the numerous state and district content standards, as well as effectively addressing multiple student learning levels simultaneously is only compounded by the responsibility to write lessons, develop exercises, labs, exams, for any teacher. Having a program that effectively makes planning and teaching easier is addresses a strong need for any science classroom. This is perhaps one of the strongest features of the *Miller & Levine Biology* (2010) program, where its strength lies in the ability for teachers of various experience levels to use the program without having the reinvent the wheel for each lesson. These facts, coupled with students' strong preference for the Miller & Levine program in comparison to control text, provides a strong case that teachers and students alike preferred to use the *Miller & Levine Biology* (2010) program.

Factors that Predicted Student Performance

Aside from previous performance (pretest), teacher and classroom characteristics had the most impact on student achievement, with teacher and student rapport having the greatest impact on students' scores. These are followed closely by student attitudes, primarily students' self-efficacy. In fact, individual student demographics had little impact on the predicted score when compared to teacher or classroom characteristics and student attitudes. For example, there was only a 1.3 point difference to between Latino and Caucasian students while a student's self-efficacy could predict a difference of up to 3.84 points. These findings are noteworthy because we understand that teachers are an essential component to the learning

process for students. Further, any program or materials that can facilitate student and teacher rapport would only enhance student performance. While we cannot definitely state that the use of the *Miller & Levine Biology* (2010) program caused student achievement to increase, we can, however, identify the trend that those who fully implemented the program had significant increases in student achievement beyond their control counterparts. Only future research that would artificially manipulate implementation level could address the causal question more directly. However, we at least understand that those students who were in classrooms where the program was used as intended had the best performance.

Reconciling Differences in Student Performance

In our analysis, we also need to reconcile differences in student performance for the two primary outcome measures—the SAT9 and the Biology Core Content Assessment. As a reminder, the SAT9 had more general science questions in comparison to the Core Content Assessment, which relied only on questions specific to the core areas of biology (i.e., Nature of Life, Ecology, Cells, Genetics, and Evolution). High implementing treatment students performed significantly better than control students on the SAT9 test, but no differences were found for the Biology Core Content Assessment. We believe the following explanation is a possibility for these discrepant results. The SAT9 requires students to answer more general life science questions, and in fact may require students to generalize the information that they learned in biology class. It is possible that the *Miller & Levine Biology* (2010) text helps students develop general science knowledge in addition to the specific biology content knowledge that was available on the Core Content Assessment. This may be due to features such as the *Big Idea* question, which directed students to reflect on the bigger picture when learning the content in each chapter. Instead of learning isolated biology concepts (similar to those in the control group), treatment students were constantly, and intentionally, being reminded that the concepts in science all tie together. While the Core Content Assessment showed that the central concepts of Biology were taught and learned by students in both the control and treatment groups, the SAT9 showed that students in the treatment group could apply their knowledge more broadly.

Study Limitations

Given that implementation level was not randomly assigned (e.g., teacher behavior and decisions about program used determined implementation level) we are less confident in our ability to state that high implementation of the program caused students to have higher achievement in biology. It is possible that better teachers were better implementers of the program, and consequently this was the most important factor influencing students' scores. However, it is clear that those students in classrooms where more of the *Miller & Levine Biology* program was implemented did have higher achievement in one measure of biology (SAT9). Future efforts to examine this causal relationship could specifically manipulate this factor to see the extent to which it predicts higher student achievement beyond other teacher and student characteristics.

References

Ball, D.L. & Cohen, D.K. (1996). Reform by the book: What is: Or might be: The role of curriculum materials in teacher learning and instructional reform. *Educational Researcher*, 25, 6-8, 14.

Baker, F. (2001). *The basics of Item Response Theory*. ERIC Clearinghouse on Assessment and Evaluation. University of Maryland, College Park, MD. Retrieved from http://echo.edres.org:8080/irt/baker/

Cordray, D. S. & Pion, G. M. (2006). Treatment strength and integrity: Models and methods. In R. R. Bootzin & P. E., McKnight (Eds.), *Strengthening research methodology: Psychological measurement and evaluation* (pp. 103-124). Washington, DG: American Psychological Association.

Forgione, P.D. (1998). *Achievement in the United States: Progress since a nation at risk?* National Center for Education Statistics Office of Educational Research and Improvement, U.S. Department of Education. April 3, 1998. http://nces.ed.gov/Pressrelease/reform/

Glynn, S. M., & Koballa, T. R. (2006). Motivation to learn in college science (pp. 25-32). In J. J. Mintzes & W. H. Leonard (Eds.), Handbook of college science teaching. Arlington, VA: NSTA Press.

Gonzales, P., Williams, T., Jocelyn, L., Roey, S., Kastberg, D., and Brenwald, S. (2008). *Highlights From TIMSS 2007: Mathematics and Science Achievement of U.S. Fourth- and Eighth-Grade Students in an International Context* (NCES 2009–001).National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Hattie, J. (2009). Visible learning: A synthesis of over 800 meta-analyses relating to achievement. London: Routledge.

Marx, R. W., Harris, C. J. (2006). No child left behind and science education: Opportunities, challenges, and risks. *The Elementary School Journal*, 106, pp. 467-477.

Rabe-Hesketh, S. and Skrondal, A. (2008). Classical latent variable models for medical research. *Statistical Methods in Medical Research* 17, 5-32.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods. 2nd edition*. Newbury Park, CA: Sage.

Sechrest, L. B., West, S. G., Phillips, M. A., Redner, R., & Yeaton, W. (1979). Some neglected problems in evaluation research: Strength and integrity of treatments. In L. B. Sechrest, S. G. West, M. A. Phillips, R. Redner, & W. Yeaton (Eds.), *Evaluation studies review annual*, (Vol. 4, pp. 15-35). Beverly Hills, CA: Sage.

Shadish, W., Cook, T. & Campbell, D (2002). *Experimental & quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Snyder, T.D., and Dillow, S.A. (2010). *Digest of Education Statistics 2009* (NCES 2010-013). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Wiggins, G., & McTighe, J. (1998). *Understanding by design*. Association for Supervision and Curriculum Development: Alexandria, VA.

Appendix A. Mother and Father Education Levels for Student Participants

	Mother's Education Level (*N* = 1319)	Father's Education Level (*N* = 1166)
Not a High School Graduate	6%	5%
High School Graduate	17%	17%
Some College	18%	13%
Bachelor's Degree	13%	10%
Master's Degree	10%	9%
Doctoral/ Professional Degree	3%	4%
Don't Know/ Not Stated	19%	25%

Appendix B. Training Description

Teacher training was comprised of two distinct sections: research study orientation and product training. All participating sites participated in training at their own school sites prior to the start of study participation. Most training sessions occurred in August 2009, while a few sites were trained in early September 2009.

Research Study Orientation: A representative from either the Cobblestone research team or a representative from the Pearson Academic Research team provided the study overview training to all participating treatment and control teachers and study liaisons. The research study orientation included a review of study activities, including timelines and procedures for pre/post testing and shipping back testing materials. The orientation also included collecting specific teacher information such as contact information, demographic information and signed teacher consent forms. Most study orientation sessions were held prior to the product training sessions so all teachers could be present, but then control teachers could be excused while treatment teachers attended the product overview sessions.

Product training: A Pearson representative (most with prior expertise in teaching Biology or another type of high school science) conducted the product overview training for approximately 6 hours during the first session. Trainers were also previously trained on how to conduct teacher training in July 2009 so training sessions would be consistent across study sites. Trainers used a power point presentation to review the program components and also demonstrate online features of the program. All trainers were familiar with product components and referred to the study implementation guidelines (see Appendix E) to ensure that teachers were aware of the most critical components of the program to implement during the study. A follow up training was held with all study sites in which trainers visited individual schools a few weeks after the school year began to reinforce usage of program components and to identify any problems that teachers were having using the new program. During follow up sessions trainers also reviewed additional online components and signed up individual students on the online system. Trainers also provided their individual contact information for teachers to follow up with them directly if they had any questions about the program or specific components.

Appendix C. Teacher Interview Protocols

Treatment Teacher Interview Protocol

1. Comment generally on your view of how your year has gone being part of the study and using the *Miller & Levine Biology* program.
2. *Differentiated instruction:* One definition of differentiated instruction is "A flexible approach to teaching in which the teacher responds to student differences in reading level, interests, and learning needs." Given this definition, how well do you feel you were able to differentiate instruction for your students this year using the *Miller & Levine Biology* materials? (e.g., prompts in the margin of teacher edition text)
3. *Understanding by Design:* Please provide specific information about what you thought of the following program components. Please be as specific as possible. To what extent did each of the following engage students, motivate students, and reinforce learning?
   a. *The Big Question*
   b. *Chapter Mystery*
   c. *Understanding by Design* strategies
4. *Digital Path:* How much of the Digital Path were you able to use in the classroom? Did students use it at home? What were the strong and weak points of the online services?
5. *Product use: For the following question, please indicate what version, Mainstream or Foundation, you used during the year. If you used both versions then please comment on the benefits and frustrations that you had when using both versions.* What components of the teacher's resources and ancillary materials did you find yourself continuously using? What were characteristics of these components that made them so appealing to use? Which parts of the program did you avoid using and why?

Be sure to have teachers comment on the following:

   a. Lab manuals (A & B)
   b. Study Workbooks (A & B)

6. What was the strongest aspect of the M&L Biology program? What was it missing?

Control Teacher Interview Protocol

1. Comment generally on your view of how your year has gone being part of the study and using your current Biology program. (Verify the control program)
2. *Differentiated Instruction:* One definition of differentiated instruction is "A flexible approach to teaching in which the teacher responds to student differences in reading level, interests, and learning needs." Given this definition, how well do you feel you were able to differentiate instruction for your students using the Biology program? Do you feel your program was able to meet the learning needs of all of your students?
3. *S tudents' Understanding:* Please describe the elements of the curriculum that best contributed to students' understanding of Biology. Are there particular things that you did as a teacher (assignments, activities, etc.) that you think helped contribute to their understanding?
4. *Control Textbook:* Please describe the best and worst features of the Biology program that you used this year.
5. Is there anything else you would like us to know about your classes this year?

Appendix D. Factors Derived from the Factor Analysis on Student Posttest Survey

Extraction Method: Principal Component Analysis
Rotation Method: Varimax Rotation with Kaiser Normalization
Six Factors Extracted

Factor 1: Intrinsic Motivation & Personal Relevance (Cronbach's alpha = .903)

1. I think about how the science I learn will be helpful to me (.818)
2. I think about how I will use the science I learn (.801)
3. The science I learn is relevant to my life (.723)
4. The science I learn has practical value for me (.709)
5. The science I learn relates to my personal goals (.697)
6. The science I learn is more important to me than the grade I receive (.569)
7. I find learning science interesting (.555)
8. I enjoy learning science (.529)
9. Understanding science gives me a sense of accomplishment (.465)
10. I like science that challenges me (.453)

Factor 2: Self Efficacy and Assessment Anxiety (Cronbach's alpha = .790)

1. I am nervous about how I will do on science tests (reversed, .820)
2. I worry about failing science tests (reversed, .811)
3. I become anxious when it is time to take a science test (reversed, .653)
4. I am concerned that the other students are betters in science (reversed, .642)
5. I hate taking science tests (reversed, .475)
6. I am confident I will do well on science tests (.432)
7. I believe I can master the knowledge and skills in the science course (.248)
8. I believe I can earn a grade of "A" in science class (.205)
9. I am confident I will do well on science labs and projects (.156)

Factor 3: Self Determination (Cronbach's alpha = .743)

1. I use strategies that ensure I learn science well (.669)
2. I prepare well for science tests and labs (.655)
3. I put enough effort into learning science (.645)
4. If I am having trouble learning science, I try to figure out why (.537)

Factor 4: Career Motivation (Cronbach's alpha = .845)

1. I think about how learning science can help my career (.827)
2. I think about how learning science can help me get a good job (.804)

Factor 5: Grade Motivation (Cronbach's alpha = .709)

1. I expected to do as well as or better than other students in science class (.760)
2. I like to do better than the other students on the science tests (.693)
3. Earning a good science grade is important to me (.615)
4. I think about how my science grade will affect my overall grade point average (.408)

Factor 6: Perception of Teacher Efficacy (Cronbach's alpha = .851)

1. My teacher inspires me to do my best in science (.893)
2. My science teacher explains concepts clearly (.875)
3. My teacher makes learning about science fun (.842)
4. My teacher expects me to do well in my science class (.709)

Appendix E. Efficacy Study Implementation Guidelines

Each teacher using *Miller & Levine Biology* (2010) program should have the minimum in place for program implementation:

Required:

- Exclusive use of the *Miller & Levine Biology* (2010) text for reading assignments, either the Mainstream or Foundation path (appropriate to level of students)
- Cover most chapters in first 5 Units of text
- Use of Study Workbooks A and B on numerous lessons
- Cover all "Key Questions" in chapter that apply to state standards
- Complete all end of Chapter assessments for those applicable to state standards
- Conduct at least three labs from each unit (i.e., *Quick Lab*, Design Your Own Lab, Real World Lab, Skills Lab)
- Work through *Chapter Mystery* (each chapter), including clues throughout
- Introduce chapter using "*Big Idea*" and revisit "Explore the *Big Idea*" throughout
- Refer to "*Understanding by Design*" ideas in Teacher's Edition (TE, Mainstream only)
- Read "UbD in *Miller & Levine Biology*" on page T8
- Students should review Visual Analogies and Build Connections, when applicable

Strongly recommended:

- Complete all Lesson assessments for those sections applicable to state standards
- Cover some *Caption Questions* throughout the chapters
- Go to Biology.com for other resources (e.g., Tutor Time, Art in Motion, etc.)
- Students use of Study Guide at end of chapter
- Use of Lab Manual A/B for labs
- Use of "*Untamed Science*" videos
- Students complete end of section assessments (mainstream)/Check Understanding (Foundation)

For those using the Mainstream text only:

- Have students complete at least one "In Your Notebook" assignment per unit (SE)
- Students complete at least one "Taking Notes" activity per unit (SE)

Optional:

- Any other component of the program, such as: Perform "Analyzing Data" exercises; Standardized test prep; "Build Vocabulary"; Other *Miller & Levine Biology* technology; Test Prep pages; Biology & Technology, etc.

Appendix F. School Site Observation Summaries

Indiana High School- Site 1. During both spring and fall observations, a total of six teachers were observed, four treatment and two control, in eight classroom sections. Three treatment teachers were utilizing the Mainstream textbook, and one was using the Foundation version. Teachers used the textbook in a variety of ways to provide a foundation of classroom learning. In treatment classrooms, researchers observed students studying vocabulary in Student Workbooks, discussing in the "*Big Idea*" and "*Chapter Mystery*", watching an "*Untamed Science*" video, participating in teacher lectures from PowerPoint slides, engaging in a chapter assessment, online work at biology.com, a quick lab ("Classroom Variation" p. 311). Non-*Miller & Levine Biology* (2010) materials used in these classrooms were labs adapted from other sources and teacher-created worksheets and PowerPoint slides. Observers in control classrooms reported that students completed worksheets and labs from control curriculum program, participated in lectures from PowerPoint slides and read sections of the textbook aloud in class.

North Carolina High School – Site 2. Researchers had the opportunity to observe five teachers at this site over seven class periods, two control teachers and three teachers implementing the *Miller & Levine Biology* curriculum. Online components were popular in treatment classrooms at this site, including videos from biology.com, and other web components such as Tutor Tubes, Art in Motion, and Visual Analogies. Other activities observed in treatment classrooms included watching an "*Untamed Science*" video, activities in Student Workbooks, referencing Mainstream textbooks, and "Analyzing Data" exercises, as well as non-*Miller & Levine Biology* materials, such as a short quiz and teacher-created worksheets aimed at specific state standards. One teacher expressed concern for meeting state standards and reports creating their own exercises to better teach required material. Control classrooms did reading from their textbook, independent student work on worksheets or handouts from workbooks, and took notes during a lecture from overhead transparencies.

Oklahoma High School – Site 3. In nine separate class sections, the research team observed three teachers at this site combined during both fall and spring observations. The treatment teachers at this site had a unique experience with the textbook in their classrooms. While one teacher used the mainstream textbook throughout the entire year, the other treatment teacher switched from Foundation to Mainstream at the end of the first semester. During the observations, students commented that the Mainstream text contained too much information and too many details. They seemed to prefer the Foundation text because it led to increased understanding (not as much information). During observations, teachers used the student edition textbooks to provide relevant information, present "Key Questions", and to prompt activities for small-group work. Treatment teacher also utilized online elements, including "*Chapter Mystery*" videos and Tutor Tubes, as well as *ExamView* software to review for upcoming exams. The teacher and students utilizing control curriculum programs did not use student edition textbooks during observations. Students in these classrooms completed crossword puzzle activities, discussed vocabulary words presented on the overhead projector, and completed a lab from previous class periods by creating their own graphs of the data.

Oregon High School – Site 4. Three teachers, two treatment and one control, were observed from this school site over seven class periods. Research team members observing this school site reported that teachers were thoroughly implementing the *Miller & Levine Biology* (2010) program. Treatment teachers used only the Mainstream version of the Student edition textbook and utilized many of the supplemental materials such as elements included on biology.com (Art Review, Visual Analogy), "*Chapter Mystery*", "*Big Idea*", PowerPoint slides, study workbook, *Quick Lab* (20.1, pg. 305 in Lab Manual), and chapter assessments. Control Teachers engaged students with vocabulary study activities, worksheets, and lab experiments.

Oregon High School – Site 5. At this high school site, observations took place in seven classroom sections, including two treatment and two control teachers. Like other sites, this site had one teacher using both student edition textbook versions, although not in the same class period, and one teacher utilizing solely the Mainstream version. Teachers implementing the *Miller & Levine Biology* (2010) program utilized a variety of elements from the program in their classrooms, including "Chapter Mysteries", *Quick Labs*, a Visual Quiz, PowerPoint slides, "*Big Idea*", "Key Questions", labs and demonstrations. The control teacher engaged students in vocabulary from the student workbook, independent exam review, and a lab exploration of a nearby meadow. Control classrooms made little use of their textbook.

Washington High School – Site 6. In eight separate sections, biology classrooms were observed for three teachers, two treatment and one control teachers. One teacher at this site was using both Mainstream and Foundation texts, however not in the same class periods. Observers in classrooms using the *Miller & Levine Biology* (2010) program reported students were working independently on activities in the Lab Manual and creating study guides for upcoming exams. Teachers presented lectures and "Chapter Mysteries" to students via PowerPoint slides. In control classrooms, students took notes on lectures from overheard transparencies and completed worksheets and diagram-labeling activities.

Appendix G. Random-Intercept Models with Covariates

To estimate the program effect, we ran a series of parallel random-intercept models with covariates using STATA, which falls under two-level linear models in our case since we have students nested within classes (Rabe-Hesketh & Skrondal, 2008). All HLM models were ran in STATA (--xtmixed procedure).

A general linear random-intercept model with covariates can be represented as follows:

$$y_{ij} = \beta_1 + \beta_2 x_{2ij} + \ldots + \beta_p x_{pij} + \zeta_j + \varepsilon_i$$
$$= (\beta_1 + \zeta_j) + \beta_2 x_{2ij} + \ldots + \beta_p x_{pij} + \varepsilon_i$$

In the above model, y_{ij} refers to the outcome of student i in cluster (or class) j; x's refer to various students, teacher/class variables (i.e., covariates). The random intercept term (i.e., ζ_j) signals the linear model is of multilevel (two-level in our study) rather than simple OLS (ordinary least square) regression. We ran parallel models for each of the outcomes.

The following is a list of variables and their operational definitions associated with student background characteristics and teacher/ classroom/ school characteristics that were used in the HLM models.

Outcome variables:

1. SAT/9 biology scaled scores posttest
2. Biology Assessment scores posttest

Student background characteristics variables:

1. Proxies for prior academic achievement
   - prior SAT/9 scaled scores pretest
   - Biology assessment scores pretest
2. Gender: female (0) and male (1)
3. Race indicators
   - African American (1)
   - Latino (1)
   - Other ethnicity (1)
   - White (reference group)
4. Primary language indicator
   - Not English (1)
   - English (reference group)
5. Mother's education: ordinal scale ranging from "less than high school" (1) to "doctoral/professional" (6).
6. Grade level indicators
   - 10th (1)
   - Junior or senior, i.e., 11th or 12th (1)
   - 9th (reference group)
7. Pre-survey composite affective variables:
   a. Student motivation
   b. Self-efficacy
   c. Student motivation for grades

Teacher/classroom/school characteristics variables:

1. Condition
   - Treatment-low (1)
   - Treatment-medium (1)
   - Treatment-high (1)
   - Control (reference group)
2. Teacher rapport with students
3. Teacher classroom management
4. Years of teaching experience (Number of years teaching)

Appendix H. Summary of the Impact of Student and Teacher Characteristics on Student Scores

Table a shows the possible impact of student and teacher characteristics on the predicted posttest SAT9 score. The "coefficient" for each variable or characteristic (e.g., SAT9 pretest, Male, Rapport) is used to determine the impact of the variable to predict an individual student's score. The "low value" and "high value" are the possible range of values that are available for each variable. They are multiplied with the coefficient and ultimately added to the intercept value (i.e., 179.77) to calculate the predicted score. For example, the response range for self-efficacy on the student survey is 1 to 5. Specifically, a score of 1 on self-efficacy would predict an additional 3.76 points added to the SAT9 score while a score of 5 would predict an additional 18.80 points added. The characteristics of each student could then be used to predict a student's SAT9 posttest scaled score.

Table a. SAT9 HLM Model: Impact of Student and Teacher Characteristics on Predicted Posttest Score

Fixed Effect	Coefficient	Lowest Possible Value	Highest Possible Value	Coefficient x Low Value	Coefficient x High Value	Difference
SAT9 Pretest* (scaled score)	0.65	541	843	351.65	547.95	196.3
Rapport*	6.69	1	5	6.69	33.45	26.76
Teacher classroom management*	4.28	1	5	4.28	21.40	17.12
Student self-efficacy*	3.76	1	5	3.76	18.80	15.04
Teacher years of teaching experience*	0.44	1	34	0.44	14.96	14.52
African American*	-13.85	0	1	0	-13.85	-13.85
English is not primary language*	11.41	0	1	0	11.41	11.41
Treatment Implementation-high*	7.89	0	1	0	7.89	7.89
Latino*	-7.45	0	1	0	-7.45	-7.45
Mother's education*	1.34	1	6	1.34	8.04	6.70
Treatment Implementation-low	4.85	0	1	0	4.85	4.85
Grade 10 indicator	-3.19	0	1	0	-3.19	-3.19
Male	2.6	0	1	0	2.60	2.60
Other ethnicity	-2.58	0	1	0	-2.58	-2.58
Junior and senior indicator	1.43	0	1	0	1.43	1.43
Student motivation	0.34	1	5	0.34	1.70	1.36
Treatment Implementation-medium	-1.24	0	1	0	-1.24	-1.24
Student motivation for grades	0.28	1	5	0.28	1.40	1.12
Intercept*	179.77	1	1	179.77	179.77	NA

From Table a, it is clear that the biggest predictor of the posttest SAT9 scaled score is the pretest score. This is followed by teacher/classroom variables (e.g., teacher rapport with students, classroom management) and student attitudes (e.g., self-efficacy) and several other student characteristics (e.g., study condition, ethnicity).

It is important to note that some variables have a range of possible scores or responses (e.g., student self-efficacy, Rapport, Classroom Management) while some variable are either "no" or "yes" (i.e., 0 or 1, respectively). The *no* or *yes* variables include ethnicity, gender, primary spoken language, grade, and treatment effect. Each of these groups includes a reference group whereby the comparison is made for the overall group (see Appendix G for all reference groups). For example, the reference group for primary spoken language is *English.* This indicates that the variable *English is not primary language* (i.e., ESL student) is compared to students that speak English as their primary language. According to the model in Table a, an ESL student would be predicted to score 11.41 points more than an English speaking student *all other characteristics being equal.* This finding may be confusing considering ESL students had lower overall scores than English speaking students. However, two key pieces of information need to be considered to understand the overall finding. First, the pretest scores, which were the largest predictor of posttest scores, for ESL students were much lower than English speaking students. This explains the overall lower

Table b. Biology Core Content Assessment HLM Model: Impact of Student and Teacher Characteristics on Predicted Posttest Score

Fixed Effect	Coefficient	Low Value	High Value	Coefficient x Low Value	Coefficient x High Value	Difference
SAT9 Pretest*	0.68	0	30	0	20.40	20.40
Rapport*	1.36	1	5	1.36	6.80	5.44
Teacher classroom management*	1.11	1	5	1.11	5.55	4.44
Student self-efficacy*	0.96	1	5	0.96	4.80	3.84
Student motivation*	0.65	1	5	0.65	3.25	2.60
Student motivation for grades*	0.5	1	5	0.5	2.50	2.00
African American*	-1.65	0	1	0	-1.65	-1.65
Teacher years of teaching experience	0.05	1	34	0.05	1.70	1.65
Treatment Implementation-medium	-1.49	0	1	0	-1.49	-1.49
Latino*	1.3	0	1	0	1.30	1.30
Junior and senior indicator	-1.28	0	1	0	-1.28	-1.28
Grade 10 indicator*	-0.83	0	1	0	-0.83	-0.83
Male	-0.5	0	1	0	-0.50	-0.50
Other ethnicity	-0.47	0	1	0	-0.47	-0.47
Treatment Implementation-low	-0.45	0	1	0	-0.45	-0.45
Mother's education	0.07	1	6	0.07	0.42	0.35
English is not primary language	0.24	0	1	0	0.24	0.24
Treatment Implementation-high	0.02	0	1	0	0.02	0.02
Intercept*	-8.26	1	1	-8.26	-8.26	NA

scores. Second, the change from pretest to posttest scores for the ESL students was larger than English speakers (ESL = 20.52, English = 10.42). Since this difference could only best be explained by primary spoken language (i.e., no other student or teacher characteristics could explain the difference better), the results show that ESL students outperformed English speaking students on the SAT9 scaled score after controlling for all other characteristics. This was the overall result of primary spoken language without looking at differences between the control and treatment groups; however, the results for the treatment and control groups are consistent to the overall findings. This finding may be the result of ESL students becoming better English speakers by the end of the school. In other words, the low performance on the pretest may have been a function of not being able to understand the language on the SAT9 more than not understanding the scientific principles on the SAT9.

Table b shows the impact of student and teacher characteristics on the predicted posttest Biology Core Content Assessment. Again, the table shows that the pretest was the most important predictor of the posttest score followed by the teacher/classroom characteristics, student attitudes, and other student characteristics. However, this model shows a different finding for Latino students than Latino students on the SAT9 such that Latino students outperform the Caucasian students after controlling for all other variables on the Biology Core Content Assessment. This finding is explained similarly to the ESL finding on the SAT9. That is, Latino students showed a much larger increase from pretest to posttest than Caucasian students. However, Caucasian students still showed larger overall test scores on the posttest than Latino students. Again, this appeared to be consistent for both control and treatment students who participated in the study.

Appendix I. Attrition and Differential Attrition Analyses

Sample attrition is defined as those students who completed pretests on any of the primary outcome measures (i.e., Biology Assessment, SAT9, and a Student Survey), yet did not complete a posttest on any of these measures. There were 1,537 participating students that completed at least one pretest and posttest. Of the 1,537 students, 1,200 completed all pretests and posttest. The original pretest sample (students that completed at least one pretest measure) included 1,859 students.

The difference between the original sample of students that took any or all of the pretests and the final number of students is 322. Of the 322 students, 169 were not included on the rosters for the second semester and most likely moved during the first semester. The remaining 153 students were either absent during posttesting or had moved during the second semester.

An overall summary of the attrition data is provided in Table a. This table shows that there were no noteworthy differences when comparing the treatment and control groups in the percentage of students that did not complete a posttest after completing a pretest assessment. Overall, the treatment and control groups had a close percentage (17% vs. 18%, respectively) of students that did not complete at least one posttest after completing a pretest. After accounting for attrition, there were 889 participating students (83% of treatment students with a pretest) in treatment classrooms and 648 participating students (82% of control students with a pretest) in control classrooms who had completed at least one pretest and posttest.

All further analyses focus only on the combination of all assessments (i.e., we have not reported individual results on the Biology Core Content Assessment, SAT9, and the Student Survey). We found that these data are sufficiently represented when looking at all assessments together. To examine whether sample attrition created differences between the treatment and control groups, we compared the original sample (those students who had completed at least one pretest) to the final sample of students (those students who had completed both a pretest and posttest) across student demographic characteristics. Table b shows the key demographics of our original sample and those students from the original sample that completed a posttest. The table shows that there are no major differences between the original sample and the final sample of students: gender is split evenly between both samples, a majority of the students are Caucasian (59.7% and 61.8%), and most speak English (85.8% and 87.9%).

Differential Attrition

Almost any experimental study has participant attrition, particularly in applied research settings (i.e., schools) where students may leave before the year is over due to circumstances outside of the control of the school, teacher, or researchers. What is important to determine, however, is whether there was differential attrition such that students in one group (treatment or control) were more likely to exit the study in comparison to the other group before completing posttest measures; two sets of analyses were conducted to test this. The first set of analyses used demographic characteristics to examine the extent to which students that completed both a pretest and posttest differ from students that completed only a pretest. The second set of analyses

Table a. Total Students with a Pretest Assessment: Students with Pretest and Posttest vs. Students Missing Posttest Only

Assessment	Condition	Students with Complete Pretest and Posttest	Students Missing Posttest Only (Attrition)	Total for Each Assessment
Biology Core Content Assessment	Treatment	826 (82%)	185 (18%)	1,725
	Control	577 (81%)	137 (19%)	
SAT9	Treatment	770 (77%)	230 (23%)	1,722
	Control	593 (82%)	129 (18%)	
Student Survey	Treatment	824 (82%)	183 (18%)	1,732
	Control	585 (81%)	140 (19%)	
At Least One Assessment Complete	Treatment	889 (83%)	183 (17%)	1,859
	Control	648 (82%)	139 (18%)	

Table b. Key Demographic Characteristics: Original Pretest Sample vs. Sample with Posttest

Demographic Characteristics (Percent %)		Original Pretest Sample *n* = 1,859	Sample with at least one Pretest and Posttest *n* = 1,537
Gender	Male	50.0	49.7
	Female	49.5	49.9
Ethnicity	Caucasian	59.7	61.8
	Latino	8.3	8.5
	Multi-ethnic/Other	10.8	10.2
	African-American	9.5	9.3
	Asian	2.0	2.3
Primary Language	English	85.8	87.9
	Other	4.8	4.9
	Unknown	9.4	7.2

** Note that the numbers provided within each group do not add up to 100% due to missing data.*

sought to determine if of the students who dropped out of the study, the treatment and control students differ in their achievement scores on the Biology Assessment and SAT9 assessments. The second set of analyses was discussed in Section Five under the Attrition section.

Based on our sample attrition analysis, there were 322 students who took at least one pretest assessment but did not take a posttest. Table c explores the demographic characteristics of these students to see whether there was any systematic differential attrition between the groups.

Table c shows that the students in the attrition group generally corresponded to the same group of the students that completed the study. For example, there were slightly more females in the treatment group that had a pretest and posttest than males (51.2% and 48.8%, respectively). This was also true for the students that dropped from the study in the treatment group (51.9% and 48.1%, respectively). Overall, Table c shows that there were not any specifics groups that displayed a surprising withdrawal of students.

Table c. Students with Complete Pretest and Posttest vs. Complete Pretest Only

Demographic Characteristics (Percent %)		Pretest and Posttest Complete *n* = 1,537		Complete Pretest Only *n* = 322	
		Treatment *n* = 889	Control *n* = 648	Treatment *n* = 183	Control *n* = 139
Gender	Male	48.8	51.5	48.1	56.5
	Female	51.2	48.5	51.9	43.5
Ethnicity	Caucasian	65.9	66.9	60.1	59.6
	Latino	11.1	6.2	8.5	10.5
	Multi-ethnic/Other	11.0	13.6	16.3	19.3
	African-American	9.4	10.7	13.7	10.5
	Asian	2.5	2.5	1.3	-
Primary Language	English	93.6	96.3	93.3	95.4
	Other	6.4	3.7	6.7	4.6

Appendix J. Mean Student Ratings of Curriculum

Textbook/ Program Component	Foundation Students Mean Rating (SD)	Mainstream Students Mean Rating (SD)	Overall Students Mean Ratings (SD)
Miller & Levine Biology Textbook (Overall)	2.89 (.74)	2.88 (.73)	2.89 (.73)
Lab Manual	2.76 (.70)	2.61 (.76)	2.63 (.75)
Study Workbook A or B	2.68 (.84)	2.44 (.88)	2.47 (.87)
The "*Big Idea*"	2.76 (.70)	2.62 (.84)	2.64 (.82)
Chapter Mystery	2.75 (.78)	2.59 (.95)	2.62 (.93)
Key Questions (blue keys)	2.86 (.71)	2.83 (.91)	2.83 (.88)
"Build Vocabulary" (highlighted words in text)	3.05 (.73)	3.02 (.83)	3.02 (.82)
Visual Analogy (e.g. the cell is a factory)	3.00 (.68)	2.95 (.84)	2.96 (.81)
Experiments & Labs	3.01 (.81)	2.81 (.84)	2.84 (.84)
"Analyzing Data" Exercises	2.66 (.76)	2.37 (.81)	2.41 (.81)
Pictures of scientific concepts	3.07 (.65)	3.00 (.75)	3.01 (.74)
Diagrams of scientific concepts	2.90 (.71)	2.93 (.77)	2.93 (.76)
Visual Guide to The Diversity of Life (colorful section in the back of the textbook)	3.03 (.75)	2.93 (.87)	2.94 (.84)
millerandlevine.com	2.71 (.72)	2.47 (.96)	2.51 (.92)
Biology.com	2.83 (.73)	2.83 (.97)	2.82 (.93)
"*Untamed Science*" videos	2.95 (.86)	2.79 (.97)	2.82 (.95)

** Note. Component ratings were provided on a scale of one to four (One = Strongly Dislike, Four = Strongly Like).*

Appendix K. Product Satisfaction Summary of the *Miller & Levine Biology* (2010) Foundation Series Textbook

The *Miller & Levine Biology* (2010) Foundation series was used in two studies: the pilot study (spring 2009) and the efficacy study (2009-10 school year). Product use and satisfaction data were collected from teacher interviews and focus groups, and student survey results, student focus groups (pilot study only) and classroom observations. The following is a summary of major findings from both studies.

Implementation of Foundation Elements

Table a provides a complete breakdown of the elements utilized by teachers in the *Miller & Levine Biology* (2010) pilot study and the efficacy study. These data were reported by teachers during weekly logs online. For the pilot study, Foundation teachers indicated that their least utilized features of the program were the Building Understanding components, Chapter Summary, and Standardized Test Prep.

Further information regarding use of the Foundation series in classrooms was drawn from classroom observations. Cobblestone researchers found that there were obvious differences between those classrooms using Mainstream versions of the text versus Foundation version of the text. Although most teachers had extremely good classroom management skills, Foundation students were more prone to off-task behaviors and slower rates of work completion—this is expected given that these students are lower academically. Teachers generally targeted instruction to the level of students that corresponded to

Table a. Percent of Possible Components of Foundations Textbook used by teacher in the Biology Pilot Study

Chapter Components		Overall Use
Chapter Concepts	**Chapter "*Big Idea*" (4)**	75%
	Chapter Assessment (4)	56%
	***Chapter Mystery* (4)**	63%
	Standardized Test Prep (4)	19%
	Chapter Summary (4)	19%
	Solve the *Chapter Mystery* (4)	50%
Section Components	**Key Questions (14)**	59%
	Build Understanding (14)	14%
	Build Vocabulary (14)	52%
	Build Connections (6)	50%
	Check Understanding (14)	50%
	Inquiry Into Scientific Thinking (4)	31%
Lab	Total # Labs Completed	1

the class level. Most teachers using Foundation materials with their students appropriately reinforced vocabulary more directly for these students; only one teacher using the Foundation materials appeared to hold extremely low expectations of students' work ability, and consequently students' behavior reflected this general attitude.

For those classrooms in which teachers used both levels of the text, teachers were generally skillful at combining both levels for students; however, logistically this appeared much more difficult than using either version alone. For teachers to use both texts simultaneously, it was obvious that this required teachers to be extremely organized in their approach to the lesson. The lack of complete crossover from the two versions of the books also made this difficult, for example, when a lab was missing from the Foundation book students had to share resources to complete the assignment.

Production Satisfaction

Teacher Reactions

Teachers using Foundation series during both the pilot study and efficacy study had similar responses regarding the reading level of the text. Most teachers who used the Foundation version of text with their students reported really liking the reading level and some stated that it kept lower level students engaged. However, some teachers still felt the reading level was still too high. One teacher commented that "for the classes that are using it, some of the reading is above them." Teachers also reported that despite the decreased amount of visual displays and text in the Foundation book, *the current amount of pictures, colors and text was too distracting for students* of a lower ability level. Teacher felt that some students had difficulty with the testing materials, specifically the chapter assessments in the Foundation text.

> *"Students got all the information they need [from the Foundations book] …and it's not overwhelming"*
>
> **–teacher using Foundations book**

Another teacher thought the Foundations text lacked content for state standards guidelines, resulting in difficulty in standardized testing for students using the Foundation series. The main complaint from teachers utilizing both Mainstream and Foundation versions was that they disliked how the elements of each were labeled differently (i.e. "Check Your Understanding"). This issue was especially poignant for teachers implementing both textbooks in the course of their day.

Vocabulary sections in the Foundations text were well-liked by students, according to teacher interviews. Teachers also felt their students enjoyed the "Speed Bumps" sections.

> *"The way things are named in the 'Foundations' book really has bothered me because it's different."*
>
> **–Treatment teacher implementing both Mainstream & Foundations**

Student Reactions

Student input on the Foundation textbook was provided during focus group interviews during the Biology pilot study and from the student survey from the Biology efficacy study.

> *"My kids love the 'speed bumps' in [the Foundations] book."*
>
> **–Treatment Teacher**

Student survey results gathered during the Biology efficacy study suggest that students using the Foundation version of the text rated their material more enjoyable than those using the Mainstream version. Foundation students reported highest ratings for pictures of scientific concepts (Mean = 3.07 out of four), "Building Vocabulary" sections (Mean = 3.05 out of four), and the Visual Guide to the Diversity of Life (Mean = 3.03 out of four).

Student Feedback from Focus Groups:

- A favorite feature in the Foundations book was **Vocabulary Builder**. Students also reported that less written information was good, and it "gets to the point"
- Some students using the Foundations version also noticed that they was less text in the book, and were happy that there was less information to read about.
- One student who used the Foundations text compared both books side by side and remarked "when looking at the mainstream book…it would kind of discourage me because it's so much [text]" and preferred that the Foundations version was "short and to the point"

Students' least favorite elements were Analyzing Data exercises (Mean = 2.66 out of four), millerandlevine.com (Mean = 2.71 out of four), and Chapter Mysteries (Mean = 2.75 out of four). Appendix J provides complete data for both Mainstream and Foundation students participating in the efficacy study.

Additional teacher comments regarding the Foundation series book:

- "It wasn't overloaded with pictures, you know sometimes the book is too busy. I did not see or feel the book was too busy. I thought the reading level was the closest to appropriate that I've had."
- Most students did "okay" using the Foundation level workbook (Workbook B). One teacher commented that some students did not like some workbook sections. "…they [students] did not like the vocabulary section where the words were defined and they had to say how they were going to remember it. They did not like to fill in those sections. And I said 'Well, what's in a definition?' you know, I would try to explain to them how to use that, that it would help you to recall the definition of something familiar, like in the word itself—they hated that section of the workbook."
- "…overall, I think they [students] enjoyed the book… the diagrams are beautifully displayed, easy to read, easy to interpret, students commented on that."
- "…the lab book [Lab Manual B]—every time we tried to do a lab in the lab book…they got lost in it. It was just too difficult."

Overall, students enjoyed using the Foundation text, particularly that there was less text than a typical textbook, and they specifically mentioned the pictures and drawings in the book. When compared side by side, students using the Foundation book preferred the look and layout, as it appeared less intimating to read in comparison to the Mainstream book. Although there were far fewer students who used the Foundation series as compared with students using the Mainstream *Miller & Levine Biology* book during the efficacy study, those using the Foundation book reported liking the program more than their Mainstream peers. Teachers reported liking the Foundation text as well, although some still expressed that the reading level in the book and in some ancillary materials was still too high for their lowest level students.

Appendix L. SAT9 Percentile Scores By Grade Level

Table a. displays the SAT9 Percentile Scores by grade level. These pretest and posttest scores were combined to provide an overall average percentile score that was used in the body of this report. While ninth and tenth grade students comprised of the majority of students in the study, there were some students that were removed from the percentile scores because we were did not have their grade information (n = 92). We also did not include students in eleventh grade (n = 33) and twelfth grade (n = 4) because reporting the averages with such small numbers would be inappropriate.

Table a. SAT9 Percentile Scores by Grade Level

Group	9th Grade Pretest Percentile	9th Grade Posttest Percentile	10th Grade Pretest Percentile	10th Grade Posttest Percentile	Weighted Average Pretest Percentile	Weighted Average Posttest Percentile
Control (9th Grade n = 262; 10th Grade n = 251)	56	57	57	63	56.5	59.9
Treatment- Overall (9th Grade n = 219; 10th Grade n = 450)	67	75	58	68	60.9	70.3
Treatment- Low (9th Grade n = 89; 10th Grade n = 187)	72	80	54	65	59.8	69.8
Treatment- Med (9th Grade n = 100; 10th Grade n = 141)	65	71	65	75	65.0	73.3
Treatment- High (9th Grade n = 30; 10th Grade n = 122)	62	78	54	63	55.6	66.0

Notes

2120 Foothill Blvd., Ste. 202 | La Verne, CA 91750 | (800) 971-3891