A RANDOMIZED CONTROL TRIAL TO TEST THE EFFECTS OF PRENTICE HALL MILLER AND LEVINE (2006) *BIOLOGY* CURRICULUM ON STUDENT PERFORMANCE: FINAL REPORT

Submitted by:

CLAREMONT GRADUATE UNIVERSITY
Institute of Organizational and Program Evaluation Research

January 17, 2007

Rebecca M. Eddy, Ph.D. & Tiffany Berry, Ph.D.
Claremont Graduate University

With Special Acknowledgement to:

John LaVelle, M.S.A.P, Claremont Graduate University
Kile Dyer, M.A., Claremont Graduate University
Xiaoxia Newton, Ph.D., Stanford University

123 East Eighth Street
Claremont, CA 91711
(909) 607-1549
(909) 621-8905 fax

TABLE OF CONTENTS

TABLES

ACKNOWLEDGEMENTS

We would like to thank all of the participating teachers and students who generously provided their time and effort to participating in this study. We appreciate you allowing us into your classrooms to observe science in action. We would also like to give a special thanks to the team from Claremont Graduate University including Laura Berger and our research associates, Kile Dyer and John LaVelle, who have worked tirelessly on many aspects of this project. We would also like to thank Dr. Xiaoxia Newton for her expertise in advanced statistics and, in particular, Hierarchical Models. We also sincerely appreciate everyone from Pearson Education including Brent McKenzie, Julia Fellows Osborne, Michelle York, and especially Marcy Baughman. Marcy, you have made this experience pleasurable, engaging, and fun! Finally, we would like to thank our husbands, Richard Eddy and Eric Berry, who have patiently endured travel schedules, long meetings, and late nights.

With sincere gratitude,

Dr. Rebecca Eddy & Dr. Tiffany Berry

Institute for Organizational and Program Evaluation Research
School of Behavioral and Organizational Sciences
Claremont Graduate University
123 E. Eighth Street
Claremont, CA 91711
(909) 607-1549

ABSTRACT

Citation: Eddy, R.M., & Berry, T. (2006). A Randomized Control Trial to Test the Effects of Prentice Hall's Miller and Levine (2006) *Biology* Curriculum on Student Performance: Final Report. Claremont Graduate University, Claremont, CA. Funding provided by Pearson Education.

Background: Under No Child Left Behind (NCLB), textbook publishers are required to scientifically validate the effectiveness of their products. Formative evaluations on the Miller and Levine's (2006) *Biology* curriculum have been conducted by internal market research teams at Pearson Education, but this is the first study to scientifically test the impact of the curriculum on Biology achievement.

Purpose: The purpose of this study is to conduct a Randomized Control Trial (RCT) on Prentice Hall's Miller and Levine (2006) *Biology* curriculum. We randomly assigned teachers to a treatment group (using the new Prentice Hall curriculum) or to a control group (using existing Biology curriculum currently in place at school). In particular, we were interested in examining: (1) How student outcomes differed for students using the Prentice Hall *Biology* program compared to control students; (2) How student outcomes differed for treatment and control students as a function of their background characteristics (e.g., gender, ethnicity, etc.); (3) The relationship between program implementation and student achievement in science; and (4) Product satisfaction among treatment teachers and students.

Setting: This study took place in five high schools across four states (California, Colorado, Ohio, and New Jersey). All high schools were fairly large (between 1,700-2,500 students each), except for two small rural schools in Ohio. The study was conducted over the 2005-2006 school year.

Study Sample: Sixteen teachers (eight treatment and eight control) and 1,401 students participated in the study.

Intervention: Teachers in the treatment group were provided with textbooks and ancillary materials for themselves and their students during the entire 2005-2006 school year. Teachers were also provided with a small budget with which to purchase supplies for laboratory activities. All treatment teachers were required to adhere to the implementation guidelines developed in cooperation between the publisher and researchers. Teachers in the control group were instructed to conduct their Biology classes as they "normally" would, or as they have in "past years".

Research Design: Teachers were randomly assigned at each study site to either treatment or control groups within five different schools across the country. Multiple process measures were collected throughout the year and two outcome measures (Biology achievement and student attitudes towards science) were collected at pretest and posttest.

Control Condition: Three different Biology curricula were used in the control group across the five sites. Teachers at one site used an older version of Prentice Hall, teachers at another site used the most recent version (2005) of a competitor curriculum,

and teachers at the three remaining sites used the 2003 version of the same competitor curriculum.

Data Collection and Analysis: In combination with descriptive statistics and qualitative analyses, Hierarchical Linear Modeling (HLM) was used to address our primary research questions. Several process and outcome measures were used throughout the study in an effort to assess the effectiveness of the Prentice Hall *Biology* curriculum.

Findings: In general, treatment teachers had an adequate level of program implementation as specified by the study design; however, their classes covered significantly more sections of their textbooks in comparison to control teachers. The quality of implementation varied in both groups, but in general was similar across treatment and control groups. Based on the HLM analysis, there was no overall difference between treatment and control groups in relation to Biology achievement or students' attitudes towards science. However, Latino students in the treatment group had significantly higher Biology posttest scores than Latino students in the control group, after controlling for student demographics and previous academic characteristics.

Conclusion: In general, the Miller & Levine (2006) *Biology* curriculum was comparable to other similar Biology curricula given that students in treatment and control groups performed equally well on the Biology achievement test. However, there may be distinct advantages for specific groups of students using the Prentice Hall curriculum, particularly those from Latino backgrounds. Teachers and students provided high satisfaction ratings for Prentice Hall Biology textbooks and ancillary materials. Future research should investigate the relationship between curriculum implementation and student achievement on a larger scale, particularly with regard to differentiating instruction.

SECTION ONE: BACKGROUND AND PURPOSE

Introduction

Over the past decade, the reform of science curricula materials has been focused on two major tasks: (1) designing science curricula consistent with the state and national standards; and (2) ensuring those materials promote students' critical thinking skills. When combined, the assumption behind these reform efforts suggests that if students learn science through a standards-based curriculum and are given the opportunity to develop and exercise critical thinking skills, students will become better equipped to learn the course material and incorporate it into their existing knowledge structures. In addition, if teachers are given the opportunity and skills to be able to differentiate instruction for all levels of students in the classroom, student learning will be promoted (Tomlinson, 1999). Given that textbooks are required to be aligned with state and national standards, the key to national science reform is being able to teach students to work with "important constructs, models, and theories to develop both critical reasoning skills and deeper understanding of the processes as well as the essential content of science" (Close, 1996).

Prentice Hall is one such publisher that has shown promise in the area of national science reform. The current version of the Biology text is visibly aligned with national science standards, and the organization and conceptual structure of the text is designed to enhance understanding and, ultimately, achievement in science. In this age of accountability, and when multiple competing options for Biology texts exist, it is imperative that empirical research is conducted to examine the extent to which Miller and Levine's (2006) Prentice Hall *Biology* textbook impacts students' attitudes towards science, interest in science, and achievement in science.

Since the *No Child Left Behind Act of 2001* (NCLB), the level of evidence required to verify that a study meets the standard of scientifically-based research has been specific. The Institute of Education Sciences has declared that the strongest evidence of effectiveness comes from "randomized controlled trials that do not have problems with randomization, attrition, or disruption…" (USDOE, 2005), The goal of the current research study was to meet the highest standard of what "works" in Biology curriculum by implementing a Randomized Control Trial (RCT) by testing the effectiveness of the Prentice Hall *Biology* curriculum (Miller & Levine, 2006). Drs.

Rebecca Eddy and Tiffany Berry from Claremont Graduate University (CGU) were hired by Pearson Education/ Prentice Hall to lead this effort.

Purpose of the Prentice Hall Miller & Levine (2006) *Biology* Study

The purpose of this study was to conduct a scientifically rigorous RCT study on the effects of the Prentice Hall Miller & Levine (2006) *Biology* curriculum on student attitudes and interest in science as well as achievement in Biology. The primary research questions motivating this study were as follows:

- How do student outcomes differ for students using the Prentice Hall *Biology* curriculum compared to students not using Miller and Levine's (2006) *Biology* curriculum?
- How do students with various characteristics (e.g., gender, ethnicity, language proficiency status) using the Prentice Hall *Biology* curriculum perform on student-related outcomes?
- What is the relationship between curriculum implementation and student achievement in science?
- How satisfied are treatment teachers and students with their Prentice Hall textbook and ancillary materials?

To prepare for the current RCT, a pilot study was conducted during the 2004-2005 school year at one school in California[1]. The main purpose of the pilot study was to identify, develop, and/or refine instruments and protocols to be used during the RCT study in the 2005-06 school year. The pilot study examined teachers' implementation of the *Biology* curriculum as well as a diverse set of student outcomes. Findings from the pilot study indicated that students increased significantly from pretest to posttest on a standards-based life science assessment; however, there were no significant differences between the treatment and control groups in terms of student achievement. However, treatment students rated the textbook and ancillary materials significantly more positively than control students. Although these data provided insight into how the Prentice Hall *Biology* curriculum affects students' achievement in Biology during high school, it did not offer conclusive evidence on the impact of the curriculum on student achievement.

[1] Please refer to "The Effect of Prentice Hall Biology on Student Performance: Pilot Study, Final Report" for a complete description of the pilot study.

Description of the Prentice Hall *Biology* Curriculum

After consultation with the Prentice Hall science editorial and marketing staff, several components were identified as features that distinguished the Miller and Levine (2006) curriculum from other competing science texts (personal communication, June 30, 2004).

- Real-world applications were embedded throughout the text in such sections as "Technology and Design" and "Careers in Science."
- Hands-on activities and labs appeared within each unit and included exercises that could be completed in the classroom or on-line.
- Student assessments were sequenced to reflect progressively deeper levels of Bloom's taxonomy (1956). These assessments were embedded throughout the text, including within each section and the unit close.
- The text emphasized the importance of differentiated instruction by developing support activities and ancillary materials for English Language Learners (ELL), advanced students, etc.

In addition to these distinguishing features, the overall organizational structure of the text was developed to promote optimal student learning. For example, at the beginning of each new chapter, an "Inquiry Activity" was presented to allow teachers an opportunity to actively engage students with an interest-grabbing activity; lesson content was reinforced through use of "Checkpoints" designed to assess students' comprehension; "Key Concepts" at the beginning of each chapter allowed teachers an opportunity to address possible misconceptions prior to integration of new knowledge; and key vocabulary words were in bold print so that students could easily identify them. Ancillary materials also provided multiple ways to engage students in laboratory simulations. Leveled workbooks and lab manuals addressed varying needs of students within one classroom. The underlying logic of the curriculum was to develop an integrated program that increased students' interest, engagement, and enjoyment in the text, while simultaneously increasing their accessibility and comprehension of difficult Biology content material. When combined, these components should lead to increased knowledge of and achievement in Biology.

SECTION TWO: DESCRIPTION OF STUDY SETTING AND SAMPLE

Site Recruitment

Pearson Education contracted separately with PRES Associates to recruit schools interested in participating in the present study. PRES was charged with recruiting teachers and schools in several large urban and suburban districts with the following characteristics: (1) ethnically and socioeconomically diverse; (2) low student mobility rates; (3) willingness to randomly assign teachers to control or experimental groups; (4) at least 500 enrolled students and a minimum of four teachers with multiple sections of college-preparatory Biology or Life Science classes; and (5) using a curriculum other than the most recent version of the Miller & Levine (2006) *Biology*. PRES contacted over 1,000 schools and districts across the country; most declined to participate in the study (even with over $7,000 of free product per participating classroom). Reasons provided for not participating in the study included: (1) district-wide adoption of competitor curriculum; (2) perception that teachers would be too busy during the school year to participate in a research study; (3) lack of interest; and/or (4) state adoption cycles that prevented sites from incorporating any new textbooks/curriculum in 2005-06. Claremont Graduate University developed additional recruiting efforts as a secondary strategy when the requisite number of sites was not secured by PRES.

Of the schools (and/or districts) that met the inclusion criteria, securing their participation occurred in one of two ways and differed slightly based on where the initial contact was made (e.g., teacher, principal, or district). When the contact person was at the district-level, a study information sheet was sent to the individual (usually the Assistant Superintendent of Instruction), with a request for interested schools in the district to participate. If initial interest was secured, the researchers followed up with the teachers, and then sought permission from the principal. If the contact person was at the teacher level, the researchers first secured support from a number of teachers and subsequently sought permission from the principal and district-level personnel. All participating teachers, site liaisons, district personnel, and CGU researchers signed a Memorandum of Understanding (MOU) to formally secure the school's participation.

Through our informational materials as well as the combined efforts of PRES, CGU, and Prentice Hall, five schools across four states (California, Colorado, Ohio, and New Jersey) agreed to participate. Although the generalizability of this study was compromised given that the five schools were not randomly selected out of all eligible

schools, these sites were the only sites that were willing to participate in the research study and fully implement the curriculum during 2005-06.

School Descriptions

California Site

The California site (Site 1—CA) was a large secondary school serving approximately 2,500 students in grades 9 through 12. The school was in a suburban community located approximately 20 miles southeast of Los Angeles. The community was primarily Latino and had a median household income of approximately $50,000. The school was built in 1953 and, therefore, many classroom portables have been added to accommodate the growth in school size. Eighty-six percent of teachers held a full teaching credential; five percent held emergency credentials. On average, the student -teacher ratio was 29 to 1. The student population was representative of the community given that the school largely served Latino students (75%). Thirty-three percent of students were eligible to receive free or reduced lunch and 13% percent of students were English Language Learners (ELL's). Based on the 2004-2005 state standardized test scores, this school was considered far below average given that the number of students classified as proficient or advanced was 32% in Reading, 15% in Mathematics, and 22% in Science.

Colorado Site

The Colorado site (Site 2—CO) was a large secondary school serving over 2,000 students in grades 9 through 12. The school resided in a suburban community located approximately 30 miles north of Denver. The community was primarily Caucasian and had a median household income of approximately $52,000. The school, constructed in 1991, was a completely enclosed two-story complex. Seventy-three percent of students were Caucasian and only 9% were eligible to receive free or reduced lunch. The organization of the school was unique in that all core academic subjects were coordinated across the school where three teachers (English Language Arts, Social Studies, and Science) had a group of 90 students each. Each 90 student group was referred to as a "pod." The teaching staff was fairly well-experienced given that, on average, teachers had 12 years of experience and 84% of teachers had tenure. Based on the 2004-2005 state standardized test scores, this school's performance was considered average given that the number of students classified as proficient or advanced was 66% in Reading and 25% in Mathematics.

Ohio Sites

There were two sites in Ohio included in the study given that they were both small schools that had only two Biology teachers each. The first site in Ohio (Site 3—OH) was a small secondary school that served approximately 550 students in grades 9 through 12. The school resided in a rural community located approximately 50 miles south of the state capital of Columbus. The community was almost all Caucasian and had a median household income of approximately $37,000. Ninety-six percent of students were Caucasian and twenty-five percent of were eligible to receive free or reduced lunch. The student -teacher ratio was 20 to 1. Based on the 2004-2005 state standardized test scores, this school's performance was designated "Excellent" given that the number of students classified as proficient or advanced was 94% in Reading, 81% in Mathematics, and 76% in Science. Given these data, students in this school were the most high-achieving in the study.

The second site in Ohio (Site 4—OH) was also located in a similar rural community, located 84 miles from the state capital of Columbus and approximately 30 minutes from Site 3—OH. The number of students in this secondary school was also approximately 550. The community was almost all Caucasian and had a median household income of approximately $31,000. Ninety-nine percent of students were Caucasian and sixteen percent of were eligible to receive free or reduced lunch. The student -teacher ratio was 19 to 1. Based on the 2004-2005 state standardized test scores, this school's performance was designated as "Continuous Improvement" given that the number of students classified as proficient or advanced was 77% in Reading, 73% in Mathematics, and 61% in Science.

New Jersey Site

The New Jersey site was a large secondary school serving over 1,700 students in grades 9 through 12. The school resided in an urban/suburban community located approximately 30 miles southeast of New York City. The community was primarily African-American and had a median household income of approximately $50,000. Seventy-five percent of students were African American and twenty-five percent were Latino. This school was considered the most socio-economically disadvantaged as indicated by the fact that more than 60% of students were eligible to receive free or reduced lunch. In addition, more than 12% of students were classified as ELL. The school schedule was unique in that all classes were on a block schedule. Students had four 80-minute periods each semester, hence students enrolled in Biology for only one

semester. Half of the participating students were enrolled in Biology in the fall semester and the other half were enrolled in Biology in the spring semester. The student-teacher ratio in most classrooms was approximately 20 to 1, and only 3% of teachers held either an emergency or conditional credential; all other teachers held full credentials. Based on the 2004-2005 state standardized test scores, this school's performance was considered below average given that the number of students classified as proficient or advanced was 60% for Reading and 35% for Mathematics.

School Demographics

Table 1 indicates the school level demographics[2] for the five participating sites. Across the five sites, one was in an urban/suburban area, one was in a suburban area, one was in a suburban/rural area, and two sites were in rural areas. The school size ranged from 543 at one of the Ohio schools to 2,461 at the California school. At three of the sites (Site 1—CO, Site 3—OH and Site 4—OH) the majority of students were Caucasian. At the California site, students were predominately Latino (74%), while at the New Jersey site, students were primarily African-American (75%). All schools served students who qualified for free and reduced lunch, however, the range was large. At the Colorado site, only 9% of students qualified for free and reduced lunch while at the NJ site 61% qualified. The median household income across all of the sites was comparable, although the rural sites in Ohio were almost $20,000 lower than the other three sites (but the cost of living in this geographic region was also much lower than the other locations). In California and the two Ohio sites, the curriculum was implemented from September 2005 - May 2006. In Colorado, the curriculum was implemented from September to late March, and then teachers taught Earth Science for the remainder of the year. In New Jersey, the teachers implemented the curriculum twice over the school year with two groups of students (once in fall and once in spring). While the length of the intervention varied between sites, length of the intervention did not vary within sites in that the control and treatment groups were considered as actively participating in the study at exactly the same intervals at each site.

To determine the academic performance level of each school, we gathered data on the percent of students in 9th grade classified as proficient or advanced on their 2004-2005 state testing in English language arts, Mathematics, and Science. The lowest-

[2] The school-level demographics were gathered from www.publicschoolreview.com.

performing school academically (compared to other schools in their respective states) was Site 1—CA and the highest performing school was Site 3—OH. It is important to note that since each state defines "proficiency" differently (as well as used different standardized tests/ measures that contribute to a "proficient" rating), the criteria used to classify schools as lower or higher performing may be specious. For example, it appears that criteria for California may be stricter than other states and therefore may not, in fact, be the site with the lowest performing students.

In summary, the schools participating in this study were diverse in location, ethnic composition, socioeconomic background, school size, and academic performance. Taken together, these sites allowed us a unique opportunity to determine the effectiveness of the *Biology* curriculum across diverse school conditions.

Table 1. School-Level Demographics of Participating Schools.

Demographic Characteristics		CA	CO	OH		NJ
Location		Suburban	Suburban / Rural	Rural	Rural	Urban/ Suburban
School Size		2,461	2,024	566	543	1,709
Ethnicity	% Caucasian	22%	73%	96%	99%	0%
	% Latino	74%	20%	0%	0%	25%
	% African - American	1%	2%	2%	0%	75%
Free/Reduced Lunch		33%	9%	25%	16%	61%
Community Measures	Age 25+ with College Degree	21%	25%	18%	13%	31%
	Median Household Income	$49,836	$52,570	$37,116	$31,318	$51,406
State Testing Results	% Proficient/ Advanced in Reading	32%	66%	94%	77%	60%
	% Proficient/ Advanced in Math	15%	25%	81%	73%	35%
	% Proficient/ Advanced in Science	22%	Not Publicly Available	76%	61%	Not Publicly Available
Length of Curriculum Implementation		Sept to June	Sept to March	Sept to June	Sept to June	Sept to January (1 semester)

Study Sample

Student-level participation was secured in a two-stage process. First, parental consent was gained through an introductory letter that contained a passive parent consent form. Second, student assent was gained through a student assent form that was attached to the student attitude survey. Less than ten students and/or parents over the five sites declined participation in the study; however, all other parents passively consented to their child's participation, and all remaining students assented to participate in the study. A summary of student demographic characteristics (and prior academic characteristics as measured via GPA and the Biology pretest score) is presented in Table 2.

Table 2. Study Sample Demographics of Participating Students.

Demographic Characteristics		Treatment (n=671)	Control (n=730)
		Percent (%)	Percent (%)
Grade Level	9th Grade	26.5	36.9
	10th Grade	64.6	55.8
	11th Grade	8.0	6.6
	12th Grade	0.9	0.7
Gender	Male	46.4	48.8
	Female	53.6	51.2
Ethnicity	Caucasian	41.0	47.4
	Latino	34.3	31.3
	African-American	21.1	15.8
	Asian	1.6	2.6
	Multi-Ethnic	0.5	0.6
	Other	1.4	2.2
Primary Language	English	81.0	81.1
	Other	13.8	12.3
	Unknown	5.2	6.6
GPA [a] (Mean) (1-4 scale)		2.92	3.11
Pretest (Mean Percent Correct) (0-100%)		56.15	56.17

[a] Statistically significant at .05 level.

The sample for the present study consisted of 1,401 students (671 treatment and 730 control) who comprised 45 teacher groups (23 treatment and 22 control). As Table 2 indicates, the treatment and control groups were fairly comparable with respect to percentage of students whose primary language was not English and the gender distribution. The grade level composition suggested that the treatment group had slightly lower percentage of 9th graders (26.5%), but slightly higher percentage of 10th graders (64.6%), than the control group (36.9% and 55.8%, respectively). The treatment group had slightly higher percentage of Latino (34.3%) and African-American students (21.1%) than the control group (31.3% and 15.8% respectively). Conversely, the treatment group had slightly lower percentage of Caucasian students (41%) than the control group (47.4%). Therefore, demographically in terms of student ethnic composition, the treatment group had a slightly higher percentage of minority students than the control group. These differences were controlled for when estimating the impact of the curriculum on student achievement.

All classes were classified by the schools as college-preparatory Biology classes with two exceptions. Students were assigned to classes based on achievement level at one school in Ohio, and for one class period in New Jersey, more than 50% of the class was considered "special needs" students. All participating classes in both groups were considered high school Biology or Life Science suggesting that the curriculum and standards were similar (i.e., the classes generally covered the same five units of the textbook).

Sample Attrition

As shown in Table 3, a total of 130 treatment students (approximately 19% of the treatment group) and 163 control students (about 22% of the control group) did not have posttest scores. This was expected given the 2000 US Census suggested that nationwide, 15%-18% of students changed schools for reasons other than grade promotion (Retrieved from http://www.census.gov/prod/2001pubs/p20-538.pdf.).

Table 3. Percentage of Students Who Did Not Have Posttest Scores

	Students with Pretest and Posttest	Students Missing Posttest	Total
Treatment	541 (80.6%)	130 (19.4%)	671 (100%)
Control	567 (77.7%)	163 (22.3%)	730 (100%)

To examine whether sample attrition created differences between the treatment and control groups, we conducted exploratory analyses in two different ways. First, taking into account those whom did not have posttest scores (i.e., excluding those who had missing posttest scores), we investigated whether treatment and control students differed with respect to key demographic and prior academic characteristics. A summary of these data is presented in Table 4.

Table 4: Key Demographic and Academic Characteristics: Treatment vs. Control Taking Missing into Account

Demographic Characteristics		Treatment (n=541)	Control (n=567)
		Percent (%)	Percent (%)
Grade Level	9th Grade	28.7	36.7
	10th Grade	64.9	57.0
	11th Grade	5.7	6.0
	12th Grade	0.7	0.4
Gender	Male	45.8	48.7
	Female	54.2	51.3
Ethnicity	Caucasian	42.2	49.0
	Latino	35.0	30.8
	African-American	18.9	14.5
	Asian	1.9	2.7
	Multi-Ethnic	0.4	0.5
	Other	1.7	2.5
Primary Language	English	84.5	86.6
	Other	15.3	12.9
	Unknown	0.2	0.5
GPA [a] (Mean) (1-4 scale)		2.97	3.16
Pretest (Mean Percent Correct) (0-100%)		56.20	57.25

[a] Statistically significant at .05 level.

As indicated in Table 4, the distribution of the key demographic and academic characteristics between the treatment and control groups taking missing into account parallels that of the original treatment vs. control groups. In other words, the sample attrition did not create any differences between treatment and control group on key background characteristic variables or prior academic characteristics such as the Biology pretest and GPA.

Table 5. Key Demographic and Academic Characteristics: Original Sample vs. Sample with Posttest Scores

Demographic Characteristics		Original Sample	Sample with Posttest
		Percent (%)	Percent (%)
Grade Level	9th Grade	32.8	31.9
	10th Grade	60.8	60.0
	11th Grade	5.9	7.3
	12th Grade	0.5	0.8
Gender	Male	47.3	47.6
	Female	52.7	52.4
Ethnicity	Caucasian	45.7	44.4
	Latino	32.8	32.8
	African-American	16.6	18.4
	Asian	2.3	2.1
	Multi-Ethnic	0.5	0.5
	Other	2.1	1.8
Primary Language	English	85.6	81.0
	Other	14.1	13.0
	Unknown	0.2	0.5
GPA (Mean) (1-4 scale)		3.07	3.02
Pretest (Mean Percent Correct) (0-100%)		56.16	56.74

Next, we ran exploratory statistical analysis to see whether the original sample and the sample that consisted of those whom posttest scores were obtained differed in their key demographic and academic characteristics. Table 5 indicates that the original sample (pretest) and the sample consisting of those with pretest and posttest scores did not differ in their key demographic and academic characteristics. In summary, student attrition did not alter the nature of the study sample.

Comparability of Groups by Teacher Characteristics

In addition to examining comparability in student characteristics across treatment and control groups, it was important to verify that teacher characteristics were also comparable. Random assignment was theoretically supposed to equate all groups (student-level and teacher-level), however, with only 16 teachers, it was possible that there were systematic differences in teacher characteristics across conditions.

Table 6. Teacher Characteristics Across Treatment and Control Groups.

Teacher Characteristics	Treatment Condition (n = 8)	Control Condition (n = 8)
Highest degree attained	4 (50%) = M.A. 3 (38%) = B.A. 1 (12%) = M.D.	5 (63%)= M.A. 2 (25%)= B.A. 1 (12%)= M.D.
Number of years teaching K-12	11.0 yrs	11.6 yrs
Number of years teaching biology	10.4 yrs	9.8 yrs
Number of years teaching at this school	8.0 yrs	8.5 yrs
Enjoyment of teaching biology (1 low, 5 high)	5.0	4.8
Familiarity with state standards (1 low, 5 high)	5.0	4.4
Personal comfort with computers (1 low, 5 high)	4.8	4.6
Comfort level using technology in the classroom (1 low, 5 high)	4.5	4.4

Table 6 illustrates that treatment teachers were comparable to the control teachers in terms of individual characteristics. The most common degree attained in both groups was a Master's Degree; one teacher in each group also possessed their Medical Degree. All teachers had been teaching high school for approximately the same amount

of time. In addition, the teachers had been teaching Biology/Life Sciences for the majority of their teaching career. All were teaching at their respective schools for an average of 8-8.5 years, and all reported enjoying teaching Biology. On average, all teachers reported a high level of comfort using computers, including accessing the internet and websites, as well as a high level of comfort using computer technologies in their classrooms.

SECTION THREE: DESCRIPTION OF EVALUATION DESIGN, MEASURES, AND PROCEDURES

Evaluation Design

A Randomized Control Trial (RCT) was conducted to test the effectiveness of Miller and Levine's (2006) *Biology* on student attitudes and achievement in Biology. Multiple process measures were collected throughout the year and two outcome measures (Biology achievement and student attitudes towards science) were collected at pretest (fall, 2005) and posttest (spring, 2006).

An equal number of teachers were randomly assigned to the treatment or control group within each site. For Site 1—CA, Site 2—CO, and Site 5—NJ, two teachers from each individual site were randomly assigned to the treatment group (using the 2006 Prentice Hall *Biology* curriculum) and two teachers were randomly assigned to the control group (using their existing curriculum). Given the smaller size of schools in Ohio, Site 3—OH and Site 4—OH were generally considered one "site". Researchers randomly assigned one teacher at each school to the treatment group and the other teacher to the control group. The assignment process involved randomly selecting teacher's names out of hat; the names were drawn first for the treatment group, followed by the control group. The researchers then informed teachers of their assigned group and subsequently obtained enrollment information for each participating class period.

Implementation Guidelines

To ensure that treatment teachers implemented the curriculum as intended by the publisher, specific implementation guidelines were developed in collaboration between CGU researchers and Prentice Hall science editorial and marketing science staff during the pilot study (see Appendix A). Guidelines were tested for feasibility during the pilot study and then were revised for use in this study. Guidelines included elements of the program that were required in all participating treatment classrooms (Required), encouraged (Highly Recommended), or optional (Optional). Optional elements also included the use of any other Prentice Hall *Biology* (2006) element not specifically outlined in the implementation guidelines. All participating treatment teachers were instructed to follow the implementation guidelines (and signed a MOU to this effect), while complying with their specific state content standards in Biology/ Life Science.

Teacher Compensation

Compensation for participation in the study was a $250 stipend for all teachers. In addition, all teachers in the treatment group were provided with a teacher's edition textbook and all available ancillary materials as well as a budget for $500 to use toward laboratory supplies. Each school received enough student edition textbooks for the number of participating students in the treatment group prior to the start of the 2005-2006 school year as well as enough textbooks for the number of students in the control group at the conclusion of the school year.

Teacher Training

Implementation of the RCT study began with teacher training held at all participating sites in August 2005 (with the exception of one site, in which training was held in early September). Teacher training was comprised of two distinct sections: research study orientation and product training. All treatment and control group teachers received the same research study training. Only treatment teachers received product training from a Prentice Hall consultant (and the researchers). Training sessions were held at each respective school site at three of the five individual schools. The two schools in Ohio were combined at one location for convenience, and teachers in California were trained at the district office.

Research Study Orientation

The first phase of the research study orientation was provided to all teachers prior to the beginning of the study at their local school sites in August or September, 2005. A detailed description of the study activities, timelines, study purpose, research questions, and expectations for participation was provided to all teachers. A second phase of the research study orientation was provided approximately six weeks after the first phase. The second phase involved reviewing study procedures and training teachers how to complete the online teacher implementation logs. A member of the research team provided direct instruction on the use of the online logs and ensured that all teachers were comfortable with completing the logs before the session concluded. All participating teachers (treatment and control) attended both sessions of the research study orientation.

Product Training with the Treatment Teachers

After the first phase of the research study orientation session, all treatment teachers received 2-4 hours of product orientation with the same Prentice Hall Science consultant at all sites. The initial training consisted of the consultant using a powerpoint presentation to review a standard training protocol based around the categories of 1) Plan 2) Teach 3) Assess. Teachers had all training materials available during the training sessions, including teacher edition textbooks, and other ancillary materials such as overhead transparencies, workbooks, lab manuals, and *Teacher Express* (including *Exam View* and *Presentation Pro*). After the consultant reviewed the standard training protocol, he allowed teachers to visit the product website to review additional resources. Teachers also registered their classes online at the www.phschool.com website to facilitate use of the online class organization and assessment scoring. Teachers were given access to the consultant as well as a technology hotline if they had technical problems with any software or programs. The researchers specifically requested that the consultant provide teachers with information regarding appropriate pedagogy that complimented the *Biology* text. This included focusing teachers on addressing students' prior knowledge; stopping at "checkpoints" to review reading comprehension; using a variety of materials and strategies to differentiate instruction for heterogeneous student groups; and using the section and chapter assessments to scaffold student understanding (designed to challenge students to think critically about the content of the information in a strategic way).

Approximately four weeks after the first product training session, the same consultant provided 1-2 hours of follow-up training (with the exception of two schools which had local follow-up trainers). The focus of this training was to answer any questions about the products and also to discuss, in-depth, the technological components of the curriculum (e.g., online system, software, and *i-text* (Interactive textbook, etc.). All treatment teachers attended the follow-up training. The session concluded after the trainer was assured that all teacher questions and concerns were addressed.

Implementation Measures

Implementation measures were developed to monitor and assess the activities in the participating classrooms throughout the year. Implementation measures included weekly teacher implementation logs, classroom observations, teacher interviews, and a

teacher survey. Since August 2005, teachers also communicated informally via email, phone calls, open-ended sections of the teacher logs and in informal interviews during the fall and spring observations. Treatment teachers were continuously encouraged to provide feedback about the Miller & Levine *Biology* curriculum throughout the study.

Teacher Implementation Log

Description of Measure: All classroom-related activities were monitored to ensure that (1) an adequate level of implementation was present in all classrooms; and (2) Prentice Hall *Biology* products were being implemented as intended by the publisher. Implementation logs were developed for control and treatment teachers separately to record both the content covered in their classes (e.g., unit, chapter, section, etc.) as well as the delivery of that content (e.g., assigned reading, lab exercises, etc.) The structure of the logs was based on the implementation guidelines. Researchers used online survey software, *surveymonkey.com*, to program the logs. The online log process also allowed researchers to remain in constant communication with study participants so that issues such as dates for training, observations, and test administration could be planned effectively.

Implementation of Measure: Participating teachers submitted online logs starting mid-September 2005 through the end of the curriculum implementation. At the end of each week, participating teachers were sent an email with a URL link to complete the log for that week. Researchers checked the status of these logs on a weekly basis to verify completion and appropriateness of responses. In only a few numbers of cases did teachers fail to complete the log for a particular week. In this instance, a member of the research team contacted the teacher directly within three days of the following week to ensure that the full log was completed. We have collected complete implementation data for all participating teachers for the entire study period via the online logs.

Classroom Observation

Description of Measure: The researchers used an observation protocol that was developed during the pilot study. The instrument was validated after several trials assessing inter-rater reliability of greater than $r = .90$ on all categories. Extensive training was provided to the observation team regarding how variables were operationalized (e.g., defining “student engagement”). The researchers then observed all pilot classrooms in teams, with careful debriefing occurring after each class session. The final instrument that was established was used in the current study. The observation protocol assessed elements in the classroom such as physical environment; level of student

engagement; allocation of classroom time; level of "cognitive demand" expected of students; and effective teacher practices.

Implementation of Measure: Every participating classroom was observed by members of the research team on one occasion during fall 2005, and again, in spring 2006. Researchers remained in the classroom for the entire class period. Teachers were instructed to organize a lesson based on a "typical" day (free from any exams, assemblies, etc.). Most teachers complied with this request and therefore "typical" lessons were observed in each classroom during fall and spring.

Teacher Interview

Description of Measure: As part of the debriefing process, a teacher interview protocol was developed for all participating (treatment and control) teachers. Questions were identical for both groups with the exception of product satisfaction questions specifically geared toward the Prentice Hall Miller & Levine *Biology* curriculum.

Implementation of Measure: In May – June 2006, all teachers were interviewed by phone (with the exception of one site in which they were interviewed in person) by a member of the research team. The interviewer followed the established protocol except where follow up questions were necessary for the purpose of clarification. Each interview lasted approximately 30 minutes.

Teacher Survey

Description of Measure: An additional part of the debriefing process and product satisfaction assessment included an online teacher survey created using the *surveymonkey* program. Questions on the teacher survey included topics such as teacher education, experience, and interest in teaching science.

Implementation of Measure: All teachers completed the teacher survey online via their personal computers in June 2006, after full completion of the RCT study.

Outcome Measures

There were two primary outcome measures in this study: a student survey and a Biology achievement assessment. The goal of the student survey was twofold. First, although random assignment theoretically eliminated any pre-existing differences between groups, speculation exists whether all pre-existing differences between groups could be eliminated in practice (e.g., Davies, Williams, & Yanchar, 2004). Thus, it was important to measure factors that theoretically related to the primary outcome (i.e., Biology achievement) so these factors could be controlled in subsequent analyses.

Second, a student survey measures changes in students' attitudes, which often contribute to changes in students' achievement. Since curriculum studies generally reveal that effect sizes in student's science achievement are typically small (Weinstein, Boulanger, & Walberg, 1982), including for activity-based elementary science programs (Bredderman, 1983), the student survey provided a tool to measure the impact of a curriculum on student factors that are often highly correlated with student achievement.

A Biology achievement test was the other outcome measure in this study. This was the primary instrument that measured the impact of the Prentice Hall *Biology* curriculum in comparison to the control curriculum. Descriptions of each measure (and their corresponding psychometric properties) are described below.

Student Survey

Description of Measure: The researchers developed a student survey for use in the Prentice Hall *Biology* pilot study. The initial set of questions was based on a thorough review of the literature that examined salient factors that may contribute to students' attitudes about or interest in science. For example, factors that have been previously linked to student interest in science and achievement include self efficacy (Schunk & Pajares, 2001); intrinsic motivation (Flink, Boggiano, Main, Barrett, & Katz, 1992); parent involvement (Fan & Chen, 2004); meta-cognitive skill including study skills (Carr, Kurtz, Schneider, Turner, & Borkowski, 1989; Flavel, 1979; Garner, 1987); and specific classroom factors such as the teacher's influence (NRC, 2000). A total of 21 questions were included in the student pretest survey (excluding demographics), and an additional 20 questions related to product satisfaction were included on the posttest survey.

To validate the instrument, an Exploratory Factor Analysis was conducted separately on pretest and posttest survey responses, using the Principal Axis Factoring extraction method with Varimax rotation. The factor solutions on the pretest and posttest survey responses were comparable, given that five factors with eigenvalues greater than one were revealed on the pretest (after discarding a few items that cross-loaded on multiple factors), while six factors were revealed on the posttest. On the posttest, there was differentiation between the intrinsic motivation items and the meta-cognitive (study) skills questions, while on the pretest, those items were grouped together. Based on the posttest EFA, 45% of the variance was explained by the rotated six-factor solution: *Attitudes Towards Science (a = .86), Study Skills and Habits (a = .78), Teacher Influence (a = .74), Self-Efficacy in Science (a = .68), Parent Involvement (a = .72), and*

Preference for Challenging Work (a = .78). As predicted, each of these factors (except parent involvement) was significantly related to students' Biology posttest scores ($p < .01$). These correlations provide support for the construct validity of the measure.

Implementation of Measure: All students participating in the study were required to complete the survey in fall, 2005 and again in spring, 2006. Researchers obtained student rosters from each school and provided each teacher with specific instructions for administration of assessments, including expected length of time and conditions for administration. There were 1,178 students who completed the pretest; 1,139 students who completed the posttest. Of this amount, there were a total of 1,019 students (520 treatment, 499 control) who completed the pretest and posttest survey.

Standards-Based Science Assessment

Description of Measure: This assessment was developed by the Partnership for the Assessment of Standards-based Science (PASS), a subdivision of the educational laboratory, WestEd. PASS has developed science assessments that are aligned to the content recommendations of the *National Science Education Standards* (NSES) and the *Benchmarks for Science Literacy* (BSL) (Comfort, 1994). Over the past decade, PASS has collected solid technical data on the reliability and validity of their assessments. In addition, scaling and equating studies were completed to produce scale scores (McCaffrey, Hamilton, & Aronson, 1998; Wilson, Delgado & Finklestein, 1998). Given that there was no specific Biology assessment available, the Biology assessment was constructed from a pool of existing life science items (each containing high inter-item reliability). The assessment included both a 30-question enhanced multiple-choice section as well as one constructed response item (worth a total of 12 points). Due to test security issues, PASS did not allow a copy of the test to be published in this report.

Implementation of Measure: A standards-based, nationally recognized Biology assessment was administered in fall, 2005 and, again, in spring 2006. There were 541 students in the treatment group and 567 students in the control group with complete pretest and posttest scores on the Biology assessment.

Summary of the Data Collection Procedures across Sites

Table 7 summarizes the timeframes for the teacher training, implementation measures, and outcome measures for this study in Month and Year format.

Table 7. Timelines for Data Collection and Activities Across Sites.

Study Implementation Procedures	CA	CO	OH		NJ
	Site 1—CA	Site 2—CO	Site 3—OH	Site 4—OH	Site 5—NJ
Study Orientation	Aug 2005	Aug 2005	Aug 2005	Aug 2005	Sept 2005
Product Training	Aug 2005	Aug 2005	Aug 2005	Aug 2005	Sept 2005
Follow-up Product Training	Oct 2005	Oct 2005	Oct 2005	Oct 2005	Sept 2005
Administration of Pretest Assessments	Oct 2005	Sept 2005	Sept 2005	Sept 2005	Oct 2005/ Jan 2006
Administration of Posttest Assessments	May 2006	Mar 2006	May 2006	May 2006	Jan 2005/ May 2006
Fall Classroom Observation	Nov 2005	Nov 2005	Nov 2005	Nov 2005	Oct 2005
Spring Classroom Observation	Mar 2006	Feb 2006	Mar 2006	Mar 2006	May 2006
Teacher Interviews	June 2006	May 2006	May 2006	May 2006	May 2006
Teacher Survey	June 2006	June 2006	June 2006	June 2006	June 2006
On-Line Logs	Sept 2005 – May 2006	Sept 2005 – Mar 2006	Sept 2005 – May 2006	Sept 2005 – May 2006	Sept 2006 – Jun 2006

Where possible, researchers attempted to maintain as much consistency across the sites that corresponded to individual school sites' academic calendars. For example, all study training was conducted prior to the start of the school year at all sites, however, because sites varied when the school year began, training for each site did not occur within the same month. Rather, training occurred at relatively the same rate for all of the sites (e.g. four of the five sites had initial training in August, 2006 whereas Site 5—NJ had this training in early September, 2006). With the exception of product training, all monitoring, implementation and outcome measures were conducted at exactly the same time at each site across treatment and control groups.

SECTION FOUR: ASSESSMENT OF CURRICULUM IMPLEMENTATION

It is difficult to interpret student outcomes appropriately without understanding what transpired within treatment and control classrooms. How the curriculum was implemented in classrooms, including an examination of the content covered, breadth, and quality of implementation, was assessed through classroom observations, teacher interviews, and online teacher logs. The following section summarizes curriculum implementation during the study.

Alignment of Treatment to Control Textbooks

The control group used three distinct versions of Biology textbooks. Comparison textbooks included a much earlier version (1998) of a competing Prentice Hall textbook at one school, and other schools used editions of competitor's textbook, with publication dates in 2005 and 2003. To determine differences in the content covered and the breadth of that coverage, we first had to align each of the control texts to the treatment textbook. The process of alignment involved compiling a list of all units, chapters, and sections covered in the treatment textbook, and then a content expert in Biology aligned each section in the control textbook with a converging theme in the treatment textbook. For example, section 1 of Chapter 37 in the treatment textbook "The Circulatory System" is similar to Section 2 of Chapter 37 in one of the control texts. Once topics from the control texts were aligned (and standardized) with those covered in the Miller & Levine *Biology* text, we compared Biology content covered (and the breadth of coverage) by teacher and across conditions.

Content Covered by Treatment and Control Teachers

Researchers instructed control teachers to use all materials and curriculum as they would in a "typical year" or as they have in "past years". This included conducting laboratory experiments, reading assignments, use of the text or any other ancillary materials. Control teachers did not receive direct product training for the control curriculum prior to the study because all control teachers were either trained or had used the control curriculum previously. Teachers in the control group were not given any additional instructions for curriculum implementation, however, all teachers agreed to follow study protocols. All participating teachers were also required to adhere to their

states' curriculum standards (including any district and school pacing guides) as a primary consideration.

We analyzed the content covered in classrooms by condition to determine comparability among groups. "Coverage" of a particular concept was defined as any of the following criteria: reading assigned in class or for homework; laboratory activities performed; discussion or writing assignment or inclusion of the topic in activity or assessment. Table 8 is a brief summary of the topics covered by the number of teachers in each group—ratings of coverage included "All" (8 of 8 teachers), "Most" (6-7 of 8 teachers), "Half" (3-5 of 8 teachers), "Few" (1-2 of 8 teachers), and "None" (0 of 8 teachers). Specifically, the topics in Table 8 comprised major areas of coverage in the first five units of the Miller & Levine *Biology* text (one of the "Required" elements on the Implementation Guidelines).

Table 8. Coverage by Topic Area.

Unit*	Topic	Treatment	Control
1	Introduction/ Scientific Method	All	All
	Chemistry: Matter, Carbon, Chemical Reactions	Most	Most
2	Ecology: Energy Flow, Cycles	All	Half
	Ecology: Climate, Biomes	Most	Half
	Populations	Most	Few
	Conservation	Half	Few
3	Cells: Structure, Function, Membranes, Diffusion	All	All
	Cells: Growth & Division	Most	Most
	Cells: Respiration, Energy & Photosynthesis	Half	Half
4	Genetics	Most	All
	Meiosis	Most	All
	DNA	Most	Half
	Applied Genetics: Human Heredity	Few	Few
5	Evolution: Darwin	Most	Most
	Evolution: Mechanisms	Half	Half
10	Body Systems	Half	Few

*Unit in Prentice Hall *Biology* text Miller & Levine (2006)—matched by topic with control texts

In general, treatment teachers covered more topics listed in Table 8, but this did vary by topic. For example, all teachers covered topics such as the scientific method and cells. However, more treatment teachers had coverage on topics such as Ecology and

DNA, whereas more control teachers had coverage on topics such as Genetics and Meiosis. In addition to the first five units of the *Biology* text, some teachers covered parts of other units throughout the book (e.g., Animals and Body Systems). The only significant area in which several teachers covered one or more section of the unit was on Body Systems (unit 10 in Miller & Levine); however, more treatment teachers covered this area compared to control teachers. These results suggest that although coverage differed slightly by topic, in general, treatment teachers covered the first five units of the text and control teachers covered the equivalent units in their textbooks.

Breadth of Curriculum Implementation Across Treatment and Control Groups

To measure "breadth of coverage", we summed the total numbers of units, chapters and sections covered for each participating teacher from the online teacher logs. "Coverage" of a particular concept was defined as any of the following criteria: reading assigned in class or for homework; laboratory activities performed; discussion or writing assignment or inclusion of the topic in activity or assessment. "Coverage" for chapters includes at least one section within a chapter as having been considered covered, and "coverage" for a unit consists of a section within a particular unit as being "covered". Table 9 provides a comparison of coverage for treatment and control groups.

Table 9. Breadth of Coverage for Treatment and Control Groups.

	Mean Number Covered		Range	
	Treatment	Control	Treatment	Control
Units	5.75	5.13	3 – 8	3 - 7
Chapters	15.38	11.63	5 - 25	8 - 17
Sections	39.00	24.50	16 - 60	16 - 34

Six of the eight treatment teachers covered all five of the recommended units during the implementation period. Some units were fully covered by most teachers (Unit 1); however, some units were covered sparsely by teachers. For example, of the six

units covered by T5, most chapters within the first five units were covered, however; only one chapter in the 6th unit was covered. Other treatment teachers had nearly complete coverage on additional units. For example, T7 had most chapters of more than seven units covered during the implementation period.

For the control teachers, a similar number of units were covered in comparison to the treatment teachers at their respective sites, with the exception of one site, in which treatment teachers exceeded control teachers in number of units covered, and far exceeded them in number of chapters and sections covered. In general, teachers at one site covered fewer sections, chapters and units in comparison to the other schools. These data seemed appropriate given that Earth science curriculum was required in classrooms for the last part of the school year at this site.

Overall, treatment teachers had more complete "coverage" of Biology topics in their classrooms throughout the year. The most specific analysis accounts for the number of sections teachers reported having covered during the implementation period. For example, on average, treatment teachers covered significantly more sections (M = 39) than control teachers (M = 24.75), $t(18) = 3.50$, $p < .01$. (Note: Degrees of freedom = 18; 16 teachers with one overall number of sections and 4 teachers from one site which reported two numbers of sections—fall and spring). It is possible that treatment teachers covered more content in their classes compared with control teachers because of their adherence to the implementation guidelines (e.g., first five units of the text were "required").

Distinguishing Features and Adherence to Implementation Guidelines

Examining coverage of topics was so important because if students did not have a similar opportunity to learn about concepts in their Biology classes, they may show differential effects on student outcomes. This was the rationale for establishing implementation guidelines for treatment teachers—to ensure a minimum amount of coverage on relevant Biology topics. Another rationale for requiring teachers to complete specific parts of the program was based on the logic that distinguishing features of the Miller & Levine text will have an impact on student performance. Distinguishing features included: (1) sections of the text that focus on real-world applications such as "Technology and Design" and "Careers in Science"; (2) hands-on activities and labs, in the classroom or on-line; (3) sequenced student assessments; (4) frequent assessments

(sections or chapter levels); and (5) strategies and materials to enhance differentiated instruction for all students. Three of these elements (2, 3, 4) were directly tied to the implementation guidelines, so an assessment regarding the distinguishing features of the program can be determined by reviewing the extent to which teachers adhered to the implementation guidelines. An analysis of the other two distinguishing features (1,5) will be assessed after a review of the implementation guidelines.

Implementation Guidelines: Hands on Activities & Labs and Assessments

All treatment teachers were provided with implementation guidelines that they agreed to follow throughout the year. Table 10 indicates the extent to which individual treatment teachers followed such guidelines (specifically the number of sections within the text that each teacher adhered to the guidelines). Most categories were self-explanatory; however, some require further explanation. For example, for the item "Complete one lab per unit", this meant that students/ teacher conducted a laboratory exercise in the classroom or for homework with one of the extended laboratory activities (Real-World Lab, Design an Experiment, or Exploration) or completed a Quick Lab. In addition, students could complete section assessments either in writing or orally.

In general, most treatment teachers followed implementation guidelines established by the researchers/ publisher. However, there were two areas in which teachers did not adhere to the guidelines—coverage of additional units other than the first five units, and completion of at least one laboratory activity per unit. Two teachers did not complete the first five units and neither of these teachers completed any additional chapters in units outside of those required. None of the treatment teachers (except for one) had students complete a lab within each unit covered. While all teachers generally completed more than five labs over the entire year, they generally completed more than one lab in each unit and no labs other units. Two teachers completed labs in nearly one-third of the sections that they covered, but only one had labs with coverage in five separate units.

These data suggest that although a distinguishing feature of the program was that laboratory exercises (either in the classroom or online) would enhance student learning, teachers did not have time/ interest in conducting more frequent lab activities. Only two treatment teachers used Virtual Labs and Lab Simulations consistently. In comparison to the control group teachers, treatment teachers conducted a similar number of laboratory exercises in their classrooms throughout the year. Therefore, it was difficult to assess the extent to which this may be a distinguishing feature of the

program when in reality teachers did not have time to integrate such activities in their classrooms.

Table 10. Curriculum Implementation According to Implementation Guidelines.

	T1	T2	T3	T4	T5	T6	T7	T8
Required Elements								
Primary Use of PH *Biology* Text	Yes 100%	Yes 100%	Yes 96%	Yes 94%	Yes 100%	Partial 64%	Yes 100%	Yes 100%
Cover First 5 Units in Text (Most chapters within a unit)	Yes 5 of 5	Yes 5 of 5	Partial 4 of 5	Partial 3 of 5	Yes 5 of 5	Yes 5 of 5	Yes 5 of 5	Yes 5 of 5
Cover 1-2 Additional Units in Text	No 1 extra	No 1 extra	No 0 extra	No 0 extra	No 0 extra	No 1 extra	Yes 2 extra	Yes 2 extra
Cover All "Key Concepts" in chapter	Yes 78%	Partial 67%	Partial 42%	Yes 88%	Yes 96%	No 13%	Yes 93%	Yes 74%
End of Chapter Assessments* (Exam View *Teacher Express* or online)	No	Yes	Partial	Yes	Yes	Yes	Yes	No
Use Adapted/Guided Reading and Study Workbook	Yes 86%	Yes 80%	Partial 54%	Yes 75%	Partial 52%	Partial 59%	Yes 99%	Yes 94%
Complete 1 lab per unit	No 5%	Yes 33%	No 15%	No 31%	No 16%	No 10%	No 9%	No 26%
Strongly Recommended Elements								
Complete Section Assessments	Partial 61%	Partial 54%	No 12%	Yes 81%	Yes 72%	Yes 72%	Yes 88%	Yes 80%
Cover Caption Questions	No 5%	No 0%	No 0%	Partial 25%	No 0%	No 0%	Partial 18%	No 7%
Use of Transparencies or Presentation Pro	Partial 35%	Yes 75%	Partial 43%	Yes 94%	Yes 64%	Partial 39%	Yes 91%	No 20%

* Exact percentages lacking, ratings based on teacher interviews; most reported frequent use of Exam View

Another distinguishing feature related to assessments. End of chapter assessments were required, and end of section assessments were strongly recommended. Most teachers reported during teacher interviews that they consistently used the *Exam View* test bank to create chapter tests for their students. This resource

was extremely helpful for teachers in terms of saving time, and targeting specific level of questions to students, etc. For the section assessments, only one of the treatment teachers failed to complete the end of section assessments that were strongly recommended. More information regarding specific product satisfaction for *Exam View* and other materials can be found in Section Six.

Real World Applications

Another distinguishing feature that was not required in the implementation guidelines was the "real world" applications of the text and ancillary materials. These "real world" sections included Issues in Biology, Technology & Society, Biology and History, and Careers in Biology. In general, very few teachers reported that they specifically integrated any of these sections into their classes. It would appear as if teachers would naturally want to integrate these topics into their class discussions, however, the reality is that very few took advantage of this during the study. Possible explanations for this could be (1) teachers did integrate them into their classes but did not report it; (2) teachers did in fact review the content of these sections in classes but did not attribute it to the named section; (3) since these elements were not required as part of the implementation guidelines, teachers did not consider them as important as other components of the program; and/or (4) the emphasis on standards-based education allows very little room for integration of other topics, regardless of how useful or relevant they might be to students.

Differentiated Instruction

The final distinguishing feature related to differentiating instruction for all students. Teachers reported how frequently they referenced the "Differentiated Instruction" section of the textbook in their logs, however, not all teachers used this section in the text. We felt that teachers could differentiate instruction for students in several other ways and we specifically asked teachers about this topic during their interviews. The following summarizes how treatment teachers were able to differentiate instruction for all students during the study. Teachers reported a variety of responses and some gave specific examples of differentiation. The following are some of those examples.

- Most treatment teachers reported that they were able to differentiate instruction for their students by using either the various levels of the Guided Reading and Study Workbook (Regular or Adapted) as well as other materials such as virtual labs, online resources, etc. For example, lower level

students were able to better understand the material with the adapted workbook than students have in previous years or with other texts.

- Several treatment teachers reported that they frequently used instructions in the margins of the Teacher Edition textbook as a way of guiding them to differentiate instruction for students. This included items such as "Differentiate Instruction" as well as other tips for teaching, addressing misconceptions, etc.
- Some teachers reported that for more advanced students, they either assigned or the student spontaneously completed additional work on their own, including critical thinking questions or standardized test preparation assignments.
- Some teachers reported that for Spanish-speaking students, it was helpful to have resources available in Spanish for them to refer to during the lesson, particularly vocabulary resources.
- Several teachers also reported that they were able to use a combination of ancillary materials to help reinforce concepts with all students, while simultaneously addressing multiple student learning needs. The following is an excerpt from a teacher interview that demonstrates how (s)/he was able to integrate several materials into lessons in order to facilitate student learning through differentiated instruction:

> *I had 4 different stations and I had basically a different work station at each lab table. One station they were doing an online lab, and they were using the virtual CD-Rom that came with the textbook, and at another station we were doing a hands on lab that was actually one of the quick labs that was in the textbook, and a third station they were doing a go online assessment were they could go at the end of the chapter and take a self assessment and it would grade it… or score the test for them. Then the fourth activity was another go online resource and they could go to different websites and then they completed a worksheet. So, I had four different stations, and depending on the level of the student, they didn't have to complete all four, I mean that was the intent for them to complete all four stations but depending upon their level they could complete one, two, three, or all four stations depending upon what they felt comfortable doing...[students] were able to grasp the material better in those sections where I could differentiate than in those when I didn't.*

Based on teacher feedback and our assessments of their use of materials, treatment teachers were able to successfully differentiate instruction for students using the various levels of workbooks, lab manuals, ancillary materials, and specific teacher

instructions on differentiation from the teacher edition textbook. We feel that this particular "distinguishing" feature in the curriculum produced a positive experience for students in these classrooms.

We also investigated whether or not teachers in the control group were able to differentiate instructions for their students. Specifically, we wanted to see if control teachers were using any strategies or materials that would facilitate this, and whether or not this was qualitatively different from the treatment teachers. The following is a summary of the responses for control teachers:

- Most control teachers reported that they were able to differentiate instruction, although a few acknowledged that they needed to do a "better" job.
- Several teachers specifically stated that cooperative group activities allowed them to help more individual students in the classroom and specifically set up lessons to incorporate this strategy.
- Several teachers reported that use of the textbook did not allow them to differentiate instruction for students, but instead relied on several outside sources (web sites, other materials) to help students of varying levels.

In general, we found that most teachers in the study reported that they were able to differentiate instruction for their students; however, there were differences between the treatment and control groups. Specifically, treatment teachers were able to accomplish a more "genuine" level of differentiation based using different levels of materials as well as relying on specific instructions from the publishers. Control teachers generally were able to accomplish differentiating instruction, but by using a variety of internet-based sources and other materials they have generated over many years. There were also similarities in the types of differentiation required within the schools. For example, both treatment and control teachers at one site mentioned that their main concern was to meet the needs of more advanced students, while teachers at the other sites were primarily concerned with lower-level students. In summary, one of the distinguishing features of the Prentice Hall *Biology* program is that teachers have a full range of options for differentiating to students without having to seek outside resources. We feel that treatment teachers in this study were able to do that. However, the extent to which all teachers using the Miller & Levine *Biology* program in the future will be able to accomplish this may depend on having all levels of materials available in their classes (e.g., Adapted and Guided Reading and Study Workbooks as well as Lab Manuals A and B).

Quality of Classroom Environment

One important way to assess curriculum effectiveness was for researchers to be physically present in classrooms. While we have objective ratings on the level of experience each teacher may have, the classroom setting was where most students engage substantially in the content of the course, and subsequently where student learning matters most. While it was impossible to make definitive statements regarding the overall quality of experience for students by observing over two occasions throughout the year, we believed that some measures of overall quality could be observed in this limited timeframe.

- Student engagement: We were concerned with student engagement in the classroom, particularly because of the fact that classrooms were typically heterogeneous environments and skilled teachers attempt to engage all students in the lesson, regardless of ability level. Members of the research team selected random points during a classroom session to reflect the percent of students that were "engaged" in the lesson. Ratings were combined across all raters and observation sessions to comprise a cumulative "student engagement" rating. For those teachers whose average student engagement was between 76% and 90%, they received a rating of "high"; for those teachers whose average student engagement was between 61% and 75%, they received a rating of "medium"; whereas those teachers whose average student engagement was between 45% to 60%, they received a rating of "low". In general, student engagement was more similar within a school than within the treatment condition. There were no differences between treatment and control teachers regarding level of engagement in their classrooms.
- Effective Teacher Practices: Teachers also facilitated student engagement by using strategies and techniques proven to be effective in the classroom. For example, teachers that expected students to engage in substantial content material and have established daily routines have been found to be generally more effective than teachers that lack these qualities (NRC, 2000). We categorized ratings of effective teacher practices into three categories: whether these practices were frequent (FR), varied (VR), or infrequent (IN), based on a rating relative to all teachers in the group. In general, there were no differences between the groups in terms of frequency of effective teacher practices. Overall, teachers at one site in particular had the highest

level of effective teacher practices, however, this was equally divided between the treatment and control groups.

- Teacher Influence: A final area related to classroom environment involved student perceptions of their teachers. Given that teachers can be highly influential in the learning process, we tracked teacher behaviors in the classroom via the student survey. Mean ratings of "Teacher Influence" (Section Five) were calculated for each teacher. This composite factor included items such as "My teacher inspired me to do my best in Biology" and "My teacher explained concepts clearly and accurately". In general, students rated their teachers in a relative positive light in relation to the likert scale of 1-5 (1 = low ratings; 5 = high ratings) for the composite variable of Teacher Influence. However, some teachers were clearly rated more favorably than others.
- Overall Classroom Environment: Based on a combination of classroom dimensions, we developed a qualitative rating of overall classroom environment that included the following dimensions: quality of learning environment (which includes cognitive demand and teacher practice), student engagement, and physical environment. Quality ratings were established based on fall and spring observations. These quality ratings were translated into five categories and labeled "Excellent" (high quality learning environment/ high quality physical space/ high student engagement); "Very Good" (high quality learning environment/ adequate physical space/ high to medium student engagement); "Good" (medium quality learning environment/ adequate physical space/ medium student engagement); "Fair" (medium quality learning environment/ low quality physical space or low student engagement); and "Poor" (low quality learning environment/ low quality physical space or low student engagement level).

In general, there were no major differences between treatment and control groups on student ratings of teacher influence or overall classroom environment. Overall, most teachers provided an adequate classroom environment in which to engage students, however, there were some teachers that demonstrated more effective teacher practices or provided an overall high quality learning environment than others. For example, one treatment teacher in particular (T1) demonstrated a high level of effective teacher practices as well as student engagement, and therefore was considered among the most effective teachers in the study. Overall, treatment and control teachers were comparable on the quality of classroom environment provided to students. There appeared to be much more consistency related to overall quality of overall classroom

environment within each site as opposed to across conditions. We cannot emphasize enough the importance of each school's specific culture when considering the activities that occur in classrooms.

SECTION FIVE: RESULTS RELATED TO ACHIEVEMENT IN BIOLOGY AND STUDENTS' ATTITUDES TOWARDS SCIENCE

In this section, we report on three of the four research questions involving student outcomes: (1) How do student outcomes differ for students using the Prentice Hall *Biology* program compared to control students?; (2) How do student outcomes differ for treatment and control students as a function of their background characteristics (e.g., gender, ethnicity, etc.); and (3) How does implementation relate to student achievement? The analytical techniques we used to investigate these questions allowed us to address the first two questions simultaneously, so the discussion pertaining to those two questions will be combined.

Descriptive Statistics on Students' Achievement in Biology

In this study, students' achievement in Biology was defined by combining the results from the 30 enhanced multiple choice (EMC) questions (one point each) and the one constructed response investigation (CRI) (which was worth 12 points total). The EMC questions were designed to assess students' understanding of important scientific facts, concepts, principles (declarative knowledge), and to probe scientific reasoning skills (procedural knowledge). The CRI item was designed to have students analyze a problem, conduct a secondary analysis, revise a hypothesis, and/or recommend solutions. Percent correct for the EMC and CRI were calculated and then averaged to obtain students' overall science achievement score (total possible score was 100%). This method was chosen so that equal weight was given to Biology content material (EMC) as well as the scientific process (CRI). These achievement outcomes were compared for several different student-level subgroups, including: (1) gender (males or females); (2) ethnicity (i.e., Caucasian, Latino, and African-American); (3) English as a primary language (i.e., yes or no); and grade level (i.e, 9^{th}, 10^{th}, or 11^{th}).

Student-level descriptive statistics are presented in Table 11. Students were included in this table only if they had matched pretest and posttest scores and there were more than 10 students in a given subgroup. This excluded five subgroups of students (i.e., 12^{th} graders, special education students from one period, students of Asian descent, mixed race students, and students with "other" ethnicity). On average, treatment students' mean percent correct on the assessment was 56% at pretest and 58% at posttest, while control students scored 57% correct at pretest and 59% correct

posttest. Every sub-group in the treatment and control group increased their scores from pretest to posttest, except for one group of African-American students in the treatment group. That group actually decreased their achievement score by approximately one point from pretest to posttest. Latino students in the treatment group and students whose primary language was not English showed the largest gain from pretest to posttest, with each group increasing 5.6 points on average. Ninth grade students in the control group also had a notable gain of 4.3 points from pretest to posttest.

Table 11. Matched Biology Achievement Pretest and Posttest Scores for Treatment and Control Groups as a Function of Student Demographic Characteristics.

Demographic Characteristics		Treatment		Control	
		Biology Pretest	Biology Posttest	Biology Pretest	Biology Posttest
Grade Level	9th Grade	58.93 (14.18)	61.64 (14.42)	58.00 (14.74)	62.28 (14.38)
	10th Grade	55.54 (15.25)	57.23 (15.28)	57.28 (15.60)	59.08 (16.40)
	11th Grade	51.80 (15.69)	54.96 (18.91)	51.72 (17.44)	52.06 (14.47)
Gender	Male	56.27 (15.47)	57.06 (14.78)	57.41 (15.61)	60.07 (15.76)
	Female	56.14 (14.87)	59.27 (15.83)	57.08 (15.23)	59.71 (15.71)
Ethnicity	Caucasian	60.70 (13.72)	60.76 (13.56)	60.09 (14.44)	63.44 (14.36)
	Latino	53.27 (14.67)	58.92 (14.70)	54.59 (15.08)	57.22 (15.27)
	African-American	49.57 (15.36)	48.55 (18.03)	49.23 (16.26)	50.18 (16.96)
Primary Language	English	56.77 (15.13)	58.12 (15.71)	57.71 (15.77)	60.65 (15.68)
	Other	53.22 (15.00)	58.89 (13.71)	54.33 (11.93)	54.98 (14.64)

Note: Subgroups with fewer than 10 cases were not included.

These data tentatively suggest that Latino students and non-Native English speakers in the treatment group may have benefited more from the curriculum than the control students within these subgroups. However, since student data was nested within

different classrooms and different schools, we used Hierarchical Linear Modeling (HLM) to test for statistically significant differences among these groups.

Differences in Biology Achievement Between Treatment and Control Groups and Across Student Demographics

Given that we randomly assigned at the teacher level and students were nested within different classrooms (i.e., non-random assignment of students into different classrooms), we used HLM to examine differences in achievement by condition (treatment and control) and by student background characteristics (e.g., gender, ethnicity, etc.). 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., teacher academic background) on students' Biology achievement. For a complete discussion of the rational and theory underlying HLM models, please see Byrk and Raudenbush (1992).

Appendix B lists the Level 1 and Level 2 variables (and their operational definitions) associated with student background characteristics and teacher/ classroom/ school characteristics that were explored prior to building the final HLM model (see Appendix C). All level 1 and level 2 variables were dichotomously coded except for three level 1 continuous variables (i.e., overall GPA last year, percent days present in school, and pretest score) and two level 2 variables (i.e., years of teaching experience and years teaching Biology). In an effort to build a parsimonious HLM model, univariate analyses were conducted comparing all level 1 and level 2 factors to the outcome variable (Biology posttest scores). Most of these factors (except for Primary Language, GPA, Biology pretest, and ethnicity) were not significantly related to the outcome, so they were excluded from the final HLM model. Additional variables, including the student survey composite variables and the two proxy indicators of student treatment (i.e., percent of the days when students were present in class and the number of months that the curriculum was implemented) may have been linearly related to students' posttest scores, however, after controlling for students' demographic and prior academic background characteristics, none of these variables remained significant.

Table 12 illustrates the descriptive statistics for the variables (five level 1 and two level 2) in the final HLM model. Table 13 details the results of the final HLM model.

Table 12. Descriptive Statistics of Variables in the HLM Model.

	Level 1 Descriptive Statistics				
Variable Name	**N**	**Mean**	**SD**	**Minimum**	**Maximum**
Primary Language (LANG)	1314	0.14	0.35	0.00	1.00
GPA Indicator (GPA)	795	3.02	0.73	0.17	4.00
Pretest Indicator (PRETEST)	1195	56.16	15.61	11.67	91.67
African-American Indicator (AFRICAN-AMERICAN)	1401	0.17	0.38	0.00	1.00
Latino Indicator (LATINO)	1401	0.31	0.46	0.00	1.00
	Level 2 Descriptive Statistics				
Treatment Condition Indicator (CONDIT)	45	0.51	0.51	0.00	1.00
Master's Degree Indicator (MASTER)	45	0.44	0.50	0.00	1.00

As shown in Table 13, the results from the HLM analyses revealed no overall difference in Biology achievement between treatment and control groups, after controlling for all other factors in the model (*coefficient*, 0.16; *s.e.*, 1.34). However, when outcomes were examined by subgroup, there were significant differences in achievement by ethnicity and primary language. For ethnicity, Latino students in the treatment group had significantly higher Biology achievement scores than Latino students in the control group by 4.32 points ($p = .052$). In addition, the cross-level interaction between the Latino indicator variable (student level variable) and condition (a classroom-level variable) suggested that Latino students in the treatment group also outperformed Caucasian students in the treatment group by roughly 2.7 points (i.e., -1.6 + 4.32 = 2.72 on a 0-100 scale), which was marginally significant (4.32, $p = .052$). [3]

For primary language, students whose primary language was English performed significantly better than students whose primary language was not English (roughly 3.9 points on a 0-100 scale, $p=0.035$). However, non-native English speakers did better when the group variable was taken into account. That is, the gap between non-native

[3] With robust standard errors, $p = .001$).

English speakers and English speakers was reduced to less than 1 point for the treatment group (i.e., -3.94 + 3.41= -0.53). Furthermore, the positive regression coefficient suggested that the non-native English speakers in the treatment group, on average, outperformed the non-native English speakers in the control group by about 3.41 points. Although this result was not statistically significant, it does suggest that Miller and Levine's curriculum may positively impact non-native English speakers.

Table 13. Results of the HLM Analysis.

Fixed Effect		Coefficient	Standard Error	Approx. T-Ratio	df	p-value
For INTRCPT1, β_{0j}	*INTRCPT2,* γ_{00}	55.24	1.68	32.97	42	0.000
	CONDIT, γ_{01}	0.16	1.34	0.12	42	0.907
	MASTER, γ_{02}	5.32	1.74	3.06	42	0.004
For LANG slope, β_{1j}	*INTRCPT2,* γ_{10}	-3.94	1.87	-2.11	649	0.035
	CONDIT, γ_{11}	3.41	2.60	1.32	649	0.189
For GPA slope, β_{2j}	*INTRCPT2,* γ_{20}	2.82	0.71	3.99	649	0.000
For PRETEST slope, β_{3j}	*INTRCPT2,* γ_{30}	0.57	0.03	17.75	649	0.000
For AFRICAN-AMERICAN slope, β_{4j}	*INTRCPT2,* γ_{40}	-3.71	3.03	-1.23	649	0.221
	CONDIT, γ_{41}	-0.48	3.71	-0.13	649	0.897
For LATINO/A slope, β_{5j}	*INTRCPT2,* γ_{50}	-1.60	1.47	-1.09	649	0.276
	CONDIT, γ_{51}	4.32	2.22	1.94	649	0.052

Differences in outcomes were also reported by certain teacher characteristics. Everything being equal, students in classes where the teacher had a Master's Degree did significantly better than students in classes where the teacher had a Bachelor's Degree (roughly 5.3 points on a 0-100 scale, $p = .004$). Given that out of the 16 teachers, nine held a Master's Degree (four in the treatment and four in the control), it was plausible that this teacher effect could potentially explain the Latino student

treatment effect noted above. However, when we examined the cross-level interaction between the Latino indicator and the Master's Degree indicator, no significant interaction was revealed. This means that the Latino student treatment effect could not be explained by being in classrooms with more highly educated teachers. There were also no other significant teacher, classroom, or school level effects.

Taken together, these results suggest that Latino students who used Miller and Levine's (2006) *Biology* curriculum outperformed Latino students who used a different curriculum. Furthermore, although not statistically significant, non-native English speakers also seemed to benefit from using Miller and Levine's text. It is also important to note that teacher's level of education impacted students' achievement in Biology; students in classes with more highly educated teachers (i.e., Master's Degree) outperformed students in classes with less educated teachers (i.e., only a Bachelor's Degree).

Changes in Student Attitudes as a Function of Condition, Primary Language Status, and Ethnicity

Responses from the student attitude survey (pretest and posttest) were also integrated into the HLM model to determine whether there was an effect of condition (treatment or control) on students' attitudes from pretest to posttest. An exploratory factor analysis (EFA) was conducted on students' posttest responses so that factor composites could be derived. Using a principal axis extraction method and a varimax rotation (for ease of interpretation of factor loadings), six factors were revealed. Composite variables were then constructed by taking an average of the items loaded on each factor. Although the final HLM model did not include the affective composites (because they were not significantly related to Biology achievement), it is important to examine changes in students' attitudes as there could be subtle, yet meaningful differences between the groups.

An Analysis of Covariance (ANCOVA) was conducted to examine differences in students' attitudes at posttest (controlling for pretest) as a function of condition (treatment vs. control), English as a Primary Language (yes vs. no), and ethnicity (Caucasian, Latino, Asian, and African American). Unfortunately, there were no significant changes in students' attitudes as a function of any of these variables. Across

condition, ethnicity, and English as a primary language, students all increased (or decreased) approximately the same amount from pretest to posttest.

The Relation Between Curriculum Implementation and Student Achievement

The next section addresses our third research question: What is the relationship between curriculum implementation and student achievement in science? The question will be addressed by combining a number of implementation measures in comparison to student achievement scores. It is particularly important to examine differences in student achievement as a function of implementation given that there were no overall differences in achievement across treatment and control groups. Since aggregate data often masks meaningful differences by teacher or classroom, we analyzed how students' test scores related to teachers' reported level of implementation (including student ratings of the composite "Teacher Influence" and overall classroom environment ratings).

Teachers were first ranked according to increases in student achievement scores. This analysis involved combining multiple classes from the same teacher into one mean change score. Although classroom level mean change scores varied widely within teacher (e.g., classes ranged from -6.44 to 15.18, or 6.63 overall), an overall mean change score was derived for all teachers. All teachers appear in ranked order according to this variable (see Table 14). Three implementation variables were also included in Table 14. The first was a mean rating of "Teacher Influence" taken from the posttest student survey. Second, we included the number of sections covered by each teacher during the year. A rating was established for each teacher based on coverage relative to other teachers in the study. For example, teachers with coverage of sections significantly greater than the median of (26 sections) were considered "high"; those with number of sections at or near the median were considered "average"; and those with number of sections covered significantly lower than the median were considered "low". Finally, we included a rating for the overall classroom environment based on survey data and classroom observations that ranged from "Excellent" to "Poor". The final column of Table 14 indicates the treatment condition for each teacher.

In general, the highest ranked teachers were evenly distributed between both groups; however, four of the top five teachers were in the treatment group. Four of the bottom five teachers were also in the treatment group. Overall, there was a general trend

in which there were more similarities within each site than between each site. For example, all treatment and control teachers at one site were ranked in the top 50% of the distribution, while none of the teachers at the another site were ranked in the top 50% and three of the four teachers at the same site were ranked in the bottom 50% of the distribution. We have reported previously that the composite factor of "Teacher Influence" was significantly related to student achievement, such that students who rated their teachers higher on teacher influence also performed better on the Biology posttest ($r = .103$, $p < .01$). This trend can also be observed in Table 14, given that student ratings of their teachers generally corresponded to their mean increases from the Biology pretest to posttest.

Table 14. Ranking of Teachers Based on Student Posttest Scores and Teacher Influence.

Teacher Rank	Overall Student Mean Change Score	Mean Rating of "Teacher Influence"	Curriculum Coverage	Overall Classroom Environment	Condition
1	8.73	3.8	Low	Fair	Control
2	7.04	3.9	High	Excellent	Treatment
3	6.63	3.7	High	Fair	Treatment
4	5.64	3.9	Average	Good	Treatment
5	5.17	3.4	High	Good	Treatment
6	3.67	3.6	Average	Good	Control
7	2.65	3.9	High	Very Good	Control
8	1.90	4.0	Average	Fair	Control
9	1.83	3.8	Average	Fair	Control
10	.81	3.7	Average	Very Good	Control
11	.60	3.8	Average	Fair	Control
12	0.00	4.1	Average	Excellent	Treatment
13	-.34	3.5	Low	Good	Treatment
14	-1.20	3.3	Low	Poor	Control
15	-4.30	3.5	High	Poor	Treatment
16	-4.97	3.6	High	Good	Treatment

For curriculum coverage there were generally no trends that aligned with the rankings—teachers with high coverage were observed in both the highest and lowest rankings, although generally speaking, teachers with the lowest coverage had student scores that were generally lower (with the exception of the teacher ranked #1). Also, classroom environments rated as “Poor” generally had lower student change scores relative to other teacher groups, but generally teacher ratings of classroom environment were not necessarily related to mean change scores. This is in part due to the variability in classes for each teacher. We generally observed different levels of classroom environment for each teacher class period, but were forced to collapse these overall ratings to produce a general rating associated with one teacher. Given that this approach generally reflects a qualitative case study methodology as opposed to a more specific quantitative analysis, we interpret these data with caution. Although we are confident in our ranking of teachers across these dimensions, further study is required on a much larger scale to investigate the extent to which implementation can predict student achievement outcomes.

SECTION SIX: PRODUCT SATISFACTION

In this section, we address our fourth research question: How satisfied are treatment teachers and students with their Prentice Hall textbook and ancillary materials? To answer this question, product satisfaction questions pertaining to the textbook and ancillary materials were included on the teacher survey and the posttest student survey. Given there were only 16 teachers across the treatment and control groups, we chose to focus on treatment teachers' satisfaction with the product rather than compare teacher satisfaction by treatment and control. However, we did compare students' perceived product satisfaction by condition, as those analyses were more appropriate given the size of our student sample. When necessary, qualitative responses taken from the teacher interviews[4] and open-ended student survey questions were included to provide support for teacher and/or student ratings. Finally, we provide suggestions for improving the textbook and/or ancillary materials based on treatment teacher and student feedback.

Perceived Product Satisfaction According to Treatment Teachers

Teachers provided satisfaction ratings on: (1) the overall teacher's edition, student's edition, workbooks, ancillaries, and technology; and (2) the frequency of use, quality, and feasibility of using specific components of the textbook (e.g., key concepts, inquiry activities, checkpoints, etc.); and (3) comparisons of Prentice Hall *Biology* to other competitor textbooks on several quality dimensions.

Satisfaction of Textbook, Workbooks, Ancillary Materials, and Technology

As shown in Table 15, treatment teachers rated all aspects of the Biology curriculum (including the textbook, workbooks and ancillaries, and all technological components) on a five point scale ranging from "Poor" to "Excellent." Overall, all teachers were very satisfied with the textbooks; all teachers rated the Teacher's Edition and 87% of teachers rated the Student Edition as "Excellent." One hundred percent of teachers rated the Guided Reading and Study Workbook "Excellent", and 87% gave that rating to the Adapted Reading and Study Workbook. Based on the teacher interviews, teachers appreciated the leveled workbooks and commented that the ancillary materials complemented one another well. As reported previously, the various levels of workbooks allowed them to better differentiate instruction for students.

[4] For a complete transcripts of teacher interviews, please contact the principal investigator.

Table 15. Satisfaction with All Aspects of the Prentice Hall *Biology* Curriculum Components

Components of the Prentice Hall *Biology* Curriculum	Number of Teachers Reporting Using Curriculum Components	Number of Teachers Responding "Excellent" (Percent "Excellent" of those who responded)
Textbook		
Teacher's Edition	8	8 (100%)
Student's Edition	8	7 (87.5%)
Workbooks and Ancillaries		
Guided Reading and Study Workbook	7	7 (100%)
Adapted Reading and Study Workbook	8	7 (87.5%)
Lab Manual A	8	5 (62.5%)
Issues and Decision Making Workbook	2	1 (50.0%)
Standardized Test Prep Workbook	1	1 (100.0%)
Summary and Vocabulary Review Workbook	2	1 (50.0%)
Technological Components		
Teacher Express	8	5 (62.5%)
i-Text CD	7	3 (42.8%)
i-Text Online	7	6 (85.7%)
www.phschool.com	7	5 (71.4%)
Transparencies	7	7 (100%)
Active Art	5	3 (60.0%)
www.millerandlevine.com	4	1 (25.0%)
BioDetectives	2	1 (50.0%)
Bio-Technology Manual	1	1 (50.0%)

Few teachers used Issues and Decision Making (n=2), Standardized Test Prep (n=1), or the Summary and Vocabulary Review (n=2), however, of the teachers who did

use the workbooks, at least half rated them as "Excellent." Regarding the technological components, teachers rated the I-text online and the transparencies positively. Sixty-two percent of teachers rated *Teacher Express* as "Excellent." This rating was supported by multiple teacher interviews that suggested that *Teacher Express* and, in particular, the *Exam View* test bank, saved them time and made planning more efficient. Of the two relevant websites, *phschool.com* was used more frequently and received higher ratings than *millerandlevine.com.* The lowest rating went to the I-text CD-ROM; teachers reported that their students had difficulties installing the programs on computers running Windows XP.

Specific Textbook Components: Frequency of Use, Quality, and Feasibility of Use

Overall, teachers rated the teacher and student edition of Prentice Hall Biology positively. To determine which specific aspects of the textbook may have related to these high marks, we asked teachers to rate each component of the textbook (e.g., key concepts, inquiry activity, checkpoints, etc.) along three dimensions. The first dimension was the frequency with which the component was used (rated on a 1-5 point scale "Never", "Rarely" (once per month), "Sometimes" (twice per month), "Frequently" (once per week), "Always" (everyday)). The second dimension was an overall rating of quality from "Poor" to "Excellent". The final dimension was on feasibility (or practicality) of use for a particular component, rated on a five-point scale from "Strongly Disagree" to "Strongly Agree". Table 16 displays the group means by textbook component along each dimension.

This table suggests that most components of the textbook was regularly used, except for the standardized test prep, extended labs, and quick labs which were used approximately once per month. Key concepts were covered most frequently, as were the caption questions and section assessments. All of these components were required elements on the implementation guidelines, so it was understandable that they were used so frequently. Teachers also commented that they frequently used (and relied on) teacher edition side notes that suggested ways of "Making connections", "Addressing misconceptions", etc.

Perceived quality of all components were rated "Good" to "Excellent", however, the key concepts and section assessments were rated the highest. In the interviews, teachers also reported that the key concepts were excellent vehicles to check for student understanding. Teachers also particularly liked the bold vocabulary words because they made terms easier for students to find in the text.

Table 16. Frequency of Use, Quality, and Feasibility of Use of Specific Components of the Textbook.

Specific Components of Textbook	Frequency of Use	Quality	Feasibility of Use
Key Concepts	4.5	4.6	5.0
Inquiry Activity	3.5	4.0	3.9
Checkpoints	4.0	4.5	4.9
Caption Questions	3.8	4.4	4.5
Analyzing Data	3.3	4.4	4.8
Section Assessment	3.9	4.6	4.6
End of Chapter Assessments	3.1	4.5	4.4
Standardized Test prep	2.5	4.4	4.4
Quick Labs	3.0	4.0	4.6
Extended or Full Labs	2.7	4.1	4.4

The feasibility of using each component differed slightly by component. Teachers reported that key concepts, checkpoints, and analyzing data were easy to implement in the classroom, however, the feasibility of the inquiry activity was rated the lowest (although the 3.9 rating suggests that it was feasible, just not as feasible as some of the other components). Teachers commented that although they liked the inquiry activities, many of them were too long and required too much time to implement fully in the classroom.

A Comparison of Prentice Hall Biology to the Competition

The treatment teachers were asked to rate the 2006 Prentice Hall *Biology* textbook in comparison with the textbooks they have used in previous years. Although teachers were instructed to compare to the text they used in the previous year, it was possible that comparisons were made to all previous Biology books they used previously. Teachers provided one of three responses (i.e., Prentice Hall is not as a good as competitor, Prentice Hall is the same as competitor, or Prentice Hall is better than competitor) across seven dimensions of quality.

Table 17 shows that out of the eight treatment teachers, seven (87%) rated Prentice Hall *Biology* better than the competitor curricula across four of the seven

dimension of quality, including reading level of text, graphs, pictures, student understanding of concepts, and breadth of coverage. Six (75%) rated depth of coverage and student enjoyment of the text as better than the competitor. The lowest rated item was "encouraging students' critical thinking skills" given that only five teachers (62%) rated this aspect as superior to competitor curricula.

Table 17. Comparison of 2006 Prentice Hall Biology to Competitor Textbooks

Dimension of Quality	Percent of Teachers Reporting "Better than Competitor" (n=8)
Reading level of text	87.5%
Graphs, pictures, etc.	87.5%
Student understanding of biology concepts	87.5%
Breadth of coverage of key biology concepts	87.5%
Depth of coverage of key biology concepts	75.0%
Student enjoyment of text	75.0%
Encourage students' critical thinking skills	62.0%

Taken together, treatment teachers were extremely satisfied with the Prentice Hall *Biology* textbook, workbooks, and ancillary materials. Many of the textbook components were used frequently, were easy to use, and were perceived as high quality. Teachers rated the *Biology* textbook superior to many of its' competitors on several dimensions, including reading level of text, graphics, and both breadth and depth of coverage. These findings support the fact that all treatment teachers reported they would "Strongly Recommend" the *Biology* text to other Biology teachers.

Differences in Students' Ratings of Product Satisfaction by Condition

Students' overall product satisfaction was measured on the posttest student survey. To enable comparisons between treatment and control students, general aspects of the treatment and competitor texts were identified rather than specifically naming components of each of the four textbooks used in this study.

Eleven independent t-tests were conducted comparing student satisfaction (for all students with available survey posttest scores) in the treatment group to the control group. Given the large number of comparisons, a stricter criterion for significance ($p < .001$) was used to reduce the chance of a Type I error. As Table 18 indicates, treatment students rated five of the 11 product satisfaction questions significantly more positively than control students. Compared to control students, treatment students reported: (1) it was easier to locate the main points of paragraphs; (2) it was less boring to read; (3) they were able to grasp the concepts without having to read the text multiple times; (4) they learned a great deal from their Biology textbook; and (5) the overall layout of the textbook was superior.

Table 18. Textbook Product Satisfaction as a Function of Condition

Overall Textbook Product Satisfaction	Treatment Mean (SD) n = 552	Control Mean (SD) n = 587	p-value
My Biology textbook explained difficult concepts in everyday language.*	3.41 (.94)	3.30 (.88)	.030
I had to read sections of my Biology textbook multiple times before I understood the concepts.*	3.10 (.98)	3.29 (.93)	.001
In my Biology textbook, it was easy to locate the main points of each paragraph.*	3.75 (.87)	3.51 (.89)	.000
My Biology textbook was boring to read.*	3.11 (1.04)	3.37 (1.03)	.000
I have learned a great deal from my Biology textbook.*	3.55 (.84)	3.33 (.83)	.000
I found myself sharing information found in this text with others not involved in the course.*	2.90 (1.12)	2.88 (1.11)	.800
Pictures in the book**	4.14 (.76)	4.12 (.80)	.648
Diagrams in the book**	4.10 (.85)	4.13 (.81)	.536
Experiments and labs**	4.01 (.91)	3.96 (.96)	.412
End of Chapter questions in the book**	3.59 (1.07)	3.47 (1.11)	.077
Overall layout of your Biology textbook**	3.94 (.90)	3.71 (.89)	.000

* Rated on a five point scale from “Strongly Disagree” to “Strongly Agree.”
**Rated on a five point scale from “Poor” to “Excellent.”

Treatment students confirmed these responses by answering the open-ended questions about the quality of their textbook (i.e., what they liked best and least). One-

fourth of the students responded that the content material was explained well, in an easy to understand format, and was full of helpful details. One student commented that the "information is written in a way that I can understand." Students also reported liking the vocabulary words and main concepts that were in bold print. A student commented, "Once you understand the main point, you get the whole section." Another student said, "the vocabulary words and main points are easy to find and help me study for tests and find answers to my homework." Students also commented on liking the pictures and diagrams in the book; one student reported "the pictures were helpful, I understand what the text is saying with a visual example" while another student commented, "the pictures make it easier to understand the concepts." Finally, the other satisfying component of the text for treatment students was the sheer number of interesting facts in the book, the extensive glossary, and that the labs and checkpoints helped students prepare for tests.

Topics Noted by Teachers and Students as "Needing Improvement"

Throughout the duration of the study, teachers and students were offered several opportunities to comment on what components of the program could be improved. The following summarizes the gist of the issues/challenges that teachers and students expressed to the researchers (either in writing or verbally).

- There were technical problems with I-text on CD rom because it was not compatible with Windows XP. However, after several attempts at loading the software, teachers identified that they had to change properties in the operating system to make the CD work.
- Not all electronic resources (in *Teacher Express*) match the printed versions of the workbooks, labs, etc. so students may have had two different versions depending on how the information was accessed. This created problems when assigning homework from both the electronic versions and hard copies.
- Some online Active Art links were not working, which was frustrating for teachers and students.
- Two teachers commented that the pages in the textbook and workbooks were very thin and would most likely not hold up over time. Several students accidentally tore pages in these resources throughout the year.
- Laboratory Instructions were confusing and tedious.

- Laboratory activities actually took longer than the time allocated in the textbook.
- Some students found the reading level of the text too difficult. Some students commented that the vocabulary words were too hard to understand and/or pronounce. Similarly, some thought that concepts were difficult to grasp or that the labs were hard to understand.
- Many of the students reported that the book was too heavy which made it difficult to carry back and forth from school.
- A few students thought the textbook emphasized the theory of evolution too much.

SECTION SEVEN: DISCUSSION

The purpose of the current study was to conduct a scientifically rigorous RCT study on the effects of the Prentice Hall *Biology* curriculum on student attitudes and interest in science as well as achievement in Biology. Teachers were randomly assigned at each study site to either treatment or control groups within five different schools across four states. Multiple process measures were collected throughout the year (online implementation logs, classroom observations, teacher interviews, and teacher surveys) and two outcome measures (Biology achievement assessment and student attitudes towards science) were collected during fall 2005 and, again, in spring 2006. Our primary research questions allowed us to investigate the impact of the curriculum on student achievement (and as a function of student characteristics), understand the relationship between implementation and achievement, and finally, assess product satisfaction among treatment teachers and students.

According to the HLM results, there was no overall difference in student Biology achievement across treatment and control groups, after controlling for students' demographics (e.g., ethnicity) and their prior academic background (i.e., GPA and pretest). Students in both conditions increased approximately the same amount (about two points) from pretest to posttest. One possible explanation for this effect was that the content and organization of treatment text (Miller & Levine) *Biology* did not appear different enough from each of the control texts to produce observable differences in student achievement across condition. Content alignment of the four textbooks suggested that each covered nearly identical content. Furthermore, an analysis of curriculum implementation and student achievement also revealed that treatment and control groups were equal on the level of curriculum implementation, ratings of teacher influence from their students, student engagement, and overall classroom environment.

However, there were reported differences between the groups on how instruction was differentiated for students. One of the key features that distinguished the Prentice Hall curriculum from its competitors was an emphasis on differentiated instruction (i.e., engaging students of different ability levels, language status, etc.). Although control teachers sought out materials to differentiate instruction to ELL students (and students of different ability levels), the level of differentiated instruction within the treatment group was much more frequent and was higher quality than what was observed in the control group. Although we cannot claim that differentiated instruction in the treatment group

(e.g., leveled workbooks, differentiated instruction techniques in the text, etc.) *caused* differences in student achievement, this implementation finding may partly explain why Latino students in the treatment group had significantly higher Biology posttest scores than Latino students in the control group, after controlling for all other variables in the HLM model. In addition, although it was not statistically significant, the data suggested that students whose primary language was not English also may have benefited from participating in the treatment condition more than those students did in the control group. To understand this effect, we examined student outcomes by condition (and by ethnicity) within the site containing the majority of the Latino students (Site 1—CA). Latino students in the treatment group outperformed Latino students in the control group at Site 1—CA, and based on the teacher interviews, treatment teachers at that site had higher quality differentiated instruction than control teachers. Perhaps future research should investigate the mechanisms underlying how teachers using the Prentice Hall *Biology* curriculum specifically engaged students with varying characteristics in the learning process.

Results regarding product satisfaction revealed that treatment teachers and students were extremely satisfied with their Prentice Hall *Biology* materials. Treatment students rated aspects of the curriculum more favorably than control students, especially with respect to the overall layout of the text, their engagement by the text, and their ability to identify the main points of paragraphs. Treatment teachers and students rated the pictures and graphics as outstanding, and all treatment teachers said they would strongly recommend the text to other Biology teachers.

Although this study provided the strongest empirical test of the impact of the Prentice Hall Miller & Levine (2006) *Biology* curriculum on student achievement, there were two primary limitations of the current study. First, the generalizability of the study was compromised given that schools were not randomly selected out of all eligible schools. Out of the several hundred recruitment calls that were made to schools across the country, these five schools were the only ones who were interested in participating and who agreed to abide by all research protocols. The results of this study only generalize to schools that share similar demographics.

The second limitation concerned our conceptualization of the research design. Evaluation scholars (Rossi, Freeman, & Lipsey, 1999) suggest that impact studies should only be conducted once the program (or intervention) is established, free from implementation errors, and has formative evidence that outcomes are reasonable and

obtainable. Given that we asked teachers in the treatment group to become familiar with the textbook and corresponding ancillary materials in a relatively short time, one could argue that implementation of the curriculum in the treatment group did not satisfy the conditions for conducting an impact study. Treatment teachers received the initial product training just prior to the start of the school year, so they had little time to prepare new lesson plans and get comfortable using the curriculum and ancillaries prior to school starting. Plus, treatment teachers were prevented from supplementing the curriculum with outside resources; all course material needed to be taken from the textbook and ancillaries, so all previous resources that teachers implemented well and felt comfortable with had to be discarded for the year they participated in the study.

Conversely, control teachers had the flexibility to supplement the text if it did not adequately cover a specified topic. Teachers knew from experience what parts of the curriculum worked or did not work well for students and had prepared lesson plans that could be adapted (rather than prepared for the first time). Although most treatment teachers reported feeling confident and prepared to implement the curriculum within the first month of school, and teachers in both conditions were comparable in their teaching experience, level of education, and content expertise, it was fundamentally unfair to compare curriculum being implemented the first time to curriculum that has been used in previous years. A more appropriate comparison would be to randomly assign teachers to two or three different conditions where new curricula is being used in all groups, and then systematically compare based on this. Alternatively, treatment teachers could be provided with new curricula the year prior to officially participating in a research study. Teachers would be given all materials with the goal of increasing their comfort and experience with the materials; no restrictions would be placed on how they used the curriculum. The following year, estimates of the effectiveness of the treatment curriculum on student achievement could be determined.

While we are confident in the present research findings, we feel that future research should be conducted to investigate the impact of the Biology curriculum on students. For example, the publisher should consider a follow-up study of all teachers who have participated in the current study to assess changes in the use of the curriculum from year to year. The study should be expanded to investigate the impact on a larger number of teachers and schools in order to conduct more specific analyses that extend beyond only eight treatment teachers. Also, a more specific content analysis should be conducted assessing the quality of lessons for students in the textbook and

ancillary materials, particularly with regard to topics such as stem cells and the theory of evolution.

REFERENCES

Bloom, B.S. (Ed). (1956). Taxonomy of educational objectives: The classification of Educational goals. *Handbook I: Cognitive Domain.* New York: Longmans/Green.

Bredderman, T. (1983). Effects of activity-based elemtnary science on student outcomes: A quantitative synthesis. *Review of Educational Research, 53*, 499-518.

Byrk, A., & Raudenbush, S. (1992). *Hierarchical Linear Models*. Sage Publications: Thousand Oaks, CA.

Carr, M., Kurtz, B.E., Schneider, W., Turner, L.A., & Borkowski, J.G. (1989). Strategy acquisition and transfer among German and American children: Environmental influences on metacognitive development. *Developmental Psychology, 25*, 765-771.

Close, D. (1996). *National standards and benchmarks in science education: A primer*. ERIC Digest: SE 058913.

Comfort, K.B. (1994). Authentic Assessment: A Systemic Approach in California. *Science and Children*, 32(2), 42-43, 65-66.

Davies, R.S. , Williams, D.D., & Yanchar, S. (2004, November). *The use of randomization in educational research and evaluation: A critical analysis of underlying assumptions.* Paper presented at the annual meeting of the American Evaluation Association: Atlanta, GA.

Fan, X., & Chen, M. (2001). Parental involvement and students' academic achievement: A meta-analysis. *Educational Psychology Review, 13,* 1-22.

Flavel, J.H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*, 906-911.

Flink, C., Boggiano, A.K., Main, D.S., Barrett, M., & Katz, P.A. (1992). Children's achievement-related behaviors: The role of extrinsic and intrinsic motivational orientations. In A.K. Boggiano and T.S. Pittman (Eds.), *Achievement and motivation: A social-developmental perspective.* Cambridge: University Press.

Garner, R. (1987). *Metacognition and reading comprehension*. Norwood, NJ: Ablex.

Klein, S., Stecher, B., McCaffrey, D., Jovanovic, J., Shavelson, R.,Haertel, E., Solano Flores, G., and Comfort, K. (1997) Gender and Racial/Ethnic Differences on Performance Assessments in Science. *Educational Evaluation and Policy Analysis*, 19(2), 83-97.

McCaffrey, D., Hamilton, L., and Aronson, E. (1998) *Scaling and Equating of the California Systemic Initiatives Assessment Collaborative Standards-based Assessments for 1996 and 1997:* Unpublished Technical Report, RAND: Santa Monica, CA

National Research Council (2000). Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). *How people learn: Brain, mind, experience, and school.* Washington, D.C.: National Academy Press.

Rossi, P. H., Freeman, H. E., Lipsey, M. W. (1999). *Evaluation: A systematic approach (Sixth Edition).* Sage Publications: Thousand Oaks, CA.

Schunk, D.H., & Pajares, F. (2001). The development of academic self-efficacy. In A. Wigfield & J. Eccles (Eds.), *Development of achievement motivation.*San Diego: Academic Press.

Tomlinson, C. A. (1999). *The differentiated classroom: Responding to the needs of all learners.* Association for Supervision and Curriculum Development: Alexandria, VA.

U.S. Department of Education (2005). Federal register: Scientifically based Evaluation Methods. Volume 70, No. 15.

Weinstein, R., Boulanger, D., & Walberg, H.J. (1982). Science curriculum effects in high school: A quantitative synthesis. *Journal of Research in Science Teaching, 19*, 511-522.

Wilson, M., Delgado, J., and Finkelstein, D. (1998) *Equating and Scaling of the CSIAC Standards-based Science Assessments 1997 and 1998.* Unpublished Technical Report, Educational Data Systems: Morgan Hill, CA.

Appendix A: Prentice Hall *Biology* Implementation Guidelines

Revised Implementation Guidelines For Use in the RCT (Miller & Levine, 2006):

Each teacher using Prentice Hall *Biology* should have the minimum in place for program implementation:

Required:

- Primary use of Miller & Levine's (2006) Prentice Hall *Biology* text for reading assignments
- Cover first five units of text (1-Nature of Life, 2-Ecology, 3-Cells, 4-Genetics, 5-Evolution) –may skip chapters within these units
- Should cover 1-2 additional units (6-10)
- Use of Guided Reading and Study Workbook or Adapted Reading and Study Workbook (each student should have a copy of both)
- Cover all "Key Concepts" in chapter that apply to state standards
- Complete all end of chapter assessments for those applicable to state standards
- One lab per unit (not necessarily from Prentice Hall textbook)

Strongly recommended:

- Complete all "Section assessments" for those sections applicable to state standards
- Cover some "Caption questions" throughout the chapters
- Frequent use of transparencies or presentation pro

Optional:

- Assign at least one writing assignment at least once every two weeks, preferably "Writing in Science" response
- Perform "Analyzing Data" exercises where applicable
- Writing in science (e.g., "Issues in Decision making")
- Standardized test prep (recommended in adoption states)
- Other Prentice Hall technology

Appendix B. Key Variables Explored Prior to Building the HLM Model

Outcome variables:

1. Overall pretest scores (proxy for prior academic achievement)
2. Overall posttest scores

Both calculated based on the average of percent correct multiple choice and constructed response items.

Student background characteristics variables:

1. Gender: female (0) and male (1)
2. Hours spent completing science homework on a normal school day
3. English as primary language: yes (0) no (1)
4. Overall GPA last year (proxy for prior academic achievement)
5. Percent days present
6. African American: yes (1) no (0)
7. Latino: yes (1) no (0)
8. pre- and post- affective composite variables (see pages 27-28 of the report for a description of the factors derived from an EFA on the student survey posttest responses).

Teacher/classroom/school characteristics variables:

1. Condition: indicator of treatment (1) or control (0) group
2. Years of teaching experience
3. Years of teaching biology
4. Percent Latino: low (0) and high (1) concentration—low where a school has less than 25% Latino students; high otherwise.[5]
5. Percent African American: low (0) and high (1) concentration—low where a school has less than 25% African American students; high otherwise.[6]
6. Percent free/reduced lunch: low (0) and high (1) percentages—low where less than 33% of the students received free/reduced lunch; high otherwise.

[5] Descriptive statistics indicated that 17 of the classes were located in schools where there's less than 20% Latino students.
[6] Descriptive statistics indicated that 27 of the classes were located in schools where there's 2% or less African American students.

7. Percent reading proficient: low (0) and high (1) percentages—median split (though the actual frequency distribution indicated that %proficient at a school was either less than 33% or 60% or more).
8. Percent math proficient: low (0) and high (1) percentages—median split (frequency distribution similar to that of reading).
9. Dose1: partial year — at least 6 months (6 coded as 1), compared to 4 months.
10. Dose2: full year (9 coded as 1), compared to 4 months.[7]
11. Urban school indicator, compared to suburban
12. Rural school indicator, compared to suburban
13. Master's degree indicator, compared to bachelor's
14. Doctoral degree indicator, compared to bachelor's
15. Medium school indicator (between 1,000 and 2,000), compared to small school (less than 1,000)
16. Large school indicator (more than 2,000), compared to small school

[7] Exploratory analysis with these two dosage indicators seemed to suggest a linear relationship (not controlling for other variables), but when these two dosage indicators were included in the HLM analysis, they weren't significant.

Appendix C: HLM Final Model

Level-1 Model

$$Y_{ij} = \beta_{0j} + \beta_{1j}*(\text{LANG-LANG..})_{ij} + \beta_{2j}*(\text{GPA-GPA}_{..})_{ij} + \beta_{3j}*(\text{OVERALL-OVERALL}_{..})_{ij} + \beta_{4j}*(\text{AA-AA..})_{ij} + \beta_{5j}*(\text{LATINO-LATINO..})_{ij} + r_{ij}$$

Where

Y_{ij} is the outcome for student *i* in classroom *j*;

β_{0j} is predicted posttest scores for students in control class with teachers who have bachelor's degree;

β_{1j} to β_{5j} are the regression coefficients that capture the relationships between student background characteristics and student outcomes;

r_{ij} is the residual associate with an individual's outcome in classroom *j*, which is assumed to be normally distributed with a mean of zero and variance of σ^2.

Level-2 Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(\text{CONDIT})_j + \gamma_{02}*(\text{MASTER})_j + U_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}*(\text{CONDIT})_j$$

$$\beta_{2j} = \gamma_{20}$$

$$\beta_{3j} = \gamma_{30}$$

$$\beta_{4j} = \gamma_{40} + \gamma_{41}*(\text{CONDIT})_j$$

$$\beta_{5j} = \gamma_{50} + \gamma_{51}*(\text{CONDIT})_j$$

Where

β_{0j} to β_{5j} are the intercept (β_{0j}) and slopes (β_{1j} to β_{5j}) from the level-1 model;

γ_{00} to γ_{50} represent the mean of intercept (γ_{00}) or slopes (γ_{10} to γ_{50});

γ_{01} to γ_{51} are level 2 regression coefficients that capture the effects of classroom-level variables on the within-classroom relationships between individual student background and student outcomes; and

u_{0j} represent the variability in β_{kj} after taking classroom characteristic variables (i.e., condition and teacher master's degree) into consideration.