Claremont

GRADUATE UNIVERSITY

THE EFFECTS OF PRENTICE HALL BIOLOGY ON STUDENT PERFORMANCE: PILOT STUDY
Final Report

Submitted by:

CLAREMONT GRADUATE UNIVERSITY
Institute of Organizational and Program Evaluation Research

August 31, 2005

Rebecca M. Eddy, Ph.D.
Tiffany Berry, Ph.D.

175 East Twelfth Street
Claremont, CA 91711
(909) 607-1549
(909) 621-8905 fax

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 like to thank Dr. Mariam Manley from PRES Associates, for all of your time and talent. We would also like to give a special thanks to the team from Claremont Graduate University including Dr. Tina Christie, Laura Berger and our research associates Kile Dyer and John LaVelle, who have worked tirelessly on many aspects of this project. We also appreciate everyone from Pearson Education including Brent McKenzie, Julia Fellows Osborne, Michelle York, and especially Marcy Baughman—you have been an absolute pleasure to work with. 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

Table of Contents

Tables

Executive Summary

Background & Study Design

- A pilot study of Miller & Levine's (2004) Biology curriculum was conducted in preparation for a Randomized Control Trial (RCT) to be conducted in 2005-06. The primary purpose of the pilot study was to finalize all instruments and protocols to be used during the RCT. In addition, research questions were addressed that focused on determining the efficacy of the Prentice Hall *Biology* curriculum.

- This study was conducted in one high-school in California. It included two treatment teachers (using Miller & Levine's 2004 Prentice Hall *Biology* curriculum) and two control teachers (using a non-Prentice Hall Biology curriculum). Data were assessed on 299 students, although only 259 were included in the final analysis.

- Outcome measures that assessed the impact on student attitides and acheivement included a student survey and a standards-based Biology assessment. Both were administered as pre and post-test instruments.

    - The student survey was administered to participating students and requested their opinions on topics such as attitudes towards science, parental involvement, meta-cognitive skills, and intrinsic motivation.

    - The standards-based Biology assessment was developed in conjunction with WestEd and included 30 enhanced multiple-choice items.

- Data collection instruments that tracked program implementation included online teacher self-reports of classroom activities completed weekly, classroom observations conducted during the fall and spring, and teacher interviews conducted at the conclusion of the study.

Research Questions

There were several research questions that guided this study and included the following:

- How do student outcomes differ for students using the Prentice Hall *Biology* program compared to students not using Miller and Levine's (2004) *Biology* program?

- How do students with various characteristics (e.g., Gender, Ethnicity, etc.) participating in the Prentice Hall *Biology* program perform on student-related outcomes?

- What is the relationship between program implementation and student achievement in science?

- What is the relationship among students' higher-order thinking, student attitudes towards science, and achievement in science?

Study Results

▶ Overall, students using the Miller & Levine (2004) *Biology* curriculum were similar in pre-test and post-test measures of attitude and performance as compared with students not using the Miller & Levine curriculum. In addition, there were no significant differences between male and female students or between ethnic groups on these measures.

▶ A factor analysis was conducted using student survey responses to both modify the existing instruments and to also assess differences between the groups on key constructs. A total of 14 items were removed from the 43-item survey and a final analysis of 5 factors emerged including intrinsic motivation & metacognitive skill; classroom environment; parent involvement; textbook satisfaction; and preference for science.

▶ Students in the using the Miller & Levine (2004) *Biology* curriculum scored higher on the post-test than students in the control condition on measures of intrinsic motivation, classroom environment, and parental involvement, however, this difference was not statistically significant.

▶ The most notable difference between the two groups was that students using the Miller & Levine (2004) *Biology* curriculum rated their textbooks significantly more favorably than those using other textbooks. However, this did not translate into higher ratings of student achievement on the science assessment.

▶ Post-test questions that were rated significantly different for students in the treatment group including text layout, diagrams and pictures, graphics, etc. Treatment group students rated the textbooks significantly more positively than control group students.

▶ Study implementation measures were adequate to investigate the extent to which teachers adhered to implementation guidelines. It is clear that the two teachers using the Miller & Levine (2004) *Biology* used the curriculum differently, although it was established that implementation guidelines should be revised.

Changes to the RCT Study

- An analysis of the study procedures and student outcomes has informed modification of several aspects of study implementation. Modifications made that will be reflected in the RCT include revision of the following:

- Teacher training: Based on analysis of training procedures and teacher feedback, training protocols were modified to focus on pedagogy during the initial training session. We will also require a follow up session that examines ancillary technologies with greater depth.

- Implementation guidelines: Based on feedback from teachers, new guidelines were established to focus on Miller & Levine (2006) text as the primary source of information. This focus stresses the importance of differentiating instruction for various learning levels and taking advantage of the textbook layout structure to maximize learning opportunities.

- Student survey: Based on a factor analysis, the existing instrument was revised to exclude 14 questions. We also revised the language of other questions to increase clarity of language for students.

- Standards-based science assessment: Given the poor performance of both groups in the pilot study on the post-test, we revised this instrument slightly to reduce language difficulty while maintaining instrument reliability.

- Teacher implementation log: We revised the content of the teacher logs to address specific elements of the Miller & Levine (2006) textbook and ancillary materials. These are expected to be helpful in answering research questions in the RCT study.

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' higher-order thinking skills. When combined, the assumption behind these reform efforts suggest if students learn science through a standards-based curriculum and are given the opportunity to develop and exercise higher-order thinking skills, students will become better equipped to learn the course material and incorporate it into their existing knowledge structures around science. 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 the big picture thereby enabling them 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"[1].

Prentice Hall is one such publisher that has shown great promise in the area of national science reform. Current versions of their science texts are visibly aligned with national science standards, and the organization and conceptual structure of the texts are designed to enhance understanding and promote higher-order thinking skills. In this age of accountability, and when multiple competing options for science texts exist, it is imperative that empirical research be conducted to examine the extent to which Miller and Levine's (2004) Prentice Hall *Biology* textbook impact students' higher-order thinking skills, attitudes towards and interest in science, and ultimately achievement in science. Given the requirements of US Department of Education's *What Works Clearinghouse*[2], it is imperative that research involves the use of experimental controls, usually referred to as Randomized Control Trials (RCTs). However, in order to conduct an effective RCT study, a pilot study was required to establish instruments and procedures to be used in a larger, more rigorous RCT study. The current report summarizes findings from this pilot study.

Purpose of the Prentice Hall Miller & Levine (2004) *Biology* Pilot Study

A pilot study of Prentice Hall *Biology* curriculum was conducted at one site in California. During the current pilot study two groups including treatment group (using Miller & Levine's 2004 Prentice Hall *Biology* curriculum) and a control group (not using Prentice Hall *Biology*) were compared. The pilot study examined teachers' implementation of the *Biology* curriculum as well as a diverse set of student outcomes. Although these data may provide insight into how the Prentice Hall *Biology* curriculum affects students' achievement in science during high school, it cannot offer definitive answers as to the effectiveness of Prentice Hall

[1] Close, D. (1996). National standards and benchmarks in science education: a primer. ERIC Digest: SE 058913.
[2] Detailed information regarding the What Works Clearinghouse can be accessed at www.w-w-c.org.

Biology to the extent that it cannot be generalized beyond the pilot site. The main purpose of the pilot study was to establish instruments and protocols to be used during the RCT study in the 2005-06 school year. We feel that findings from the pilot study only hint at the effects on student learning from Prentice Hall *Biology* as we have established research questions regarding the differences between those using Miller & Levine's *Biology* curriculum and those who did not. The pilot study focused on systematically tracking curriculum implementation, measuring students' achievement in science, and investigating the relationship between these elements with an assessment of the students' attitudes towards/interest in science, higher-order thinking skills, and product satisfaction.

Pilot Study Description & Design

The Prentice Hall *Biology* pilot study was conducted from September, 2004 to June, 2005. One high school site, including four teachers (two in the treatment group and two in the control group) were recruited to participate. Because of the nature of this study, teachers selected which group they wanted to participate in. Teachers were not randomly assigned to conditions and therefore we have limited ability to assess whether or not teacher self-selection in groups did or did not impact the results. It would have been ideal to randomly assign teachers to either the treatment or control group; however, since establishing cause was not the main purpose of the study, it was not necessary to use random assignment to answer the research questions. It is still possible, however, to describe differences between the two groups even when lacking the ability to assign causality to those differences. To assess these differences, qualitative and quantitative instruments were developed and tested within the pilot study. These instruments and protocols are currently being modified/ adapted for use in the RCT study.

The primary questions motivating this research included the following:

- How do student outcomes differ for students using the Prentice Hall *Biology* program compared to students not using Miller and Levine's (2004) *Biology* program?

- How do students with various characteristics (e.g., Gender, Ethnicity, etc.) participating in the Prentice Hall *Biology* program perform on student-related outcomes?

- What is the relationship between program implementation and student achievement in science?

- What is the relationship among students' higher-order thinking, student attitudes towards science, and achievement in science?

The pilot study design entailed using multiple measures including those that address student achievement and those that addressed program implementation. Two instruments were administered as a pre-post test at the beginning and the end of the 2004-05 school year. Although we recognize that statistical artifacts such as regression toward the mean are common in such situations where measures are administered multiple times, we feel that this design was the best way to track student growth in the study. Other measures (e.g., teacher logs) were reported throughout the entire year. Next is a description of each type of measure and corresponding characteristics.

Outcome Measures

Standards-Based Science Assessment

Participation in the study required students to complete two measures at pre-test and post-test. The first of these measures is a science assessment that was administered to all participating students. A standards-based, nationally recognized Biology assessment was identified in August 2004 to measure student learning in Biology. The Partnership for the Assessment of Standards-based Science (PASS) is 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). Over the past decade, PASS has collected solid technical data on the reliability and validity of their assessments, which have been shown to possess reliability coefficients of greater than .90. 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 contained high inter-item reliability.

This Biology pre-test assessment was administered in Fall of 2004 and the post-test was administered in spring of 2005. Two sections of the assessment were available for use including a 30-question enhanced multiple choice section and a constructed response section. However, due to teacher difficulties, students in the pilot study completed only the enhanced multiple choice section of the assessment. A copy of how the enhanced multiple-choice questions of the Biology assessment addressed the state and national standards can be found in Appendix 1. Both sections of this assessment will be administered in the RCT.

Student Survey

All students participating in the pilot study were required to complete a self-report survey that addressed attitudes towards science, higher order thinking skills, and product satisfaction (specifically textbooks). The results of this survey were analyzed using a factor analysis. The factor analysis served multiple purposes: (1) Revision of the existing instrument to ensure its reliability (2) Determination of any differences between ratings given between students in various groups (treatment versus control, ethnicity, etc.) (3) Identification of factors related to student achievement on the PASS assessment. A more specific description of the student survey is provided next.

Student Attitudes toward Science: The first portion of the survey addressed students' attitudes/interest in science. It was important to measure students' attitudes towards/interest in science for two primary reasons. First, students' attitudes towards science have been related to student achievement in science (TIMSS, 1999)[3]. Therefore, these factors must be measured so that we understand which factors contribute to variability in science achievement scores. Second, since the Prentice Hall *Biology* texts offer innovative technologies to engage students, it is important to measure whether these students' attitudes towards/interest in science change to a greater extent than students not participating in the Prentice Hall *Biology* curriculum for reasons attributable to factors other than textbook content.

Higher Order Thinking Skills: Based on a review of the Prentice Hall *Biology* materials, it is clear that these texts were developed to promote students' understanding of the material by encouraging students to practice their higher order thinking skills. Factors previously identified in research related to cognition and learning (National Research Council, 2000)[4] are aligned with several facets of the Prentice Hall *Biology* program. For example, major concepts are organized appropriately to demonstrate how they relate to other concepts, unit assessments promote deeper level of understanding than just identification of key terms, and teacher resources provide ways to address prior student misconceptions of knowledge. Given that Prentice Hall *Biology* materials are visibly aligned with several of the guiding principals that foster effective thinking strategies, we feel it is possible that this is the reason students may have success using the Prentice Hall curriculum. We also need to understand if the text is effective at fostering these skills in students.

Product Satisfaction: The final area of assessment on the student survey assessed to what extent students rated specific aspects of their textbooks favorably. For example, if students feel more favorable about features of their textbooks (e.g., "My textbook explained difficult concepts in everyday language", or "I enjoy reading my Biology textbook", etc.) it is possible that this factor could have contributed to student learning or their engagement of the material.

---

[3] Trends in International Mathematics and Science Study. Retrieved July 25, 2004 at http://nces.ed.gov/timss/

[4] 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.

Implementation Measures

The key to interpreting outcomes is in understanding how well the Biology curriculum has been implemented in each classroom. Implementation measures included teacher interviews, weekly teacher implementation logs, and two classroom observations (fall and spring). We understand that experienced teachers will vary their method of content delivery, however, participating teachers must also cover a minimum level of material from the Miller and Levine (2004) text to consider the program "implemented". There are two main goals in tracking implementation: (1) to ensure that teachers are implementing the Prentice Hall *Biology* program as intended by the publisher, and (2) to ensure that teachers are creating conditions in their classrooms that supports the higher order learning principles being tested. For example, specific classroom norms, such as connecting the current lesson to prior materials learned, can either facilitate or hinder student learning.

Teacher and student use, including in-classroom and additional activities, were monitored to ensure that an adequate level of implementation was present in all classrooms. For example, it would be inappropriate to compare student assessments with the assumption of equal implementation if in fact one teacher has less than the minimal level of implementation of the science program while all others have much higher levels of implementation. Measures used in the pilot study addressed these issues.

Teacher Implementation Logs

Each week teachers participating in the study were required to complete online logs that described activities in their classrooms. Teachers documented both the content covered in their classes (unit, chapter and section) as well as the delivery of that content (assigned reading, lab exercises, etc.). In addition, teacher logs were useful as a source of teacher reflection on their own practice or for providing informal feedback regarding use of the products or other issues with program implementation.

Teacher Interviews

After completion of the pilot study, all teachers were interviewed individually. Interview questions inquired about several facets of the research study including the experience of participation in the study such as communication from researchers, training procedures, logistics details, etc. Teachers who used the Prentice Hall materials were also required to answer more specific questions about product satisfaction, textbook content, and other relevant issues.

Classroom Observations

All participating classrooms were observed at a minimum of two occasions by the research team. Classroom observation protocols were adapted from a variety of sources and modified for use in the pilot study. The final revised instrument will be used in the RCT study. The instrument used included descriptive information; information about activities that occurred in the classroom (e.g., review of chapter, quiz, etc.); materials used (e.g., textbooks, overhead slides, etc.); extent to which students were "engaged" in the lesson; and the cognitive demand required in the classroom (e.g., interaction between teacher and students, level of Bloom's (1956)[5] taxonomy engaged).

Together, these data sources allow us to understand the activities that occurred in participating classrooms throughout the pilot study year. The most useful part of tracking implementation allows us to provide context for the quantitative results. For example, during the study it was discovered that a subgroup of students in the participating classrooms should not be included in the greater analysis due to specific differences between these students and most other students. This scenario is discussed next as well as the results of the pilot study.

Pilot Study Results

Throughout this report, research question will be addressed using quantitative data including student surveys, student assessments, and teacher implementation logs. In addition, these data will be informed by qualitative data such as student reports of textbook satisfaction, classroom observations, and teacher interviews that will help to interpret the findings.

Based on initial results of teacher interviews, student surveys and student assessments, it was clear that there were two distinct groups participating in the pilot study. The first group included students in 9th and 10th grades that completed a Biology course in sequence with their classmates. The second group included students in the 11th and 12th grades who were either taking Biology out of sequence from their classmates or taking Biology for a second or third time. An initial analysis of student achievement results indicates that these were two distinct groups. For this reason, we proceeded with data analyses on the first group only, that is, only those students in the 9th and 10th grades. After eliminating the 11th and 12th grade students, a total of 259 students participated in all aspects of the study. The demographic characteristics of these participating students are described next.

Demographic Characteristics

Table 1 summarizes the demographic characteristics of students in the treatment and control groups. Demographic description of participants includes gender, grade level and ethnicity. There were approximately equal numbers of male and female students in both groups. Most students (76% in the treatment group and 80% in the control group) were in 10th grade, and the rest were in 9th grade.

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

Finally, the sample was ethnically diverse as one-third of the students were from Asian descent, the next most represented group was African American students (more than 20% for the treatment group and almost 28% in the control group); the next most-represented group were students who identified themselves as Hispanic (18.6% and 16.9% in the treatment and control groups, respectively). Finally, the percent of Caucasian students was 17.8% in the treatment group and 11.8% in the control group. These results demonstrate that students in the treatment and control groups were similar at the outset of the study.

Table 1. Demographic Characteristics of Pilot Study Participants

	Response Options	% Treatment	% Control
Gender	Male	51.6%	48.2%
	Female	48.4%	51.8%
Grade Level	9th	24.6%	19.7%
	10th	75.4%	80.3%
Ethnicity	Caucasian	17.8%	11.8%
	African American	20.3%	27.9%
	Hispanic	18.6%	16.9%
	Asian	32.2%	33.8%
	Other/ Unknown	11.1%	9.6%

Figure 1 summarizes self-report ratings from students regarding time spent on homework each day. These results indicate that the groups are similar in terms of how time is spent on homework on a typical day.

Figure 1. Reported Hours Spent on Homework Daily

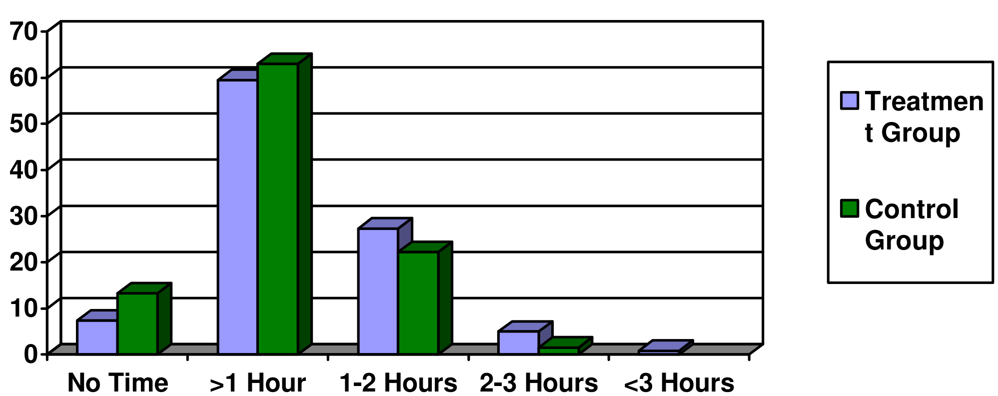


Table 2 summarizes parent education level for participating students. These variables are important to compare between the two groups, particularly mother's education level. There were no significant differences between these variables when comparing treatment and control groups.

Table 2. Parent Education Level of Pilot Study Participants

	Response Options	% Treatment	% Control
Mother's Reported Education Level	Not a High School Graduate	16.9%	11.9%
	High School Graduate	12.7%	20.0%
	Some College	28.8%	33.3%
	Bachelor's Degree	4.2%	2.2%
	Master's Degree	2.5%	1.5%
	Doctoral/ Professional Degree	.8%	-
	Don't Know/ Not Stated	34.1%	31.1%
Father's Reported Education Level	Not a High School Graduate	15.3%	10.4%
	High School Graduate	18.6%	30.4%
	Some College	23.7%	16.3%
	Bachelor's Degree	5.1%	5.2%
	Master's Degree	2.5%	3.7%
	Doctoral/ Professional Degree	-	.7%
	Don't Know/ Not Stated	34.8%	33.3%

The reason why it is imperative to compare treatment and control groups on characteristics specified in terms of demographic characteristics as well as amount of homework assigned each day is because we wanted to ensure that the groups were similar prior to full implementation of the Biology curriculum. While there may be differences observed in terms of teacher characteristics we can at least be assured that student characteristics were similar at the outset of the study.

Research Questions

The remaining sections of this report will address each of the research questions from the pilot study. Specifically, we will describe the combination of measures that addresses each question as well as describe research methodology that was required at each step. Overall, findings reveal that there were only minor differences between the treatment and control groups on measures of student achievement and attitudes towards science. However, there were significant differences between students regarding textbook satisfaction ratings, such that students in the treatment group rated their textbooks more favorably than students in the control group despite the fact that this did not translate into higher ratings of student achievement on the science assessment.

How do student outcomes differ for students using the Prentice Hall *Biology* program compared to students not using Miller and Levine's (2004) *Biology* program?

The first major research question compares outcomes of students in the treatment group to those in the control group. We conducted comparisons on these two groups for both pre-tests and post-tests as well as comparisons on individual teachers within each group, where appropriate.

Standards-Based Biology Assessment

Figure 2. Pre-test and Post-test Scores for Treatment and Control Groups on the Standards-Based Biology Assessment

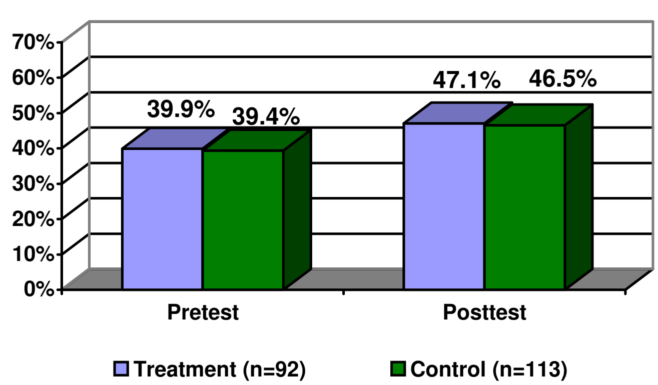
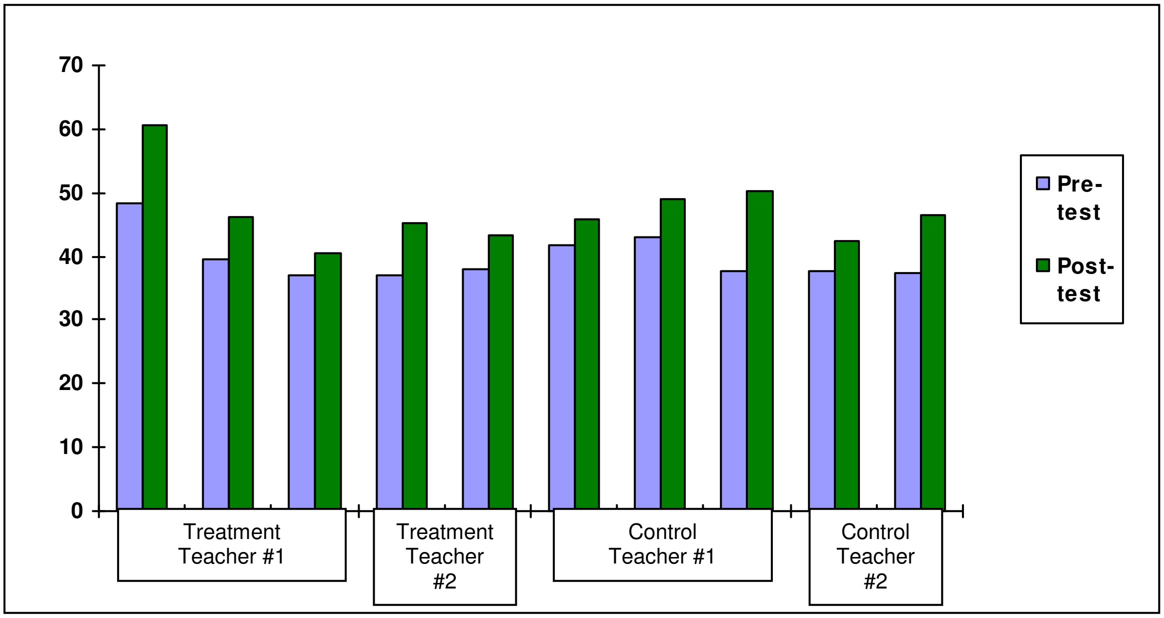


Figure 2 illustrates scores on the standards-based Biology assessment in terms of percent. A repeated-measures Analysis of Variance (ANOVA) was conducted to investigate whether there were systematic differences between the treatment and control groups from pre-test to post-test. Results indicate there were significant increases from pre-test to post-test for both groups $F(1, 203) = 62.35$, $p<.01$. This result means that significant learning occurred between pre-test and post-test for students in both groups. However, there are only slight differences between the two groups overall where students in the treatment group scored slightly higher (M= 47.1%) than students in the control group

(M=46.5%) though this difference was not statistically significant $F(1, 203) = .01$, $p > .05$. We also analyzed student performance on the assessment for class period and each teacher. Figure 3 illustrates that there are differences between class periods in their overall performance from pre-test to post-test.

Figure 3. Pre-test and Post-test Scores on the Standards-Based Science Assessment by Participating Teacher



As demonstrated in Figure 3, classes varied greatly from pre-test to post-test, even within each of the participating teachers' classes. There were an equal number of classes that had significant increases from pre-test to post-test. These within-group differences tend to wash out any effects of a given treatment group.

We propose some possible explanations for these results:

- First, we must again provide the caveat that these teachers were not randomly assigned to either the treatment or control groups, and therefore it is quite possible that there is a self-selection bias that may mask the true impact of the Biology curriculum.

- Second, it is possible that even though the Miller & Levine (2004) *Biology* curriculum may in fact be a superior product, the full impact of using the new curriculum for teachers will not be realized until after the first year of implementation. There will necessarily be more difficulty in implementing a new curriculum for the first time (treatment teachers) as opposed to practicing the same curriculum that has been in place for years (control teachers). The treatment teachers necessarily have a distinct disadvantage for this reason.

- Third, we considered that the instrument itself was not a good indicator of students' true knowledge. Given that this result seems feasible, we reviewed the instrument and made changes for the RCT. The full explanation can be found in the last section of this report.

- Fourth, a final explanation could be that the impact of a Biology curriculum has such a small effect on student achievement and there are not significant enough differences between the two texts. At this point, is it unclear whether or not this is the most feasible explanation. We expect that we will be able to have a clearer indication of the most likely scenario when analyzing data from the RCT with a greater number of students and teachers that are randomly assigned to the treatment conditions.

<u>Student Survey</u>
Next, we considered if there were differences between the treatment and control groups on the student survey—also administered as a pre-test and post-test measure. The 43-item survey was first analyzed using a factor analysis procedure that allowed us to (1) test constructs such as parent involvement and student favorability of science and (2) to provide a stronger interpretation of these constructs as factors rather than individual survey questions.

An initial exploratory analysis was analyzed using only the pre-test data. From this, 6 distinct factors emerged, however, not all factors had reliable alpha coefficients (generally considered reliable at >.70). However, we re-ran these data incorporating additional questions related to textbook satisfaction on the post-test. The results of the new model are given in Table 3.

Table 3. Descriptive Statistics for Treatment and Control Students on Each Composite Factor Derived from the Student Survey

Composite Factors	Treatment Group Post-test Mean (Std. Deviation)	Control Group Post-test Mean (Std. Deviation)
Meta-cognitive Skill and Intrinsic Motivation (8 items, alpha = .80)	3.12 (.65)	3.01 (.70)
Classroom Environment (7 items, alpha = .73)	3.92 (.50)	3.77 (.56)
Parent Involvement (4 items, alpha = .76)	3.16 (.96)	2.93 (.90)
Textbook Satisfaction (5 items, alpha = .75)	3.58 (.63)	3.06 (.76)
Lack of Science Ability (5 items, alpha = .70)	2.43 (.68)	2.49 (.71)

Meta-cognitive Skill & Intrinsic Motivation:
Ratings were lower on the post-test survey for questions relating to meta-cognitive skill and intrinsic motivation for both groups. On this factor, there were no significant differences between the two groups ($p > .05$); however, the students in the treatment group rated these items higher at post-test. This means that both groups did not increase self ratings of intrinsic motivation after completing a year of Biology. Meta-cognitive skill and intrinsic motivation appears to have remained constant for students participating in this study.

Classroom Environment:
Students in the treatment and control groups rated items related to the classroom environment exactly the same on pre-test surveys, however, students in the treatment group rated them marginally significantly higher ($p = .06$) on the post-test survey than students in the control group. This result indicates that the classroom environment such as effective teacher behaviors and routines were similar for both groups.

Parent Involvement:
There were no differences among reports of parent involvement for the treatment or control groups on either the pre-test or the post-test, although these means decreased slightly from pre-test to post-test. As previously mentioned, it is expected that some ratings above the mean would decrease in this fashion, as this is a typical statistical artifact. For this variable, the important point is not that the groups were different (as we would not expect them to be) but that four of the five questions related to parent involvement can be combined into one highly reliable composite factor.

Textbook Satisfaction:
Questions related to textbook satisfaction were developed only for the post-test. There were a total of five items that could be combined to create a composite factor. There were significant differences between the treatment and control groups on this factor such that students in the treatment group (M = 3.58) rated their texts much more favorably than students in the control condition (M = 3.06), $t\,(df = 176) = 4.9$, $p < .01$. This result will be discussed in greater detail in the section related to Product Satisfaction.

Lack of Science Ability:
The final composite item relates to students' view of their own ability and favor of science. This item was reversed scaled, in which higher ratings indicated that students felt less confident in their own abilities related to science. In general, students reported their own abilities more favorably on the post-test than on the pre-test; however, this was largely dependant on which Biology teacher they had. Students of one treatment teacher and one control teacher rated this composite more favorably than the other two teachers, and therefore this was not a treatment effect, but rather a teacher effect, though not statistically significant.

Summary

In general, there were few differences between the treatment and control groups on measures of attitude and science achievement. While student scores generally increased from pre-test to post-test, main differences existed within ratings of product satisfaction. Our hope is that modification of the existing science assessments and surveys will help to define these differences more clearly in the RCT study.

How do students with various characteristics (e.g., Gender, Ethnicity, etc.) participating in the Prentice Hall *Biology* program perform on student-related outcomes?

Student Assessment

The next research question addresses how students with different characteristics performed on the science assessment and whether or not there were differences in terms of student achievement. Figure 4 includes an analysis by gender. A repeated-measures ANOVA was conducted to compare male and female students on science assessment pre-tests and post-tests. Results indicate there were no significant differences between male and females students on the science assessment, $F(1, 201) = .65$, $p > .05$., although male students scored slightly higher in both groups.

Figure 4. Pre-test and Post-test Scores for Treatment and Control Groups on the Standards-Based Biology Exam for Male and Female Students

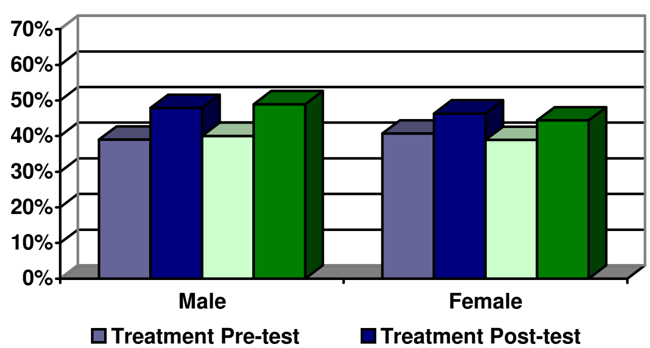
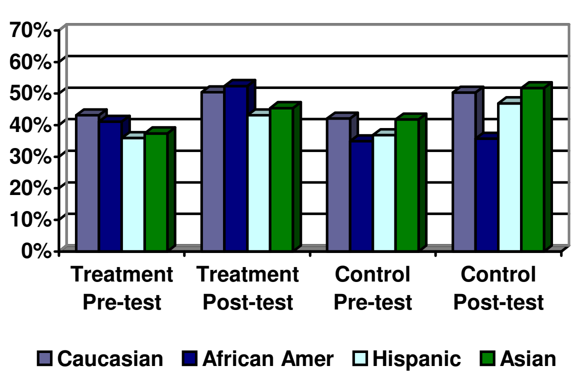


Figure 5. Post-test Scores for Treatment and Control Groups on the Standards-Based Biology Assessment by Ethnic Group



Another repeated measures ANOVA was conducted to investigate differences between ethnic groups on the pre and post-tests by treatment condition. Mean percentages for each ethnic group and each assessment are illustrated in Figure 5. Both the treatment and control groups increased overall on the scores from the Biology assessment.

Students in the ethnic group scoring the highest on the post-test assessment for the treatment group were African American students (M = 52.4, n=26) and the group scoring highest on the post-test assessment in the control group was Asian students (M = 51.7, n=46). No group scored statistically lower than any other group on the science assessment either at pre-test or post-test. These results are encouraging; however, when each group is disaggregated by ethnicity, the sample size is small and unstable. Small sample size is a concern for making predictions for any variable. We will ameliorate the issue of small sample size in the RCT because there will be large samples of most major ethnic groups participating in the study.

Student Survey

Male and female students were also compared on how they responded to the student attitudes survey. In addition, we tracked growth by ethnic group as well. Table 4 demonstrates how each group scored on the pre-test and post-test student survey.

Table 4. Mean Ratings on Post-test Composite Factors for Gender and Ethnic Groups

Group	Composite 1		Composite 2		Composite 3		Composite 4		Composite 5	
	Treat.	Cont.	Treat.	Cont.	Treat.	Cont.	Treat.	Cont.	Treat.	Cont.
Male	3.0	2.9	3.9	3.8	3.0	2.8	3.5	3.1	3.1	3.1
Female	3.3	3.0	4.1	3.8	3.3	3.1	3.7	3.1	3.0	3.2
Caucasian	2.9	2.9	3.7	3.8	3.0	3.4	3.4	2.9	2.8	2.9
African American	3.3	2.8	4.2	3.9	3.7	3.2	3.9	3.0	2.5	3.3
Hispanic	2.8	3.1	3.9	3.9	2.9	2.9	3.5	3.2	3.4	2.9
Asian	3.3	3.2	4.1	3.8	3.1	2.7	3.6	3.2	3.3	3.2

The largest differences observed between each of the treatment groups are between the African American students, though none of these differences are statistically significant, ($p > .05$). For example, African American students in the treatment group consistently rated all items higher than African American students in the control group, with the exception of Lack of Science Ability. Given that this item has a reverse scale, this is a positive result for these students. Males and females were similar on most composite factors where the general trend was to rate most items higher in the treatment group, especially Textbook Satisfaction.

Summary

In general, there were few differences observed between male and female students and those with varying ethnic backgrounds. The RCT study will also include analyses related to socio-economic status as well as language status.

We are encouraged by the positive trend in the data regarding treatment group students, however, obviously more study should be done to determine whether these trends are meaningful.

What is the relationship between program implementation and student achievement in science?

To address this research question, we looked at data related to program implementation (e.g., how well the curriculum was implemented in the study classrooms). Implementation was considered based on the teacher implementation logs as well as classroom observations. We also reviewed our primary measure of science achievement—the standards-based Biology assessment. Given that this was the pilot study, we also considered both sources of data critically to allow for the possibility of modifying the instruments to use in the RCT.

<u>Teacher Implementation Logs</u>
Table 5 shows the level of implementation for the two treatment teachers for the year. We have not reported this specific information for control teachers for several reasons: (1) there are not exact parallels in which we can report these with the competitor's product (2) one control teacher was on medical leave for several weeks and consequently we do not feel that the long-term substitute teacher was able to provide the same level of information as the control teacher would in weekly logs. However, we were able to determine the chapters covered for each teacher in both treatment groups and data are reported for both groups, although not in terms of parallel descriptions.

The first set of results shown in Table 5 indicates that Treatment teacher A covered more material in their classroom as compared to Treatment teacher B. We would expect that increased content coverage would result in students' increased achievement on the science assessment. However, results indicate that although students in Treatment teacher A classes scored higher than those students in the other treatment classes, this difference was only marginally significant, t (df=92) =1.87, p= .06. The mean score on the post-test assessment for Treatment teacher A was =.49; whereas the mean score on the post-test assessment for students in classrooms with Treatment teacher B was =.44.

Table 5. Percent of Weeks Treatment Teachers Implemented Components of the Implementation Guidelines

Implementation Guidelines	Treatment Teacher A	Treatment Teacher B
Daily use of Miller & Levine's (2004) Prentice Hall *Biology* text	77%	46%
Start each chapter with "Inquiry Activity"	38%	76%
Cover all Key Concepts in chapter	54%	27%
Cover some "Checkpoints"	3%	34%
Cover some "Caption Questions" throughout each chapter	8%	46%
Perform most "Quick Labs"	4%	4%
Complete all "Section Assessments"	12%	27%
Perform "Analyzing Data" exercises where applicable	0%	0%
One full lab per chapter	24%	30%
Complete parts of all end of chapter assessments	8%	4%
Use Guided Reading and Study Workbook	69%	19%
Use Summary and Vocabulary Review Book	19%	0%
Standardized Test Prep	8%	0%
Reviewed new vocabulary	92%	42%

The following Tables 6 and 7 also contain data reported in the teacher logs. Table 6 is a description of technologies teachers used during the year and Table 7 is a description of technologies students used during the year. Results indicate that teachers used various technologies more or less often depending on their personal preference, however, the general trend was again that Treatment teacher A used more ancillary materials and technologies and required his/her students to use more ancillary products as compared with Treatment teacher B. What is also interesting to note is that both teachers reported that they were adhering closely to implementation guidelines established at the onset of the study. Even after reminders to once again to adhere to these guidelines, Treatment teacher B persisted in avoiding some elements of these guidelines. Given the difficulty in imposing so many specific and various guidelines on teachers, we modified the existing guidelines for the RCT. These new guidelines are found in Appendix A.

Table 6. Percent of Weeks Treatment Teachers Used Supplemental Materials and Technology

Technologies the Teachers Used	Treatment Teacher A	Treatment Teacher B
Teacher's Edition I-TEXT CD-ROM	42%	4%
Reading and Study Workbook A	84%	0%
Lab Manual A	27%	12%
Standardized Test Prep Workbook	27%	0%
BioDetectives: Investigations in Forensics	8%	0%
Issues and Decision Making	7%	39%
Biotechnology Manual	8%	0%
Overhead Projector	96%	84%
Television/Video Tape/DVD	42%	38%
Other website or computer programs*	12%	7%

* Examples of “Other Website or Computer Program” include the following: Section Review spiral-bound book; Biology Project from U. of Arizona website...used Onion Root tip mitosis activity to learn to identify phases of mitosis; Power Point version of Jeopardy prepared by a student; Power point, search engines: google images, yahoo, alta vista

Table 7. Percent of Weeks Treatment Students Used Supplemental Materials and Technology

Technologies the Students Used	Treatment Teacher A	Treatment Teacher B
Reading and Study Workbook A	69%	0%
Lab Manual A	12%	8%
Standardized Test Prep Workbook	4%	0%
Bio Detectives: Investigations in Forensics	4%	0%
Issues and Decision Making	31%	4%
Other website or computer programs*	15%	8%

* Examples of “Other Website or Computer Program” include the following: Cell Biology Onion Root Tip activity from U. of Arizona Biology Project website; Teachersdomain.org - Info/video on Stem cells; various biome websites found by each student during research; Various websites and search engines, etc.

From Tables 5-7, we have concluded that Treatment teacher A had a more thorough level of program implementation. However, what is most notable about these data is the distinct lack of technologies and ancillary materials used by both teachers. We believe that the reason for this stems from inadequate training at the onset of the study and lack of support during the year. We have addressed both of these issues to ensure that there are not similar problems throughout the RCT study. However, it is also worth noting that it is necessarily more difficult for teachers to integrate an entire new program within one year. This is a more difficult task than the control teachers were assigned—mainly, implementing class as they always have. Given these facts alone, we expect that the treatment teachers were at a distinct disadvantage.

One additional rating we requested of teachers in the logs was to estimate time allocations for each activity category. The results of this analysis for all participating teachers are found in the following table.

Table 8. Mean Percentage of Classroom Time Teachers Allocated to Classroom Tasks

Mean Percentage of Class Time Spent on Classroom Tasks	Treatment Teacher A	Treatment Teacher B	Control Teacher A	Control Teacher B
Lecture to Class	16.2%	13.7%	13.9%	22.7%
Experiments/Labs	24.8%	30.3%	22.2%	35.4%
Student Discipline/Interruptions	2.9%	.09%	1.9%	1.0%
Routines (e.g.: checking homework, taking role, etc.)	5.7%	4.3%	5.0%	3.8%
Student Independent Work Time	21.2%	29.8%	25.9%	8.3%
Watching Video/Multi Media Presentations	3.8%	5.8%	9.1%	10.2%
Chapter/Unit Reviews	5.2%	2.1%	9.9%	5.4%
Administration of Quiz or Test	11.4%	6.0%	9.7%	10%
Other School Activities (e.g., assemblies, fire drill, etc.)	.07%	2.3%	.09%	.04%
Other	8%	4.9%	1.4%	2.7%
Number of Minutes of Homework Assigned for the Week	Mean # minutes=108	Mean # minutes = 9.6	Mean # minutes =141	Mean # minutes =52.4

These results indicate that all teachers differ in their time allocations. What is most interesting is the mean number of minutes each assigned for homework in which this number ranged from an average of 9 minutes to 141 minutes per week.

While we believe that the reporting of class activities in this way could be refined for use in the RCT, one additional problem that may continue to arise in the RCT is the difference in coverage across different classes but even within each participating teacher. Differences between classes are evident in the data from Figure 3. We have not made the assumption that everything reported in the teacher logs has occurred in the same manner in all participating classes, even for the same teacher. However, it would have been impossible to require teachers to report class periods separately. For this reason, we have supplemented the reported teacher log information with classroom observations to provide greater context to the classroom environment.

Classroom Observations

Two sites visits were conducted throughout the year specifically for the purpose of monitoring program implementation. Classroom observations also allowed us the opportunity to see the program implemented first-hand. This was important because it allowed us to view lessons and then later review what was reported by the teachers for that lesson. We can conclude from these data is that there is a level of subjectivity in self reports of program implementation. Classroom observation protocols were established and modified during the pilot study. Elements of the observations include environmental features such as temperature and noise level; distribution of time on various activities; student engagement level; and finally ratings of cognitive demand in the classroom.

There is one major caveat to consider with regard to these data: we have relied on these measures (particularly ratings of cognitive demand in a classroom) but understand that this may not be representative of what occurs in the classroom on a daily basis. Because of the artificial nature of this event, we cannot rely on classroom observations to provide definitive information beyond giving specific context to teachers' reports of special circumstances in their classes. For example, during one of the observations, we learned that one class period for one of the treatment teachers was not expected to perform well on the assessment given a high number of special education students in the classroom. We also observed first-hand how some students were disruptive during the lessons. This is an example of how the classroom observations help us to interpret the findings rather than just rely on student test scores.

Overall, there was an "adequate" level of implementation in each classroom; however, we would not classify the classroom experience as highly cognitively demanding for any of the participating teachers, based on two days of observation. We expect that this is the norm in Biology classrooms, however, we would also expect that students need a cognitively demanding environment to maximize learning opportunities. The surprising result is that we found similar teacher behaviors and student engagement throughout various class periods for the same teachers. This contradicts what we would have expected given the results from Figure 3, in that class performance varied within teacher. The only explanation that we have available is based on teacher comments where they explained that certain class periods have fewer numbers of high-achieving students due to the times when advanced placement classes are offered in other subjects. This could explain the within-teacher differences observed.

Table 9. Classroom Observation Summary

Classroom Observation Measure	Treatment Teacher A	Treatment Teacher B	Control Teacher A	Control Teacher B
Classroom Environment (Adequate/ Inadequate)	Adequate	Adequate	Adequate	Adequate
Level of Student Engagement (Fall)	70%	60%	80%	50%
Level of Student Engagement (Spring)	75%	85%	80%	50
Level of Cognitive Demand (Fall) Range: 1-4	1.8	1.5	1.5	1.0
Level of Cognitive Demand (Spring) Range: 1-4	1.2	2.4	2.0	1.6
Effective Teacher Behaviors (Fall) Percent out of 18 items	50%	30%	39%	30%
Effective Teacher Behaviors (Spring) Percent out of 18 items	50%	30%	50%	28%

<u>Summary</u>

In general, there was an “adequate” level of implementation, however, not nearly to the depth of presentation that would have been expected in a highly cognitively demanding classroom environment. There are mixed results related to implementation and student achievement with regard to the research questions. We could not detect any significant differences among the treatment groups regarding implementation; however, classroom observations did reveal differences among each of the teachers regarding presentation of material. We expect to be able to make more generalizations in the RCT with a larger number of participating teachers and their classes.

What is the relationship among students' higher-order thinking, student attitudes towards science, and achievement in science?

The final research question addresses the relationships between students' thinking skills, attitudes towards science and achievement in science. Table 10 includes a correlation matrix for both treatment and control groups.

Table 10. Correlation Coefficients for the Relationship Between Student Assessment Post-test Scores (PASS) and Student Composite Factors

	Post-test Composite 1	Post-test Composite 2	Pre-test Composite 3	Post-test Composite 4	Post-test Composite 5	PASS Post- test
Post-test Comp 1	--					
Post-test Comp 2	.469**	--				
Post-test Comp 3	.291**	.290**	--			
Post-test Comp 4	.482**	.428**	.314**	--		
Post-test Comp 5	-.384**	-.088	.126	-.202**	--	
PASS Post-test	.251**	.041	-.110	.084	-.333**	--

**Correlation is significant at the $p< .01$ level

Composite Factor 1: Meta-cognitive Skill & Intrinsic Motivation
Composite Factor 2: Classroom Environment
Composite Factor 3: Parent Involvement
Composite Factor 4: Textbook Satisfaction
Composite Factor 5: Lack of Science Ability (reverse scaled)

Results in Table 10 indicate that the Composite Factor 1 (Meta-cognitive Skill/ Intrinsic Motivation) is highly correlated with all other composite factors and the standards-based science assessment (PASS post-test). Students with higher intrinsic motivation and advanced thinking skills generally perform better on student assessments, have higher levels of parent involvement and generally have greater interest and skill in science and hence would rate Factor 5 lower. These students also rated the textbook features more favorably. Most of the other results also follow this same pattern of correlation, with Factor 5 being negatively correlated with four of the five other items.

The only result not anticipated was the lack of significant relationships between the PASS post-test score and parent involvement. Students with highly involved parents tend to score higher on measures of academic achievement. Our hope is that the newly-revised instrument will be sensitive enough to detect these differences and produce the expected results.

Product Satisfaction

Student Satisfaction Ratings

Students using the Prentice Hall *Biology* textbook rated the Composite Factor 4 significantly higher than students in the control group. The following table summarizes the results for all questions related to product satisfaction for the Miller & Levine's (2004) *Biology* curriculum. The rating scale ranges from 1 = Strongly Disagree to 5 = Strongly Agree.

Table 11. Textbook Satisfaction Ratings for Treatment and Control Group Students

Survey Question	Treatment Students Mean Rating (Std. Dev.)	Control Students Mean Rating (Std. Dev.)
My Biology book explained difficult concepts in everyday language.	3.4 (.96)	3.1 (.92)
In the Biology book it was easy to locate the main points of each paragraph.	3.8 (.85)	3.7 (.96)
I enjoyed reading my Biology book.	3.1 (.93)	2.3 (1.0)
I learned a great deal from my Biology book.	3.7 (.77)	3.3 (1.0)
I found myself sharing information found in this text with others not involved in the course (e.g., Family members or friends).	3.4 (1.1)	2.7 (1.2)
I had to read sections of my Biology book multiple times before I understood the concepts.	3.2 (1.0)	3.3 (1.1)
Rating of the extent to which the following sections of the Biology textbook helped students understand the concepts.		
Pictures, diagrams, charts, tables	4.3 (.77)	4.2 (.69)
Experiments/ labs	4.0 (.90)	3.5 (1.2)
End of chapter questions	3.7 (.90)	3.4 (1.2)
Overall layout of the text	4.0 (.79)	3.7 (1.0)

Overall, students rated the Prentice Hall *Biology* textbook more favorably than the non-Prentice Hall Biology textbook. In particular, students reported that they enjoyed reading their textbook and would be more likely to share this information with others. Students also found the layout of the text as well as experiments/ labs particularly helpful in the Prentice Hall text.
We also requested that students provide additional information regarding what they liked best and least about their book. The following is a short summary of those comments.

Students using the Prentice Hall *Biology* textbook reported the following:

- More than half of students that provided specific comments rated that layout of the text as a favorite feature. This included bold/ highlighting of vocabulary terms and main points.
- The book was considered easy to read and key concepts were easy to find. Other students reported that explanations were thorough and detailed.
- Half of students providing feedback reported that they liked pictures and diagrams the best.
- One-third of students that provided comments regarding what they liked least reported that the book was too heavy, too long and difficult to carry.
- Some students reported that the reading was too difficult.
- One-third of students reported that labs were not helpful.

<u>Teacher interviews</u>
Teachers using the Prentice Hall *Biology* program participated in one formal and several informal interviews in which they responded to questions regarding their use and satisfaction of the Biology program.

It is clear that the two treatment teachers used the Prentice Hall Biology textbook and supplementary materials differently. One explanation is that the content of their Biology courses varied slightly. One treatment teacher taught sections of General College Preparatory Biology classes while the other teacher taught sections of Agricultural Biology. While each had to address the California state content standards, the focus of the curriculum varied slightly. Given that there were differences in product use, we inquired about the reasons why they did or did not use certain elements of the program or conduct specific laboratory exercises.

A summary of their comments is a follows:

- Some lab exercises were too time-consuming and/ or unrealistic to complete within one class period.

- Differentiating instruction for students was helpful with bold statements in the textbook that highlighted key concepts. However, this may facilitate too much reliance on getting to the main idea and not enough practice in thinking critically about the main idea and developing one's own conclusion from the material. (This idea is supported by learning research in that to advance critical thinking skills students must make inferences about the content and avoid "spoon-feeding" the main points.)

- Teachers reported that they were happy with several aspects of the study such as receiving a wide variety of resources and communicating with researchers. Teachers reported that they were able to adhere to most implementation guidelines with the exception of Analyzing Data and the Pacing Guide. They reported that Analyzing Data exercises may or may not have been relevant to class lessons. These teachers also used their district pacing guide instead of the Prentice Hall pacing guide.

- One teacher reported technical difficulties with using the I-text, while the other teacher never attempted to use this product.

- One teacher also had technical difficulties using the test bank and selecting out non-desired questions.

- Favorite ancillary materials included Issues and Decision-Making, the Guided Reading and Study workbook, and the Lab Manual.

- Teachers reported that they liked the content of the overhead slides, however, they did not like the fact that all slides were together in one box absent labels. One suggestion was to have slides available in a binder presentation with appropriate tabs so teachers can easily find needed slides.

Changes to Randomized Control Trial Study

Teacher training

We feel that some of the results from the teacher logs and feedback have indicated that teachers needed more adequate training to be able to use various technologies with the textbook. In addition, training occurred close in time from the start of the new school year. To address these issues, we have established more rigorous training protocols for teachers participating in the RCT. In addition, we have ensured that all participating teachers have received a copy of the Teacher's Edition textbook as soon as we had a firm commitment of their participation in the study. Also, the content of the initial training was focused much more on introduction of various products than on pedagogy. Since we contend that understanding the pedagogy is a critical piece for student success, we have focused on making this a key element in training. We have also established follow-up training to ensure that teachers understand how various technologies work so they can get the best use from them.

Implementation Guidelines

It is clear from an analysis of the teacher logs that all implementation guidelines set forth prior to the pilot study were difficult to adhere to for the treatment teachers. Although they reported in interviews that they were able to address all the study guidelines, it was clear from their self reports in the weekly implementation logs that this was not the case. In addition, after further analysis of the guidelines and a discussion of study priorities, it was decided that new implementation guidelines should be developed that focus on the Miller & Levine text as the primary source of information. In addition, because the focus of the design of the Miller & Levine text is based on principles differentiating instruction for a variety of skill levels, new guidelines were developed with most salient features of the text in mind. The new implementation guidelines can be found in Appendix 2.

Student Survey

Based on the results of our factor analysis, we eliminated items that did not map onto specific constructs being tested. The language of some of the items was modified that might have been confusing or difficult for some students to answer. The modified survey has fewer questions and takes less time to administer. In addition, we have ensured that all teachers have received surveys to administer during the first week of class to capture a true pre-test score.

Standards-Based Science Assessment

Despite the fact that there was significant growth from the pre-test to post-test for both groups, we are still somewhat unsettled that mean post-test scores were below 50%. To address this, we revisited the content of the assessment itself. We analyzed reliability coefficients for each question included on the science assessment. Items that are considered reliable yield coefficients of between .40 and .70. It was clear from this analysis, that there were specific items that did not meet this requirement. In addition, upon further examination of the instrument we felt that the language of some questions was too difficult for lower level readers to interpret. We concluded this based on our own analysis of the difficulty level of the assessment as well as teacher comments that indicated that the assessment may not be a reflection of their students' actual knowledge due to confusing language. Given the issues with the assessment, we worked with the test publisher to modify the assessment to create an instrument that will more easily assess knowledge of life science concepts, but is not too difficult for lower level readers. In fact, the new science assessment to be used for the RCT study will include the modified enhanced multiple choice items as well as one constructed response item written at the 8^{th} grade reading level.

Teacher implementation log

The teacher log will reflect our expanded knowledge of the activities and materials used in participating classrooms from the pilot study. For example, we will no longer request information regarding only "Quick Labs" and "Extended Labs" for treatment classrooms. Instead, we will request information on specific activities in the Miller & Levine (2006) text such as “Real World Lab”, “Design an Experiment”, or “Exploration.” In addition, we have mapped the entire book to create a pull-down menu for teachers to select from instead of explanation of lessons and individual chapter sections. This allows teachers to complete the logs quickly and easily and provides a better system of data analysis for the researchers.

Attrition Rate

At the outset of the study, we estimated that more than 350 students would participate in all aspects of the project. After some students voluntarily withdrew from the study, we proceeded with pre-testing with the student survey and assessments for as many students as possible. From the time of the pre-test to the post-test, many participating students had dropped or switched classes, and did not take the post-test assessments. Data for the post-test was obtained from a total of 299 students, less the 40 students that we dropped from the 10^{th} and 11^{th} grades. After matching all pre-test and post-test data for all measures, we were left with a total of 163 students with complete data on all measures. While this attrition rate caused some concern initially, we have addressed this by developing a system to minimize attrition for the RCT study.

Conclusions

- There were few differences observed between the treatment and control groups on science achievement and attitudes towards science. There were also few differences between other groups such as males and females and various ethnic groups.

- Where there were techinically significant differences between treatment and control group teachers we hesitate to distinguish these results as specifically a “treatment effect” primarily because groups were self selected and one teacher in the treatment group clearly had a higher level of program implementation than the other.

- There were differences observed in terms of teacher implementation of the Miller & Levine (2004) *Biology* program. We also observed varying levels of cognitive demand in classrooms during site observation visits, although overall participating teachers were similar.

- Students in the treatment group rated textbook satisfaction higher than those students in the control group. Specific features they noted were graphs and charts in the text as well as the book organization which allows for easily identifying key points.

- Several modifications will be made for the RCT study to be conducted in 2005-06. These modifications include revision of instruments such as student surveys, science assessments, and teacher implementation logs.

Appendix 1: Standards-Based Science Assessment (PASS)

The following chart describes how enhanced multiple-choice items by form number on the Biology assessment corresponds to the California Standards and National Standards.

	California Standards		National Standards (NSES)	
Item #	**Form 1**	**Form 2**	**Form 1**	**Form 2**
1	1a	1a	LS 1.2	LS 1.2
2	1c	1c	LS 3.1	LS 3.1
3	1f	1f	SI 1.3	LS 1.5
4	1f	1f	SI 1.3	SI 1.3
5	1f	1g	LS 1.5	LS 1.2
6	1g	3a	LS 1.2	LS 2.2
7	3a	4b	LS 2.2	LS 2.1
8	4c	4c	LS 2.3	LS 2.3
9	4b	4d	LS 2.1	LS 1.3
10	6a	I&E d	LS 3.1	SI 1.2
11	6b	5c	LS 4.4	ST 2.2
12	6c	6a	LS 4.3	LS 3.1
13	6d	6b	LS 5.2	LS 4.3
14	6f	6c	LS 4.2	LS 4.3
15	6b	6d	LS 4.5	LS 5.2
16	8b	6f	LS 3.1	LS 4.2
17	8d	7c	LS 3.2	LS 3.3
18	8e	8b	SI 1.2	LS 3.1
19	9a	8d	LS 1.4	LS 3.2
20	9a	83	LS 6.2	LS 2.3
21	9c	9a	SI 2.6	LS 5.3
22	10b	9a	LS 1.1	LS 6.2
23	10c	9c	LS 2.3	SI 2.6
24	I&E j	10b	SI 1.2	LS 1.1
25	I&E a	10c	SI 1.3	LS 2.3
26	I&E c	I&E j	HNS 1.2	SI 1.2
27	I&E d	8e	LS 2.3	SI 1.2
28	I&E d	I&E d	SI 1.2	HNS 1.2
29	I&E g	I&E d	SI 5.6	LS 2.2
30	I&E I	I&E n	LS 3.2	ST 2.4

Appendix 2: Prentice Hall Biology Implementation Guidelines

Implementation Guidelines Used in the Pilot Study (Miller & Levine, 2004):

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

- Daily use of Miller & Levine's (2004) *Prentice Hall Biology* text
- 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)
- Start each chapter with "Inquiry Activity"
- Cover all Key Concepts in chapter
- Cover some "Checkpoints", and Caption Questions throughout each chapter
- Perform most "Quick Labs"
- Complete all "Section Assessments"
- Perform "Analyzing Data" exercises where applicable
- One full lab per chapter (not necessarily from book, could be from lab manual)
- Assign at least one writing assignment at least once every two weeks, preferably "Writing in Science" response
- Complete parts of all end of chapter assessments
- Use Guided Reading and Study Workbook (currently available) or Adapted Reading and Study Workbook (for lower level readers, will be published this fall)
- Recommended: Standardized Test Prep in end of chapter assessments

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