



THE EFFECTS OF PRENTICE HALL *LITERATURE (PENGUIN EDITION)* CURRICULUM  
ON STUDENT PERFORMANCE: RANDOMIZED CONTROL TRIAL  
Final Report

Submitted by:

CLAREMONT GRADUATE UNIVERSITY  
Institute of Organizational and Program Evaluation Research

September 28, 2007

Tiffany Berry, Ph.D. & Rebecca M. Eddy, Ph.D.,  
Dreolin Fleischer, M.A., Mitra Asgarian, M.A., Yassaman Malek, M.A.

With Special Acknowledgement to:  
Xiaoxia Newton, Ph.D., University of California, Berkeley

123 East Eighth Street  
Claremont, CA 91711  
(909) 607-1540  
(909) 621-8905 fax

# TABLE OF CONTENTS

TABLE OF CONTENTS .....	2
TABLES AND FIGURES .....	5
ACKNOWLEDGEMENTS .....	6
ABSTRACT .....	7
SECTION ONE: BACKGROUND AND PURPOSE .....	9
Introduction .....	9
Purpose of the Prentice Hall <i>Literature (Penguin Edition)</i> RCT .....	9
Description of the Prentice Hall <i>Literature (Penguin Edition)</i> Curriculum .....	10
SECTION TWO: DESCRIPTION OF STUDY SETTING AND SAMPLE .....	14
Site Recruitment .....	14
School Demographics .....	15
<i>California High School-Site 1</i> .....	15
<i>California Middle School-Site 2</i> .....	15
<i>California Middle School-Site 3</i> .....	16
<i>Colorado High School-Site 4</i> .....	16
<i>Colorado Middle School-Site 5</i> .....	16
<i>Illinois High School-Site 6</i> .....	17
<i>Illinois Middle School- Site 7</i> .....	17
Study Sample .....	19
Sample Attrition .....	22
Comparability of Groups by Teacher Characteristics .....	24
SECTION THREE: DESCRIPTION OF EVALUATION DESIGN, MEASURES, AND PROCEDURES .....	26
Evaluation Design .....	26
Implementation Guidelines .....	26
Teacher Compensation .....	27
Teacher Training .....	27
<i>Research Study Orientation</i> .....	27
<i>Treatment Teacher Product Training</i> .....	28
Implementation Measures .....	28
<i>Teacher Implementation Log</i> .....	28
<i>Classroom Observation</i> .....	29
<i>Teacher Interview</i> .....	30
<i>Treatment Teacher Survey</i> .....	30
Outcome Measures .....	30
<i>English Language Arts Assessment</i> .....	31
<i>Writing Assessment</i> .....	31
<i>Student Survey</i> .....	32
Summary of the Data Collection Procedures across Sites .....	33

SECTION FOUR: ASSESSMENT OF CURRICULUM IMPLEMENTATION.....	35
Description of Treatment Curriculum Implementation.....	35
Adherence to Treatment Implementation Guidelines.....	39
Description of Control Curricula Implementation.....	43
Alignment of Treatment to Control Textbooks.....	44
Comparing Coverage Across Treatment and Control Groups.....	44
<i>Reading Selections Coverage Across Treatment and Control Groups</i> .....	44
<i>Skills Coverage Across Treatment and Control Groups</i> .....	46
Comparing Teacher Quality Across Treatment and Control Groups.....	46
<i>Teacher Quality</i> .....	46
<i>Teacher Influence</i> .....	48
<i>Differentiated Instruction</i> .....	48
 SECTION FIVE: RESULTS RELATED TO STUDENTS' ATTITUDES AND ACHIEVEMENT	
IN ENGLISH LANGUAGE ARTS.....	50
Differential Attrition.....	50
Performance on Outcome Measures.....	51
<i>ITBS and ITED Percentile Rank Scores</i> .....	51
<i>Writing Assessment Scores</i> .....	53
<i>Student Survey</i> .....	55
Differences in English Language Arts Achievement Between Treatment and Control	
Groups.....	55
Implementation and Achievement in English Language Arts.....	60
Student Attitudes towards English Language Arts and Achievement.....	61
 SECTION SIX: PRODUCT SATISFACTION.....	63
Teachers' Ratings of Product Satisfaction with Core Textbook Components.....	63
<i>Grade 7 Teacher Perceptions of Breadth and Depth</i> .....	65
<i>Grade 9 Teacher Perceptions of Breadth and Depth</i> .....	65
Students' Ratings of Product Satisfaction by Group.....	66
Favorite and Least Favorite Reading Selections.....	67
Topics Noted by Teachers and Students as "Needing Improvement".....	69
 SECTION SEVEN: DISCUSSION.....	73
Major Findings.....	73
Product Satisfaction.....	75
Study Limitations.....	77
Future Research.....	78
 REFERENCES.....	80
 Appendix A. Skill Areas Covered in Treatment Textbooks.....	81
Appendix B. Information for Interested Schools.....	82
Appendix C: Variables Explored during Data Analysis.....	86

Appendix D: HLM Final Model-Grade 7 ITBS Reading .....	88
Appendix E: HLM Final Model-Grade 7 ITBS Language .....	90
Appendix F: HLM Final Model-Grade 9 ITED Reading Total .....	92
Appendix G: HLM Final Model-Grade 9 ITED Spelling .....	94

## TABLES AND FIGURES

Table 1. School Level Demographics for Participating Sites.....	18
Table 2. Sample Size by Grade Level and by Group.....	20
Table 3. Study Sample Demographics of Participating Students.....	20
Table 4. English Language Arts Course Attendance by Grade Level and by Group.....	21
Table 5. Percentage of Students Who Did Not Have Posttest Scores on the ITBS or ITED	22
Table 6. Key Demographic Characteristics: Treatment vs. Control Taking Missing into Account.....	23
Table 7. Key Demographic Characteristics: Original Sample vs. Sample with Posttest.....	24
Table 8. Teacher Characteristics Across Treatment and Control Groups.....	25
Table 9. Timelines for Data Collection and Activities Across Sites.....	34
Table 10. Breadth of Implementation Coverage for Grade 7 Treatment Teachers.....	36
Table 11. Breadth of Implementation Coverage for Grade 9 Treatment Teachers.....	37
Table 12. Grade 7 Treatment Teacher Fidelity to Implementation Guidelines.....	40
Table 13. Grade 9 Treatment Teacher Fidelity to Implementation Guidelines.....	41
Table 14. Examples of Teacher Practices that Represent Scores on the Classroom Management and Teacher Quality Scales.....	47
Table 15. HLM Results for Grade 7 Reading and Language.....	57
Table 16. HLM Results for Grade 9 Reading and Spelling.....	59
Table 17. Level of Program Implementation and Student Achievement Scores.....	60
Table 18. Overall Product Satisfaction of the Prentice Hall <i>Literature (Penguin Edition)</i> Core Curriculum Components.....	64
Table 19. Students' Textbook Product Satisfaction as a Function of Group.....	66
Table 20. Students' Textbook Product Satisfaction for Treatment Students.....	67
Table 21. Teachers Report of Students' Favorite and Least Favorite Grade 7 Reading Selections.....	68
Table 22. Favorite and Least Favorite Grade 9 Reading Selections.....	69
Figure 1. English Language Arts Achievement Pretest and Posttest Scores (Grade 7).....	52
Figure 2. English Language Arts Achievement Pretest and Posttest Scores (Grade 9).....	53
Figure 3. Writing Assessment Pretest and Posttest Scores (Grade 7).....	54
Figure 4. Writing Assessment Pretest and Posttest Scores (Grade 9).....	54

## ACKNOWLEDGEMENTS

We would like to thank all of the participating schools, principals, teachers, and students who generously provided their time and effort to participating in this study. We appreciate you allowing us into your schools and classrooms to observe you in action! A special thanks to our exemplary research team at Claremont Graduate University is warranted, for the contribution, dedication, and persistence of Dreolin Fleischer, Mitra Asgarian, and Yasie Malek was truly remarkable on all accounts. Several other graduate students also offered invaluable assistance on this project: Jennifer Pinzon, Garrick Wahlstrand, Nancy Aguirre, Carli Straight, Linda Meyer, Stacy Hawkins, Kimaada Brown, and Sarah Sadeghi. Thank you to the two teachers from Torch middle school, Shirley Burns and Wendy Anson, who provided content expertise in Language Arts at the beginning of this research effort. We would also like to thank Dr. Xiaoxia Newton for her expertise in advanced statistics and, in particular, hierarchical models. We would like to thank the Penguin authors that visited our schools and provided excellent presentations to participating students and teachers, including Gary Soto, Pat Mora, and John Phillip Santos. We also sincerely appreciate everyone from Pearson Education including Rebecca Graziano, John Ade, Megan Chill, 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. Tiffany Berry & Dr. Rebecca Eddy

Institute for Organizational and Program Evaluation Research  
School of Behavioral and Organizational Sciences  
Claremont Graduate University  
123 E. Eighth Street  
Claremont, CA 91711  
(909) 607-1540

## ABSTRACT

**Citation:** Berry, T., & Eddy, R.M., Fleischer, D., Asgarian, M., & Malek, Y. (2007). The Effects of Prentice Hall *Literature (Penguin Edition)* Curriculum On Student Performance: Randomized Control Trial 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 Prentice Hall *Literature (Penguin Edition)* 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 students' English language arts achievement.

**Purpose:** The purpose of this study is to conduct a Randomized Control Trial (RCT) on Prentice Hall's *Literature (Penguin Edition)* curriculum. We randomly assigned teachers to a treatment group (using the new Prentice Hall curriculum) or to a control group (using existing Literature curriculum currently in place at school). In particular, we were interested in examining: (1) How do student outcomes differ for students using the Prentice Hall *Literature (Penguin Edition)* program compared to other literature programs?; (2) How do student outcomes differ for treatment and control students as a function of their background characteristics (e.g., gender, ethnicity, etc.) and teacher characteristics (e.g., experience)?; (3) What is the relationship between program implementation and student achievement in reading and language arts?; and (4) What is the relationship among students' attitudes toward reading and language arts and their achievement in reading and language arts (e.g., vocabulary, reading comprehension, spelling, language and writing skills)?

**Setting:** This study took place in three high schools and four middle schools across three states (California, Colorado, and Illinois). The schools ranged in size and location. The study was conducted over the 2006-2007 school year.

**Study Sample:** Thirty-one teachers (15 treatment and 16 control) and 1,722 students participated in the study. This sample included a total of 922 students in the treatment group (424 in grade 7 and 498 in grade 9) and 800 students in the control group (349 in grade 7 and 451 in grade 9).

**Intervention:** Teachers in the treatment group were provided with textbooks and ancillary materials for themselves and their students during the entire 2006-2007 school year. 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 Literature 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 each participating school. Multiple process measures were collected throughout the year and three outcome measures (English language arts achievement, writing achievement, and student attitudes survey) were collected at pretest and posttest.

**Control Condition:** Six distinct English Language Arts curriculum were used in the control group across seven sites. Two sites used an earlier version (2002) of a competing Prentice

Hall textbook, while the remaining sites used five different editions of competitors' textbooks with publication dates ranging from 1985 to 2002.

**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 *Literature (Penguin Edition)* curriculum.

**Findings:** Findings suggest that students in the treatment group did not outperform students in the control group, regardless of grade level or English Language Arts achievement outcome. As anticipated, several student survey composites (e.g., self-efficacy, motivation, etc.) did significantly predict reading achievement, which is consistent with the research literature (see Baker & Wigfield, 1999, as an example). Also, African American grade 9 students and Latino grade 7 students had lower reading achievement than students across other ethnicities. Unfortunately, there were also no cross-level interactions between ethnicity and group; ethnic minorities in the treatment group did not outperform ethnic minorities in the control group at either grade level.

**Conclusion:** Students using the Prentice Hall *Literature (Penguin Edition)* curriculum had comparable achievement to students using competitor English Language Arts curricula. Implementation analyses revealed that differences in implementation favored the treatment group over the control group and grade 7 teachers over grade 9 teachers. The treatment group generally provided high satisfaction ratings for the Prentice Hall Literature textbooks and ancillary materials when compared to the control group. Similarly, grade 7 treatment teachers and students rated the curriculum much more favorably than grade 9 treatment teachers and students. Taken together, this study suggests that the *Literature (Penguin Edition)* curriculum may be much more conducive to lower grade levels than higher grade levels.

## SECTION ONE: BACKGROUND AND PURPOSE

### Introduction

The need to determine whether educational curricula and products are effective at improving student achievement has continued to grow throughout the past decade. Since the *No Child Left Behind Act of 2001* (NCLB, 2001), 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...” (What Works Clearinghouse<sup>1</sup>). Drs. Tiffany Berry and Rebecca Eddy from Claremont Graduate University (CGU) were hired by Pearson Education to conduct a research study that meets the highest standard of “what works” in Language Arts curriculum by implementing a Randomized Controlled Trial (RCT) to test the efficacy of the Prentice Hall *Literature (Penguin Edition)* curriculum.

### Purpose of the Prentice Hall *Literature (Penguin Edition)* RCT

The purpose of this study was to assess the extent to which Prentice Hall’s *Literature* curriculum impacted student learning in English Language Arts in grades 7 and 9. The current version of the text was visibly aligned with state and national reading and language arts standards. The *Literature (Penguin Edition)* textbook was also designed to offer differentiated instruction that was linked to identical content, to introduce a new scope and sequence of reading skills, and to offer an emphasis on the writing and revision process through segments written by Penguin authors. Given these innovations, the *Penguin Edition* curriculum differed markedly from the earlier version of the Prentice Hall Literature series (i.e., *Timeless Voices, Timeless Themes*). Thus, this study provided the first empirical test of the Literature curriculum and was intended to supply Prentice Hall with valuable feedback that may be useful for future revisions.

The RCT study addressed the following four research questions (each developed collaboratively by CGU and Prentice Hall):

- How do student outcomes differ for students in grades 7 and 9 using the *Literature (Penguin Edition)* program compared to other literature programs?

---

<sup>1</sup> <http://www.whatworks.ed.gov/>

- How do student outcomes differ for treatment and control students as a function of their background characteristics (e.g., gender, ethnicity, etc.) and teacher characteristics (e.g., experience)?
- What is the relationship between program implementation and student achievement in reading and language arts?
- What is the relationship among students' attitudes toward reading and language arts and their achievement in reading and language arts (e.g., vocabulary, reading comprehension, spelling, language and writing skills)?

Ultimately, the purpose of the proposed research study was to address each of the research questions related to the Prentice Hall *Literature* program. In addition to the RCT study, a modified pilot study was conducted (from May, 2006 – August, 2006) to test all materials CGU developed for use in the RCT with a sample of students similar to the RCT target population. The pilot study also provided feedback about the study protocols and measures from current English Language Arts (ELA) teachers. The current report will exclude the results of the pilot study, as a separate report detailing the pilot study was delivered to Prentice Hall in December, 2006<sup>2</sup>.

### **Description of the Prentice Hall *Literature (Penguin Edition)* Curriculum**

The *Literature (Penguin Edition)* textbooks were structured the same for grade 7 and grade 9. Each book had six units. Within each unit there was an introduction and then two unit parts that were identical with regard to organization. Within each unit introduction, a Penguin author introduced the unit according to a particular theme. For example, Laurence Yep introduced the theme “exploring drama” in a section entitled *From the Author’s Desk* in Unit 5 of the grade 7 *Literature (Penguin Edition)* textbook. A short biography about the Penguin author was provided and then students were given more information about the main topic (i.e., literature form) of the unit. Students were given an opportunity to check their understanding of the unit introduction in a section entitled *Check Your Understanding*. There was at least one model selection in every unit introduction. The model selection was always a reading selection written by the Penguin author, which the author introduced and then answered questions at the conclusion of the selection. There was also a *Writing Workshop*:

---

<sup>2</sup> Please refer to the report “*A Modified Pilot Study of Prentice Hall Literature Penguin Edition (2007)*” for a completion description of the pilot study and its findings.

*Work in Progress* section that gave students an opportunity to practice writing. The final section of the unit introduction was the *Apply the Skills* section. At this point students were primarily asked to recall and respond to the reading selection they just read.

After the unit introduction, both part 1 and part 2 of each unit were structured the same. The unit part began with a *Reading and Vocabulary Skills Preview*, followed by *Build Skills* page related to the paired reading selections, which were unique to the *Literature (Penguin Edition)* textbook. Teachers could select which of the two reading selections they wanted their students to read. Each reading selection varied in level, so teachers could choose to give the lower-level students in their class one story and the higher-level students in their class the other story. Both stories covered the same skills; this helped ensure that students were given the same opportunity to learn identical skills even if they did not read the same reading selection. The *Build Skills* page included a Reading Skill (e.g., predicting), review of Literary Analysis (e.g., plot), and finally provided a Vocabulary Builder section that was intended to introduce vocabulary words students would encounter in the paired reading selections. Next, there was a *Build Understanding* page for the first reading selection that included a background piece of information related to the story, a quick reading/writing connection activity, and then a short biography of the author.

The model selection included support notes in the margins that connected to the Reading Skill, Literary Analysis, and Vocabulary Builder introduced on the *Build Skills* page. A few small pictures were placed in the margins that went to help to promote critical viewing. A box at the end of the reading selection discussed a social studies connection in a section entitled *Literature in Context*. After the reading selection was finished there was an *Apply the Skills* page, which included a subsection entitled *Thinking About the Selection*, as well as the components from the *Build Skills* page (i.e., Reading Skill, Literary Analysis, Vocabulary Builder) and two new sections entitled Writing and Extend Your Learning. This cycle was repeated with the second of the two paired reading selections and then again with another pair of reading selections. The drama and poetry unit were grouped slightly differently because these units included a number of acts or poems respectively, not just pairs of reading selections.

In addition to the paired reading selections were sections on *Reading Information Materials*, *Comparing Literary Works*, *Writing Workshop*, *Monitor Your Progress*, and at the end of the second unit part various workshops (e.g., spelling, writing, communications) and further reading related to the unit or a section for students to do on their own (intended for

higher-level students who may move quickly through the materials and want to do something more).

A separate skill area was covered in each unit part for a total of 12 overarching skill areas per textbook. A full description of the skill areas by grade and unit part can be found in Appendix A. The skill areas were similar across grades, covering the same skills (e.g., cause and effect, compare and contrast, author's purpose, etc.), however there were three skill areas distinct to each grade level. For grade 7 the distinct skill areas were analyzing persuasive texts, reading fluently, and cultural and historical context. For grade 9 the distinct skill areas were context clues, fact and opinion, and purpose for reading.

The *Literature (Penguin Edition)* curriculum included not only a student edition and teacher edition textbook, but also a number of ancillary materials. For example, students were given Reader's Notebooks (either a regular Reader's Notebook, an adapted version of the Reader's Notebook, or an English learners version). The Reader's Notebooks provided students with reading support for each reading selection in the student edition textbook. There was also a Reading Kit, which provided additional strategies and support for struggling readers and a Skills Development Workbook that provided practice and reinforcement of skills. Teachers were provided with Teaching Resources, also called Unit Resource Books (one book per unit), which provided teachers tools to differentiate instruction. A General Resources booklet that provided timesaving resources for assessment and class management was also provided to participating teachers. In addition, there were several audio/visual resources (e.g., Listening to Literature Audio CD, Spanish/English Summaries Audio CD, From the Author's Desk DVD, Teacher Express CD-ROM, etc.) and online resources (e.g., Teacher Online Access Pack, Success Tracker, etc.) for teachers and students to use.

After consultation with the Prentice Hall literature editorial and marketing staff, several components were identified as features that distinguished the *Literature (Penguin Edition)* curriculum from other competing language art curricula including:

- The collaboration with Penguin authors who introduced each unit part and provided insight into their writing process and lives, as well as provided model selections for the unit introductions.
- The inclusion of paired reading selections, which encouraged teachers to differentiate instruction. Students were able to learn the same content and skills using differentiated materials appropriate for their diagnosed level. For example, to select an appropriate paired selection, teachers were supposed to use the diagnostic

test to determine the reading level of each student and then decide which of the paired reading selections would best suit each student. The teacher edition of the textbook provided an *Accessibility at a Glance* table prior to each paired reading selection that provided details concerning the context, language, concept level, literary merit, lexile score, and overall rating.

- The scope and sequence of reading skills were geared toward mastery. In addition, the materials encouraged reinforced learning (i.e., the textbook details what the students are going to learn, teaches them, then asks them to do activities that use what they have learned).
- Built in progress monitoring allowed teachers to assess specific skills that need to be retaught and then provided the materials to reteach those particular skills without reteaching the entire unit.
- The formatting was unique with support notes in the margin (rather than interrupting the text on the page) and single column reading selections.
- Focused instruction that targeted student mastery of skills through diagnosing, teaching, assessing, and reteaching.

## **SECTION TWO: DESCRIPTION OF STUDY SETTING AND SAMPLE**

### **Site Recruitment**

Schools in several large urban and suburban districts with the following characteristics were contacted: (1) ethnically and socio-economically diverse; (2) low student mobility rates; (3) willingness to randomly assign teachers to control or treatment groups; and (4) at least 750 enrolled students or a minimum of four teachers with multiple sections of college-preparatory English Language Arts. A team of three research assistants contacted over 500 schools and districts across the country beginning recruitment in early April and concluding in August. Strategies used to secure sites included targeting key states as identified by Prentice Hall; using an established recruitment protocol used in other studies; leveraging all previous and current contacts of the principal investigators, and following up with leads from the Prentice Hall sales force. Most districts/schools declined to participate in the study (even with over \$6,000 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 2006-07.

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., principal or district). When the contact person was at the district-level, a study information sheet (Appendix B) 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, researchers followed up with the principal, and then sought permission from the teachers. If the contact person was at the principal level, the researchers first secured support from a number of teachers and subsequently sought permission from the 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.

Seven schools across three states (California, Colorado, and Illinois) agreed to participate. Although the generalizability of this study was compromised given that the seven schools were not randomly selected out of all eligible schools, these sites were the only

sites contacted that were willing to participate in the research study and fully implement the curriculum during 2006-07.

### **School Demographics**

As indicated in the previous section, a total of seven schools across three states participated in the RCT study. The following provides a brief snapshot of each participating school.

#### *California High School-Site 1*

The California site was a large secondary school serving approximately 2,800 students in grades 9 through 12. The school resided in a suburban community located approximately 45 miles east of Los Angeles. Thirty-four percent of the school community was Caucasian/White while the Latino population comprised 32% of the school. The median household income was approximately \$66,000. Ninety-six percent of teachers held a full teaching credential; less than one percent held emergency credentials. On average, the student -teacher ratio was 20 to 1. Eight percent of students were eligible to receive free or reduced lunch and three percent of students were English Language Learners (ELL's). Based on the 2005-2006 state standardized test scores, this school was considered average given that the number of students classified as proficient or advanced was 57% for English Language Arts, 31% for Mathematics and 44% for Science.

#### *California Middle School-Site 2*

The first California Middle School site was a small middle school serving about 500 students in grades 7 and 8. The school resided in a suburban community located approximately 80 miles east of Los Angeles. Forty-four percent of the school was Latino, while 45% of the school was comprised of Caucasian students. The median household income was approximately \$32,000. The school building was originally housed in a high school; the 2006-2007 school year was the first year it was converted into a middle school. The English Language Arts classes were unique because it was the only subject at the school which was taught in 90 minute intervals. Sixty-four percent of the students were eligible to receive free or reduced lunch. The teaching staff was fairly new given that, on average, teachers had six years of experience and 10% teachers were in their first year. All of the teachers held full teaching credentials. On average, the student -teacher ratio was 30 to 1. Twelve percent of the students were English Language Learners. Based on the 2005-2006 state standardized test scores, this school's performance was considered

average/slightly below average given that the number of students classified as proficient or advanced was 40% for Language Arts and 34% for Mathematics.

#### California Middle School-Site 3

The second California Middle School site was also a small middle school serving about 800 students in grades 7 and 8. The school resided in a suburban community located approximately 80 miles east of Los Angeles. Forty-five percent of the community was Latino and earned an average household income of approximately \$32,000. Fifty-one percent of the students were eligible to receive free or reduced lunch. The teaching staff was fairly experienced given that, on average, teachers had 17 years of experience and there were no first year teachers. One hundred percent of teachers held teaching credentials. The English Language Arts classes were unique because it was the only subject at the school which was taught in 90 minute intervals. On average, the student -teacher ratio was 31 to 1. Based on the 2005-2006 state standardized test scores, this school's performance was considered average given that the number of students classified as proficient or advanced was 49% for Language Arts and 38% for Mathematics. Twelve percent of students were English Language Learners.

#### Colorado High School-Site 4

The Colorado site was a large secondary school serving over 2,400 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 percent of students were Caucasian and only nine percent 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 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 2005-2006 state standardized test scores, this school's performance was considered above average in Reading and below average in Mathematics given that the ninth grade students classified as proficient or advanced was 70% and 28%, respectively. Six percent of students were English Language Learners.

#### Colorado Middle School-Site 5

The second Colorado site was a midsize middle school serving over 470 students in grades 7 and 8. The school resided in a suburban community located approximately 20

miles north of Denver. The school served primarily Caucasian students (60%) and had a median household income of approximately \$52,000. Thirty-nine percent of students were eligible to receive free or reduced lunch. The teaching staff was fairly well educated given that fifty-eight percent of teachers held at least a masters degree. Based on the 2005-2006 state standardized test scores, this school's performance was considered average given that the number of students classified as proficient or advanced was 58% in Reading and 34% in Mathematics. Fourteen percent of students were English Language Learners.

#### Illinois High School-Site 6

The Illinois site was a large secondary school serving over 2,100 students in grades 9 through 12. The school resided in a suburban community located approximately 20 miles north east of St. Louis, Missouri. The school served primarily Caucasian students (66%) and African-American students (32%). The median household income was approximately \$34,000. Twenty-nine percent of students were eligible to receive free or reduced lunch; zero percent of students were English Language Learners. The teacher student ratio was about 19 to 1. 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 54% in Reading and 50% in Mathematics.

#### Illinois Middle School- Site 7

The second Illinois site was unique as it was the first year that it operated as a middle school by merging two former middle schools into one location. Prior to 2006-07, the school building was previously the local high school, which was subsequently moved into a newly constructed campus approximately one mile away for the 2006-07 school year. Since the middle school was new, no school demographic data was available at the time of this report. Therefore, demographic data was averaged from the two middle schools that were merged into the new middle school. Based on these data, the participating middle school served approximately 1,500 students in grades 6 through 8. The school resided in a suburban community located approximately 20 miles north east of St. Louis, Missouri. The school served primarily Caucasian (60%) and African-American (38%) students; approximately 50% of the students were eligible for free and reduced lunch. The teacher student ratio was about 15 to 1. The organization of the school was unique in that the each floor of the school was a "team" with a different bell schedule. Based on the 2004-2005 state standardized test scores, this school's performance was considered slightly above average given that the number of students classified as proficient or advanced was 64% in Reading and 70% in Mathematics.

**Table 1. School Level Demographics for Participating Sites**

Demographic Characteristics		California			Colorado		Illinois		
<b>Site</b>		Middle School	Middle School	High School	Middle School	High School	Middle School*	Middle School*	High School
<b>Location</b>		Suburban/Rural	Suburban/Rural	Suburban	Suburban	Suburban	Suburban	Suburban	Suburban
<b>School Size</b>		539	830	2,794	471	2,491	863	648	2,167
<b>Ethnicity</b>	<b>Percent Caucasian</b>	44%	43%	34%	60%	70%	56%	67%	66%
	<b>Percent Latino</b>	45%	45%	31%	30%	22%	<1%	<1%	1%
	<b>Percent African American</b>	4%	4%	15%	3%	3%	43%	32%	32%
	<b>Percent Other Ethnicity</b>	7%	8%	20%	7%	5%	<1%	<1%	1%
<b>Economic Measure</b>	<b>Percent Eligible for Free &amp; Reduced Lunch</b>	64%	51%	8%	39%	9%	61%	43%	38%
<b>Community Measures</b>	<b>Age 25+ with College Degree</b>	18%	18%	37%	32%	25%	25%	33%	25%
	<b>Median Household Income</b>	\$32,254	\$32,254	\$66,668	\$52,117	\$52,570	\$34,253	\$50,069	\$34,253
<b>State Testing Results** (Test Used)</b>	<b>Percent Proficient/Advanced in Language Arts/Reading</b>	40% (CST)	49% (CST)	57% (CST)	58% (CSAP)	70% (CSAP)	58% (ISAT)	70% (ISAT)	54% (PSAE)
	<b>Percent Proficient/Advanced in Writing</b>	N/A	N/A	N/A	42% (CSAP)	47% (CSAP)	N/A	N/A	N/A
	<b>Percent Proficient/Advanced in Math</b>	34% (CST)	38% (CST)	31% (CST)	34% (CSAP)	28% (CSAP)	68% (ISAT)	72% (ISAT)	50% (PSAE)
	<b>Percent Proficient/Advanced in Science</b>	N/A	N/A	44% (CST)	N/A	N/A	73% (ISAT)	78% (ISAT)	49% (PSAE)

\*These two middle schools were merged into one location for the 2006-07 school year. Consequently, the 2005-2006 school year data provided in this table are from the individual IL middle school accountability report cards.

\*\*All 2005-2006 test results except for the IL high school, which are 2004-2005 test results.

Please Note: School size, ethnicity for all states but CO, and the state testing results data came from individual School Accountability Report Cards from the 2005-2006 school year. The economic and community measures came from [www.publicschoolreview.com](http://www.publicschoolreview.com), which uses the 2000 Census data.

Table 1 indicates the school level demographics for the seven participating sites. Across the seven schools, most were located in suburban areas. The middle schools ranged in size from 470 at the Colorado school to 1,400 at the Illinois school. All of the high schools were relatively large (greater than 2,000 students each). At the California schools there were an equal number of Caucasian and Latino students. The Colorado schools served primarily Caucasian students, while the Illinois schools served both Caucasian and African-American students. All schools served students who qualified for free and reduced lunch, however, the range was large. At the Colorado high school, only nine percent of students qualified for free and reduced lunch while at the Illinois schools, approximately 50% qualified. The median household income across all of the sites was comparable, although the California high school was about \$15,000 higher than the two Colorado sites (which were about \$20,000 higher than the rest of the schools in the study). Across all schools, the curriculum was implemented from September 2006 – June 2007.

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 *Literature* curriculum across diverse school conditions.

All participating classes were classified by the schools as college-preparatory/general English language arts classes, and site recruitment efforts focused on eliminating any courses that were considered honors/advanced or basic/remedial. For the content of curriculum coverage, all schools adhered to their own district and state curriculum standards. While there were some differences between states in terms of content standards that were expected to be covered during the year, the standards were identical for both treatment and control groups within each of the sites, and therefore this controlled any major variation in content coverage throughout the year. A more specific discussion of curriculum content can be found in Section Four of this report.

## **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. Twenty-eight students and/or parents over the seven 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. Table 2

summarizes the sample size by grade level and by group (i.e., treatment vs. control). The sample consisted of 773 grade 7 students (424 treatment; 349 control) and 949 grade 9 students (498 treatment; 451 control). These 1,722 students came from 31 teacher groups (thirteen 7th grade; eighteen 9th grade). There were an additional 200 students in our sample, however, we excluded all students if they (1) moved from treatment to control or vice versa or (2) if they entered the study after fall 2006. It is also important to note that the exact number of students that completed assessments varies depending on the outcomes of interest. For example, a student could have failed to complete one subtest (e.g., Reading comprehension on ITED assessment), yet could have completed all other subtests and measures (e.g., student survey and writing assessment) and therefore, is included as a member of the sample even though they lack complete data on every measure.

**Table 2. Sample Size by Grade Level and by Group**

		Treatment	Control	Total
Grade Level	Grade 7	424	349	773
	Grade 9	498	451	949

Table 3 illustrates the frequency distributions of students' gender, ethnicity, and primary language status by grade level and by group. A higher proportion of grade 7 students (55.7%) in the treatment group were female students than grade 7 students (46%) in the control group. Grade 9 students were comparable in terms of gender distribution in the treatment and control groups. In terms of student ethnic composition, Table 3 also indicates that a higher proportion of grade 7 students (43.6%) in the treatment group were Caucasian than in the control group (39%). The same was true for grade 9 students (50% Caucasian in treatment vs. 41% Caucasian in control). On the other hand, a higher proportion of grade 9 students (26.2%) in the control group were Latino than in the treatment group (18.3%). Finally, Table 3 indicates that most students considered English their primary language, and this was comparable for treatment and control groups, but there was a slighter higher percent of students in grade 7 whose primary language was English as compared with students in grade 9.

**Table 3. Study Sample Demographics of Participating Students**

Demographic Characteristics		Grade 7		Grade 9	
		Treatment (n=424)	Control (n=349)	Treatment (n=498)	Control (n=451)
		Percent (%)	Percent (%)	Percent (%)	Percent (%)
Gender	Male	44.3	54.0	47.9	46.6
	Female	55.7	46.0	52.1	53.4
Ethnicity	Caucasian	43.6	39.0	50.0	41.0
	Latino	26.7	26.1	18.3	26.2
	Multi-ethnic/Other	15.8	17.7	13.5	16.4
	African-American	9.0	10.0	11.0	12.0
	Asian	2.8	2.3	6.0	3.5
Primary Language	English	87.5	83.4	97.0	92.7
	Other	10.3	11.7	1.8	5.8
	Unknown	2.2	4.9	1.2	1.5

We also compared students' English language arts course attendance (percent of days attended school), which was fairly similar between the grade levels. These results can be found in Table 4.

**Table 4. English Language Arts Course Attendance by Grade Level and by Group**

		Treatment	Control
Grade Level	Grade 7	92.7%	93.4%
	Grade 9	94.9%	94.6%

In summary, while there were minor differences in relation to some of the demographic characteristics of participating students across treatment and control groups, we are reasonably confident that the groups are similar in composition. We have further tested this as it relates to distribution of pretest scores, the results of which can be found in Section Five of this report. In addition, of primary importance was to assess the extent to which student attrition may have impacted the characteristics of the study sample. Analyses related to demographic characteristics are provided in the following section, and analyses related to attrition as it may have impacted test scores can be found in Section Five.

## Sample Attrition

Sample attrition is defined as those students who completed pretest scores on the primary outcome measures (i.e., ITBS for grade 7 and ITED for grade 9), yet did not complete posttesting on these measures. As shown in Table 5, attrition rates were fairly low for across groups given that a total of 53 grade 7 treatment students (approximately 12.8% of the treatment group) and 44 control students (about 13.9% of the control group) did not have posttest scores. Similarly, a total of 38 grade 9 treatment students (approximately 8.2% of the treatment group) and 35 control students (about 7.9% of the control group) did not have posttest scores. A total of 1,488 students had complete pretest and posttest scores on the ITBS and the ITED. The remaining 64 students did not complete the pretest, but may or may not have completed the posttest assessments. While the two grade levels vary in their rates of missing posttest scores, the rates between treatment and control groups are consistent within grade level and were not significantly different between the treatment and control group for both grade 7 and grade 9.

**Table 5. Percentage of Students Who Did Not Have Posttest Scores on the ITBS or ITED**

		Students with Complete Pretest and Posttest (n=1,488)	Students Missing Posttest Only (n=170)	Total (n=1,658)
<b>Grade 7</b>	<b>Treatment</b>	361 (87.2%)	53 (12.8%)	414 (100%)
	<b>Control</b>	272 (86.1%)	44 (13.9%)	316 (100%)
<b>Grade 9</b>	<b>Treatment</b>	448 (92.2%)	38 (8.2%)	442 (100%)
	<b>Control</b>	407 (92.1%)	35 (7.9%)	486 (100%)

After examining the number of students in each group that were missing posttest scores, we further explored the demographic background characteristics of these students in both treatment and control groups. 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 who did not have posttest scores, we investigated whether treatment and control students differed with respect to key demographic characteristics. A summary of these data are presented in Table 6. Second, we compared the original sample to the final sample students who had posttest scores

across student demographic characteristics. It is worth noting that the exact number of students missing posttest scores varies in Tables 5 and 6. Students who were considered to have “missing” posttest scores on the primary outcome of interest (i.e., Reading subtest score for ITBS and ITED) are presented in Table 6 (n=1,430), whereas the number of students listed in Table 5 (n=1,488) reflects those students with complete pretest and posttest scores on any subtest of the ITBS or ITED.

**Table 6. Key Demographic Characteristics: Treatment vs. Control Taking Missing into Account**

Demographic Characteristics		Grade 7		Grade 9	
		Treatment (n=360)	Control (n=268)	Treatment (n=429)	Control (n=373)
		Percent (%)	Percent (%)	Percent (%)	Percent (%)
Gender	Male	42.8	52.2	48.5	45.3
	Female	57.2	47.8	51.5	54.7
Ethnicity	Caucasian	44.6	41.0	50.8	43.7
	Latino	27.4	28.4	18.2	25.5
	Multi-ethnic/Other	15.5	17.5	13.3	13.9
	African-American	8.3	7.8	11.0	12.6
	Asian	3.3	3.0	6.5	4.3
Primary Language	English	89.7	84.7	97.9	94.1
	Other	10.0	13.1	1.6	5.6
	Unknown	<1.0	2.2	<1.0	<1.0

As shown in Table 6, the treatment and control groups for grade 7 students were fairly comparable in terms of ethnic distribution and English as primary language after accounting for students that did not complete the posttest. The treatment group had a higher proportion of female students than male students, while the control group had slightly higher proportion of male students than female students. For grade 9 students, the treatment and control groups were relatively comparable in terms of gender distribution and proportion of students whose primary language was English. However, the treatment group had somewhat higher proportion of Caucasian but relatively lower proportion of Latino students than the control group. Taking into account those who had missing posttest scores,

the treatment and control groups for grade 7 students were fairly comparable in terms of ethnic distribution and English as primary language. The treatment group had a higher proportion of female students than male students, while the control group had a slightly higher proportion of male students than female students. However, gender did not turn out to be a statistically significant predictor of student outcomes, all things considered. Therefore, the slight difference in gender distribution among treatment and control groups is not of concern in this study.

**Table 7. Key Demographic Characteristics: Original Sample vs. Sample with Posttest**

		Original Sample	Sample with Posttest
Demographic Characteristics		(n=1,722)	(n=1,538)
		Percent (%)	Percent (%)
Grade Level	Grade 7	n=773	n=666
	Grade 9	n=949	n=872
Gender	Male	47.8	46.8
	Female	52.0	53.2
Ethnicity	Caucasian	43.8	45.3
	Latino	24.0	24.1
	Multi-ethnic/ Other	17.8	16.1
	African-American	10.6	10.3
	Asian	3.8	4.3
Primary Language	English	90.8	92.4
	Other	7.0	6.8
	Unknown	2.3	<1.0

Next, we ran exploratory statistical analysis to see whether the original sample and the final sample of those who had posttest scores differed in their key demographic characteristics (see Table 7). These analyses revealed no systematic differences between the two groups in terms of background characteristics. That is, the distribution of the key demographic characteristics between the treatment and control groups taking missing into account parallels that of the original sample. In summary, student attrition did not alter the nature of the study sample.

In addition to examining comparability in student characteristics across treatment and control groups, it was important to verify that teacher characteristics were also comparable. Although random assignment was theoretically supposed to equate all groups (student-level and teacher-level), with 31 teachers it was possible to have systematic differences in teacher characteristics across conditions.

**Table 8. Teacher Characteristics Across Treatment and Control Groups**

<b>Teacher Characteristics</b>	<b>Treatment Condition (n = 15)</b>	<b>Control Condition (n = 16)</b>
Highest degree attained	4 (27%) = M.A. 8 (53%) = T.C. 3 (20%) = B.A.	7 (44%) = M.A. 4 (25%) = T.C. 5 (31%) = B.A.
Number of years teaching K-12	12.5 yrs (range = 1-30)	6.1 yrs (range = 1-19)
Number of years teaching English Language Arts	11.9 yrs (range = 1-30)	5.5 yrs (range = .5-19)

M.A.=Master’s degree; T.C.=Teaching Credentials; B.A.=Bachelor’s degree

Table 8 illustrates that treatment teachers, despite random assignment, were different across groups on the following three teacher characteristics: highest degree attained, number of years spent teaching K-12, and number of years teaching ELA. Highest degree attained was the most equivalent of the characteristics with all teachers having at least attained a bachelor’s degree, but treatment teachers were slightly more likely to have their teaching credentials, while control teachers were slightly more likely to have their Master’s degree. Treatment teachers had a wider range of years spent teaching K-12. On average, treatment teachers had spent roughly twice as much time teaching K-12 than control teachers, yet control teachers were more highly educated. Similarly, treatment teachers, on average, spent just over twice as many years teaching ELA as control teachers did. These differences in teacher characteristics were taken into account when conducting statistical analyses related to differences in student achievement across groups.

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

### **Evaluation Design**

A Randomized Control Trial (RCT) was conducted to test the efficacy of the Prentice Hall Literature (*Penguin Edition*) curriculum on student attitudes and achievement in English Language Arts. Multiple process measures were collected throughout the year and three outcome measures (English Language Arts achievement, writing achievement, and student survey assessing attitudes towards Language Arts) were collected at pretest in fall 2006, and again at posttest in spring 2007 (the posttest also included items related to product satisfaction).

An equal number of teachers were randomly assigned to the treatment or control group within each site. At each of the seven sites, all of the college prep English Language Arts teachers were randomly assigned to either the treatment group (implementing the 2007 *Literature* English Language Arts curriculum) or were randomly assigned to the control group (using the existing curriculum). The assignment process involved randomly selecting teacher's names out of a 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 and to ensure consistency of implementation across sites, specific implementation guidelines were developed in collaboration between CGU researchers and Prentice Hall English Language Arts team. The implementation 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 *Literature (Penguin Edition)* 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 English Language Arts.

## **Teacher Compensation**

Compensation for participation in the study was a \$300 stipend for all teachers. Teachers who also served as the site liaison received an additional \$300. In addition, all teachers in the treatment group were provided with a teacher's edition textbook and all available ancillary materials. All participating teachers received a \$500 Pearson Education/Prentice Hall product voucher to use at the conclusion of the study. Each participating school received enough student edition textbooks for the number of participating students in the treatment group prior to the start of the 2006-2007 school year as well as enough textbooks for the number of students in the control group at the conclusion of the school year.

In addition to direct compensation, all schools were presented with an opportunity to have a Penguin author visit their school for a special presentation for students and teachers. All participating schools (with the exception of the California High School) took advantage of this opportunity. These visits generally involved the author meeting with teachers and school administrators, and then most participating treatment and control group students (in some cases, the entire school) would attend a one-two hour presentation by the author which included the author reading his or her work and a question and answer session with students.

## **Teacher Training**

Implementation of the RCT study began with teacher training held at all participating sites in August/ September 2006. 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 the product training from a Prentice Hall consultant (and/or the researchers). Training sessions were held at each respective school site at four of the seven individual schools. The two middle schools in California, Colorado (middle and high schools), and Illinois (middle and high schools) were combined at one location within each state for convenience.

### *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, 2006. 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.

### Treatment Teacher Product Training

After the first phase of the research study orientation session, all treatment teachers received 3-4 hours of product orientation with a Prentice Hall English Language Arts consultant at all sites (except for the IL sites in which the Principal Investigator provided the product development training). 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*). The consultant reviewed the curriculum, implementation guidelines, and all ancillary materials. The product website was also shown and teachers learned how to access additional resources online. A consultant returned to each site (except the California High School) to provide a question and answer session after the teachers had an opportunity to use the products in the classroom.

## **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 satisfaction survey. Since August 2006, 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 *Literature* curriculum throughout the study.

### Teacher Implementation Log

*Description of Measure:* We had three primary goals in developing an on-line teacher log system: (1) capture as accurately as possible both the content covered in their classes (e.g., unit, part, and paired selection) as well as the delivery of that content (e.g., read story aloud in class, read story independently, etc.); (2) reduce the strain on teachers by making the process user-friendly and efficient; and (3) collect data in a way that was meaningful to researchers and could be reported back easily. To that end, several revisions of the teacher

implementation logs were developed and pilot-tested until the best model emerged. Ultimately, four versions of the weekly implementation logs were developed for this study. This included two versions for treatment teachers (for both the 7<sup>th</sup> and 9<sup>th</sup> grade levels), as well as two versions for control teachers. The logs differed based on reading selections per grade level. The logs were developed using Remark survey software and were presented to teachers via an email message that directed them to a link to complete the correct log. In addition, 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. A copy of the teacher log protocol for treatment and control teachers is available in Appendices H and I, respectively.

*Implementation of Measure:* Participating teachers submitted online logs starting mid-September 2006 through the end of the curriculum implementation. At the end of each week, participating teachers were sent a reminder to complete their on-line log for that week. Researchers checked the status of these logs on a weekly basis to verify completion and appropriateness of responses. A few of the teachers consistently failed to complete their logs on time, at which point the researchers called and emailed the teacher as well as contacted the site liaison in an effort to encourage the teacher to complete his/her logs. In only a few numbers of cases did teachers fail to complete a log by the end of the study (less than five percent of teachers).

### Classroom Observation

*Description of Measure:* The researchers used an observation protocol that was developed and modified during the pilot study. Extensive training was provided to the observation team regarding how variables were operationalized (e.g., defining “student engagement”). The observation protocol assessed elements in the classroom such as curriculum materials used, physical environment, level of student engagement, allocation of classroom time, and effective teacher practices.

*Implementation of Measure:* Every participating classroom was observed by at least one member of the research team on one occasion during fall 2006, and again, in spring 2007. 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 *Literature* curriculum.

*Implementation of Measure:* In May – June 2007, all teachers were interviewed by phone 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. Interviews with treatment teachers lasted approximately 30 minutes, while interviews with control teachers lasted 15-20 minutes

### Treatment Teacher Survey

*Description of Measure:* An additional part of the debriefing process and product satisfaction assessment included an online teacher survey created using the Remark software. Questions on the teacher survey included background demographic information as well as elicited unit-specific feedback about the quality of the program, the depth of the content covered, and open-ended questions about the selections read within the unit.

*Implementation of Measure:* Sixty-seven percent of grade 7 treatment teachers (4 of 6) and 78% of grade 9 treatment teachers (7 of 9) completed the survey online via their personal computers from March – June 2007.

## **Outcome Measures**

There were three primary outcome measures in this study: English Language Arts assessment, writing assessment, and a student survey. 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., ELA 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. Although very few empirical studies on Language Arts curricula exist, it is clear that curriculum studies generally reveal small effect sizes (except for very targeted interventions to increase students' performance in a specified domain) (Datnow, Borman, & Stringfield, 2000). Therefore, the student survey provided a tool to measure the

impact of a curriculum on student factors that are often highly correlated with student achievement.

An ELA achievement test along with a writing assessment were the other two outcome measures used in this study. These were the primary instruments that measured the impact of the Prentice Hall *Literature* curriculum in comparison to the control curriculum. Descriptions of each measure (and their corresponding psychometric properties) are described below.

#### English Language Arts Assessment

*Description of Measure:* A standardized, nationally recognized reading and language arts assessment aligned to state and national standards was identified to measure student learning in reading and language arts. The Iowa Test of Basic Skills (ITBS) was administered to students in grade 7, while the Iowa Test of Educational Development (ITED) was administered to students in grade 9. Both tests were developed by the University of Iowa's College of Education and contain strong psychometric properties<sup>3</sup>. The ITBS measured students' skills in reading comprehension and language, while the ITED measured reading comprehension, vocabulary and spelling. Although assessed separately in the assessment, scores from the reading comprehension and vocabulary subtests can be appropriately combined to assess an overall rating of "reading" when analyzing scores from the ITED. These assessments included a varied set of items that captured a range of student abilities, not just students performing at grade level. National percentile rank scores were used as the dependent variable for each instrument.

*Implementation of Measure:* All students participating in the study were required to complete the assessment in fall, 2006 and again in spring, 2007. There were 1,658 students who completed the pretest for their grade level (e.g., ITBS in grade 7 and ITED in grade 9) and 1,488 students with complete data on both the pretest and posttest (809 treatment and 679 control).

#### Writing Assessment

*Description of Measure:* Given that the *Penguin Edition* curriculum emphasized the writing process, it was important to assess participating students' writing abilities at pretest and again at posttest. All participating students in both grades 7 and 9 had to write a timed persuasive essay. A persuasive prompt was selected because it was the type of writing

---

<sup>3</sup> Detailed technical information on the ITBS and the ITED can be found in the *Guide to Research and Development* published by Riverside Publishing. Please contact the Principal Investigator for more information.

aligned with most state content standards. The prompt was taken from the Iowa Writing Assessment. The essays were hand scored by trained scorers using a rubric provided by Riverside Publishing and a score was provided on each of the following quality dimensions: Content, Organization, Voice, and Convention.

*Implementation of Measure:* All students participating in the study were also required to complete the Writing Assessment in fall, 2006 and again in spring, 2007. Teachers were instructed to administer the Writing Assessment per the specific instructions outlined in the testing manual. There were 1,697 students completed the pretest Writing Assessment; 1,654 students completed the posttest Writing Assessment. Of this amount, there were a total of 1,459 students (815 treatment, 644 control) who completed the pretest and posttest Writing Assessment.

### Student Survey

*Description of Measure:* The pretest student survey included demographic questions, such as their gender, primary ethnicity, primary language, and parent's/caregiver's highest level of education. Next, students were asked to report the number of hours they spend each day, on average, doing a particular activity (e.g., watching television, playing video games, doing homework, playing a sport, etc.). Students then responded to 20 questions under the heading of "Attitudes Towards English Language Arts," seven questions under the heading of "Parent Involvement," and five questions under the heading of "Classroom Environment" using a five-point likert scale from 1 (Strongly disagree) to 5 (Strongly agree). At the end of the survey, students were asked to answer five true/false questions having to do with the writing process, as well as rate their like or dislike of certain literary genres (e.g., fiction, nonfiction, biography, etc.) on a five-point likert scale from 1 (Dislike a lot) to 5 (Like a lot). The posttest student survey included the same questions, with the exception of a few items that did not map onto the constructs of interest based on an Exploratory Factor Analysis (EFA) that was conducted on the pretest survey. Additionally, on the posttest survey, 10 questions were added that specifically assessed students' product satisfaction with the control and/or treatment curriculum.

To validate the instrument, an Exploratory Factor Analysis (EFA) was conducted separately on pretest and posttest survey responses, using the Principal Components 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 each revealed on both the pretest and posttest. However, the fifth factor on both surveys had only one item with a strong factor loading ("I study hard for English Language

Arts tests”), so that factor was discarded leaving four remaining factors. Based on the posttest EFA, 57% of the variance was explained by a rotated four-factor solution: *Intrinsic Motivation* ( $a = .87$ ), *Self-efficacy* ( $a = .71$ ), *Teacher Influence* ( $a = .82$ ), and *Parent Involvement* ( $a = .74$ ). Each of these constructs has received empirical validation in the research literature as well, especially in relation to motivation (Flink, Boggiano, Main, Barrett, & Katz, 1992) and self-efficacy (Schunk & Pajares, 2001) in predicting reading achievement in school. As predicted, each of these factors (except parent involvement) was significantly positively related to students’ English Language Arts achievement in grade 7 (Reading and Language scores on the ITBS) and in grade 9 (Vocabulary, Reading Comprehension, and Spelling on the ITED). These correlations, along with Cronbach’s alpha coefficients, help provide support for the validity and reliability of the survey instrument.

*Implementation of Measure:* All students participating in the study were required to complete the survey in fall, 2006 and again in spring, 2007. Researchers obtained student rosters from each school and provided each teacher with specific instructions for administration of assessments, including expected length of time and optimal conditions for administration. There were 1,672 students who completed the pretest student survey; 1,661 students completed the posttest survey. Of this amount, there were a total of 1,455 students (802 treatment, 653 control) who completed the pretest and posttest survey.

### **Summary of the Data Collection Procedures across Sites**

Table 9 summarizes the timeframes for the teacher training, implementation measures, and outcome measures for this study in Month and Year format. 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. With the exception of product training and classroom observations, all monitoring, implementation and outcome measures were conducted at the exact same time across treatment and control groups.

**Table 9. Timelines for Data Collection and Activities Across Sites**

Study Procedures	CA			CO		IL	
	HS	MS (1)	MS (2)	HS	MS	HS	MS
Product Training	August 2006	Sept. 2006	Sept. 2006	Sept. 2006	Sept. 2006	August 2006	August 2006
Teacher Log Training	Sept. 2006	Sept. 2006	Sept. 2006	Sept. 2006	Sept. 2006	Sept. 2006	Sept. 2006
Follow-up Product Training	N/A	October 2006	October 2006	October 2006	October 2006	October 2006	October 2006
Administration of Pretest Assessments	Sept. 2006	Sept. 2006	Sept. 2006	Sept. 2006	Sept. 2006	Sept. 2006	Sept. 2006
Administration of Posttest Assessments	May 2007	May 2007	May 2007	May 2007	May 2007	May 2007	May 2007
Fall Classroom Observation	Nov. & Dec. 2006	Nov. 2006	Nov. 2006	Nov. 2006	Dec. 2006	Nov. 2006	Nov. 2006
Author Visit	N/A	May 2007	May 2007	April 2007	April 2007	April 2007	April 2007
Teacher Interviews	June 2007	June 2007	June 2007	June 2007	June 2007	June 2007	June 2007
Spring Classroom Observation	April & May 2007	May 2007	May 2007	March & May 2007	May 2007	April 2007	April 2007
On-Line Logs	Sept. '06 -May '07	Sept. '06 -May '07	Sept. '06 -May '07	Sept. '06 -May '07	Sept. '06 -May '07	Sept. '06 -May '07	Sept. '06 -May '07

## **SECTION FOUR: ASSESSMENT OF CURRICULUM IMPLEMENTATION**

To interpret student outcomes appropriately, it was important to measure implementation within treatment and control classrooms. Through the classroom observations, teacher interviews, and online teacher logs, we were able to examine the depth and breadth of the content covered as well as the quality of implementation. The following section provides an analysis of the (1) implementation of the treatment curriculum (specifically focusing on the breadth of coverage and fidelity to implementation guidelines); (2) implementation in control classrooms; and (3) comparison between treatment and control classrooms.

### **Description of Treatment Curriculum Implementation**

The *Literature (Penguin Edition)* textbooks were structured the same for grade 7 and grade 9 with each book containing six units. Within each unit there was an introduction and then two unit parts that were identical in regards to organization. A separate skill area was covered in each unit part for a total of 12 overarching skill areas per textbook. These skill areas are provided in Appendix A. To assess breadth of coverage, Table 10 (grade 7) and Table 11 (grade 9) provide an overview of the number of model selections and paired reading selections each teacher covered. Please note that coverage of Reading Information Materials (RIM) selections or Comparing Literary Works (CLW) reading selections were not included in the reading selection tally for the tables below. In other words, a teacher who covered RIM or CLW reading selection(s), but did not use any of the paired reading selections in a particular unit part, will have a check mark indicating that they were in the unit part but will not have a reading selection listed. Cells with a “0” indicate that teachers did not report working within that particular unit part.

**Table 10. Breadth of Implementation Coverage for Grade 7 Treatment Teachers**

	State	CA			CO	IL	
	Teacher	T1	T2	T3	T4	T5	T6
Unit 1: Fiction and Nonfiction	Introduction	2MS	2MS	2MS	2MS	2MS	2MS
	Part 1: Context Clues	4RS	3RS	2RS	3RS	3RS	4RS
	Part 2: Author's Purpose	4RS		1RS	3RS	3RS	3RS
Unit 2: Short Story	Introduction	0	1MS	1MS	1MS	1MS	1MS
	Part 1: Predicting	1RS	1RS	1RS	3RS	2RS	1RS
	Part 2: Making Inferences		0	0	2RS	3RS	0
Unit 3: Types of Nonfiction	Introduction	0	1MS	0	1MS	1MS	1MS
	Part 1: Main Idea	0	1RS	1RS	3RS	3RS	2RS
	Part 2: Fact and Opinion	0		1RS	2RS	2RS	1RS
Unit 4: Poetry	Introduction	4MS	4MS	4MS	4MS	4MS	4MS
	Part 1: Drawing Conclusions	7RS	6RS	1RS	6RS	4RS	1RS
	Part 2: Paraphrasing	4RS		0	3RS	0	0
Unit 5: Drama	Introduction	2MS	2MS	0	2MS	2MS	2MS
	Part 1: Purpose for Reading	2RS		2RS	2RS	2RS	0
	Part 2: Summarizing	1RS	1RS	1RS	1RS	1RS	0
Unit 6: Oral Tradition	Introduction	0	3MS	0	3MS	3MS	0
	Part 1: Cause and Effect	1RS	0	0	2RS	1RS	0
	Part 2: Comparison and Contrast	0	0	0	2RS	2RS	0

Note: RS= Reading Selection; MS = Model Selection

Six grade 7 treatment teachers participated in the study. All six teachers covered unit 1 in its entirety (i.e., the introduction, part 1, and part 2). Most teachers covered units two through five. Within those units both the introduction and part 1 were more likely to be covered than part 2. Only two of the six grade 7 teachers covered unit 6 in its entirety, which makes unit 6 the least covered unit. The drop-off in coverage may be an indicator that the pace suggested by Prentice Hall (covering one unit every six weeks) was ambitious for teachers using the *Literature (Penguin Edition)* curriculum for the first time. On average,

grade 7 treatment teachers covered nine of the twelve possible unit parts and nearly five of the six possible unit introductions, which was sufficient coverage for the purposes of this study.

**Table 11. Breadth of Implementation Coverage for Grade 9 Treatment Teachers**

	State	CA					CO		IL	
	Teacher	T7	T8	T9	T10	T11	T12	T13	T14	T15
Unit 1: Fiction and Nonfiction	Introduction	2MS	2MS	2MS	2MS	2MS	2MS	2MS	2MS	2MS
	Part 1: Make Predictions	2RS	2RS	2RS	3RS	2RS	2RS	3RS	4RS	3RS
	Part 2: Author's Purpose		0	1RS	2RS			2RS	2RS	3RS
Unit 2: Short Stories	Introduction	1MS	1MS	1MS	1MS	0	1MS	0	1MS	1MS
	Part 1: Make Inferences	1RS	1RS	2RS	3RS	0	0	0	3RS	2RS
	Part 2: Cause and Effect	0	2RS	0	1RS	0	0	0	2RS	1RS
Unit 3: Types of Nonfiction	Introduction	0	0	0	1MS	0	1MS	1MS	1MS	1MS
	Part 1: Main Idea and Supporting Details	0	0	0	1RS	0	2RS	2RS	2RS	2RS
	Part 2: Analyze Persuasive Texts	0	0	0	0	0	3RS	2RS	2RS	0
Unit 4: Poetry	Introduction	2MS	0	2MS	2MS	2MS	2MS	2MS	2MS	2MS
	Part 1: Read Fluently	0	0	0	7RS		4RS	4RS	7RS	8RS
	Part 2: Paraphrase	0	0		0	0	5RS	3RS	3RS	4RS
Unit 5: Drama	Introduction	1MS	1MS	1MS	1MS	1MS	0		1MS	1MS
	Part 1: Summarize	0	5RS	0	5RS	1RS	5RS	5RS	5RS	5RS
	Part 2: Draw Conclusions	0	1RS	0	0	0	0	0		1RS
Unit 6: Heroism	Introduction	1MS	1MS	1MS	1MS	1MS	0	0	1MS	1MS
	Part 1: Cultural and Historical Context	1RS	1RS		1RS	1RS	2RS	0	2RS	2RS
	Part 2: Compare and Contrast	0	1RS	0	0	0	0	0	1RS	0

Note: RS= Reading Selection; MS = Model Selection

Nine grade 9 treatment teachers participated in the study. Grade 9 teachers frequently covered the unit introductions. With the exception of unit 3, no fewer than seven of the nine teachers covered the remaining five unit introductions. On average, grade 9 treatment teachers covered nearly five of the six possible unit introductions. Also of note, unit 3 was not covered by a majority of the CA teachers. California teachers were clear with researchers from the beginning that they would skip unit 3 in order to make time for required novels. Teachers chose to skip unit 3 because they perceived the unit to be redundant with unit 1 (both cover nonfiction) and not as relevant to their state standards. All teachers covered units 1 and 2 first then either went directly onto units 4-6 in that order or went in the opposite order by skipping to unit 6, then unit 5, and finally unit 4. The reason for the reordering of the units was because *The Odyssey* was required reading for all CA grade 9 students and so some teachers were anxious to teach unit 6 (which contained *The Odyssey*). Plus, mythology was heavily tested on the California Standards Test, so teachers wanted to make sure they covered that unit prior to their state testing.

Overall, there was a difference of three unit parts between the number of unit parts, on average, the CA treatment teachers covered (nearly six) versus the number of unit parts, on average, the non-CA treatment teachers covered (nine). The slower pace of the CA treatment teachers was due, in part, to time spent covering required novels (e.g., *Animal Farm* and *To Kill a Mockingbird*) and their use of materials and lessons from previous years teaching. For example, two treatment teachers supplemented the treatment curriculum materials for *The Odyssey* with mythology packets they had used previously. They did this because they felt the packets provided necessary background information and support for the text that was not provided by the treatment curriculum.

The two grade 9 Colorado teachers skipped most or all of unit 2 and unit 6. Unit 2 was skipped to read the required novel *Animal Farm*, although teachers covered the grammar and vocabulary from unit 2. Colorado teachers skipped unit 2 because they thought it was the most ancillary and least relevant to CO state standards. The teacher who worked within unit 6 covered the *The Odyssey* only. Teachers from all states were more likely to cover part 1 of unit 5 rather than part 2 and, in particular, cover *Romeo and Juliet* in part 1 of unit 5. *Romeo and Juliet* was required reading for most grade 9 students and because of this teachers spent most of their time while in unit 5 covering *Romeo and Juliet*.

The intended pacing, as outlined on Pearson’s website<sup>4</sup> and communicated directly to the researchers, is one unit every six weeks. Within the current study, treatment teachers spent approximately 32 weeks in school (on average), which did not leave enough time to complete six units using the intended pacing (i.e., 36 weeks). In addition, not all weeks were spent implementing the curriculum due to standardized testing, novel-based lessons, or other such interruptions to the curriculum (e.g., field trips, assemblies, student discipline issues, teacher absence, etc.).

### **Adherence to Treatment Implementation Guidelines**

The purpose of the implementation guidelines (Appendix G) was to ensure that treatment teachers would fully implement the *Literature (Penguin Edition)* curriculum as was intended by its developers.

Treatment teachers were provided with implementation guidelines they agreed to follow throughout the year. Table 12 indicates, of the reading selections covered for a particular unit, the percentage of those reading selections in which at least one required component was covered. For example, the required section entitled *Build Skills* consisted of three components: *Reading Skill*, *Literary Analysis*, and *Vocabulary Builder*. The coverage of at least one of these components was considered coverage for the entire required section. To interpret the table properly, please note that the unit of analysis is the reading selection (not the percentage of teachers). In other words, the percentages presented in Table 12 represent the percent of at least one required item from a particular curriculum component was covered out of all the reading selections covered by grade 7 teachers. For example, in unit 1 introduction, the grade 7 treatment teachers covered a total of twelve reading selections. In all 12 reading selections (or 100% of the time), at least one component of *From the Author’s Desk* (e.g., Author Introduces, Author Insight, etc.) was covered. However, within the same unit introduction, at least one component of the *Apply the Skills* (e.g., Research the Author, Quick Review, etc.) was covered 83% of the time by teachers (or in 10 out of the 12 reading selections). Table 13 is organized the same; the only difference is grade level.

---

<sup>4</sup> [http://www.phschool.com/sales\\_support/Product\\_Sites/Literature/faq.htm#3](http://www.phschool.com/sales_support/Product_Sites/Literature/faq.htm#3)

**Table 12. Grade 7 Treatment Teacher Fidelity to Implementation Guidelines**

<b>Introduction: Required Sections</b>	<b>Unit 1 (n=12)</b>	<b>Unit 2 (n=5)</b>	<b>Unit 3 (n=4)</b>	<b>Unit 4 (n=6)</b>	<b>Unit 5 (n=10)</b>	<b>Unit 6 (n=9)</b>	<b>Average Across Units</b>
<i>From the Author's Desk</i> (e.g., Author Introduces, Author Insight, etc.)	100%	100%	100%	100%	100%	100%	100%
Intro: <i>Apply the Skills</i> (e.g., Research the Author, Quick Review, etc.)	83%	40%	50%	83%	80%	67%	67%
Intro: <i>Ancillary Materials</i> (e.g., Diagnostic Test)	83%	40%	25%	17%	40%	67%	45%
<b>Unit Part Reading Selections: Required Sections</b>	<b>Unit 1 (n=33)</b>	<b>Unit 2 (n=15)</b>	<b>Unit 3 (n=16)</b>	<b>Unit 4 (n=33)</b>	<b>Unit 5 (n=13)</b>	<b>Unit 6 (n=8)</b>	<b>Average Across Units</b>
<i>Build Skills</i> (e.g., Literary Analysis, Vocabulary Builder, etc.)	94%	80%	94%	97%	92%	100%	93%
<i>Build Understanding</i> (e.g., reading/writing connection)	76%	87%	81%	49%	62%	88%	74%
Point of Use Notes	64%	47%	6%	6%	23%	25%	29%
<i>Build Language Skills</i> (e.g., Vocabulary Skill, Grammar Lesson, etc.)	73%	47%	25%	9%	46%	25%	38%
<i>Apply the Skills</i> (e.g., Literary Analysis, Vocabulary Builder, etc.)	94%	93%	94%	91%	85%	100%	93%
<i>Ancillary Materials</i> (e.g., Accessibility-at-a- Glance charts from TE, etc.)	79%	27%	31%	27%	46%	100%	52%
<b>Additional: Required Sections</b>	<b>Unit 1 (n=12)</b>	<b>Unit 2 (n=9)</b>	<b>Unit 3 (n=10)</b>	<b>Unit 4 (n=9)</b>	<b>Unit 5 (n=10)</b>	<b>Unit 6 (n=5)</b>	<b>Average Across Units</b>
<i>Reading Informational Materials</i> (e.g., Prereading page, etc.)	83%	44%	40%	33%	40%	20%	43%

**Table 13. Grade 9 Treatment Teacher Fidelity to Implementation Guidelines**

<b>Introduction: Required Sections</b>	<b>Unit 1 (n=18)</b>	<b>Unit 2 (n=7)</b>	<b>Unit 3 (n=5)</b>	<b>Unit 4 (n=16)</b>	<b>Unit 5 (n=8)</b>	<b>Unit 6 (n=7)</b>	<b>Average Across Units</b>
<i>From the Author's Desk</i> (e.g., Author Introduces, Author Insight, etc.)	100%	86%	100%	100%	100%	100%	98%
Intro: <i>Apply the Skills</i> (e.g., Research the Author, Quick Review, etc.)	22%	29%	40%	38%	50%	71%	42%
Intro: <i>Ancillary Materials</i> (e.g., Diagnostic Test)	33%	14%	20%	25%	0%	14%	18%
<b>Unit Part Reading Selections: Required Sections</b>	<b>Unit 1 (n=31)</b>	<b>Unit 2 (n=15)</b>	<b>Unit 3 (n=19)</b>	<b>Unit 4 (n=36)</b>	<b>Unit 5 (n=29)</b>	<b>Unit 6 (n=11)</b>	<b>Average Across Units</b>
<i>Build Skills</i> (e.g., Literary Analysis, Vocabulary Builder, etc.)	84%	80%	95%	83%	86%	91%	87%
<i>Build Understanding</i> (e.g., reading/writing connection)	42%	27%	26%	47%	21%	27%	32%
Point of Use Notes	45%	20%	26%	25%	28%	82%	38%
<i>Build Language Skills</i> (e.g., Vocabulary Skill, Grammar Lesson, etc.)	26%	0%	42%	22%	24%	36%	25%
<i>Apply the Skills</i> (e.g., Literary Analysis, Vocabulary Builder, etc.)	71%	87%	42%	69%	76%	64%	68%
<i>Ancillary Materials</i> (e.g., Accessibility-at-a- Glance charts from TE, etc.)	71%	40%	53%	75%	38%	18%	49%
<b>Additional: Required Sections</b>	<b>Unit 1 (n=18)</b>	<b>Unit 2 (n=8)</b>	<b>Unit 3 (n=7)</b>	<b>Unit 4 (n=12)</b>	<b>Unit 5 (n=9)</b>	<b>Unit 6 (n=9)</b>	<b>Average Across Units</b>
<i>Reading Informational Materials</i> (e.g., Prereading page etc.)	56%	38%	71%	50%	44%	11%	45%

Averaging across units for both grade 7 and 9, teachers' fidelity to implementation was fairly similar across grade level. However, there were four notable required sections that were covered at least 25% less by grade 9 teachers than grade 7 teachers. Those sections included the *Apply the Skills* section of the unit introductions (42% compared to 67%), the *Apply the Skills* sections within the unit parts (68% compared to 93%), *Ancillary Materials* in the unit introductions (18% compared to 45%), and the *Build Understanding* section of the unit parts (32% compared to 74%). The least frequently covered required section, on average, for grade 7 teachers was *Point of Use Notes* (29%) and the most frequently covered item, on average, was *From the Author's Desk* in the unit introduction (100%). For grade 9 teachers the least frequently covered item, on average, was the *Ancillary Materials* in the unit introductions (18%) and the most frequently covered item, on average, was *From the Author's Desk* in the unit introductions (98%).

One required section that was covered fairly infrequently in both grade 7 and grade 9 was the section entitled *Build Language Skills*. One explanation for this limited coverage was that this particular section connected back to both of the paired reading selections, so a teacher who only read one of the two paired reading selections with their class may have thought that their students could not complete this page without having reading both reading selections.

It is important to mention that the grade 9 teachers covered fewer required components in unit 4 (the poetry unit) compared to the other units. Although it is not entirely clear why required items were not sufficiently covered, some teachers indicated that they supplemented this unit with their own outside materials more so than they did with the other units. One teacher also mentioned that the writing requirements in unit 4 were rather weak, so he/she spent time having her students actually write poetry while working through unit 4. Of course, it was also possible that some teachers may not have been comfortable with poetry, and therefore, only minimally covered the unit.

A portion of the *Literature (Penguin Edition)* curriculum that was not required, but was strongly recommended was Comparing Literary Works (CLW). This section typically included two reading selections, but could include up to four reading selections (e.g., the poetry unit included four short poems rather than two short stories). Unlike the paired reading selections within the unit parts, the intent of the CLW reading selections was to be read together. The purpose for this was to compare and contrast the selections. There was one CLW section per unit part for a total of 12 possible CLWs. Grade 7 teachers covered anywhere from two to seven CLWs over the course of the year with an average of four

CLWs covered. When teachers did use a CLW with their class, they tended to always read all reading selections involved (i.e., sporadically they might read only one or two of the reading selections, but mainly they read both). Grade 9 teachers covered anywhere from four to eight CLWs over the course of the year with an average of five CLWs covered. When teachers did use CLW with their class, they tended not to read all of the reading selections involved.

### **Description of Control Curricula Implementation**

The control group used six distinct English Language Arts (ELA) textbooks. Comparison textbooks included an earlier version (2002) of a Prentice Hall textbook at the two middle schools sites in CA. The remaining five sites used editions of competitors' textbooks, with publication dates ranging from 1985 to 2002. The oldest textbook (publication date of 1985) was quite different from the other control textbooks. The 1985 textbook was almost entirely made-up of reading selections, which were accompanied by roughly two pages of story-related activities (e.g., discussion questions, a language and/or vocabulary section, a writing assignment, and a brief blurb about the author). The other control textbooks provided more support for the reading selections, as well as additional elements such as workshops for the students (e.g., writing, language, reading, etc.), student models, revision of writing and editing support, comparing literature sections, standardized test practice, and other such elements. Also, except for the oldest control textbook, the other control textbooks all were organized by themes (e.g., themes related to making choices, family, reality, strength, love, identity, etc.). This differed from the treatment textbooks, which were organized by genre and reading skill rather than theme.

Two control teachers covered more outside reading selections than reading selections from their textbooks. However, it was much more typical for control teachers to read a mixture of reading selections from a variety sources. Examples of outside reading included novels such as *To Kill a Mockingbird*, by Harper Lee for ninth grade and *The Pearl*, by John Steinbeck; *The Outsiders*, by S.E. Hinton; *Animal Farm*, by George Orwell; *The Magician's Nephew*, by C.S. Lewis; *The Breadwinner*, by Deborah Ellis for seventh grade. Control teachers included poetry that was not part of their textbook, for example: *Mending Wall*, *The Road Not Taken*, *Fire and Ice*, and *Stopping by Woods on a Snowy Evening* all by Robert Frost; *The Raven* and *Annabel Lee*, both by Edgar Allan Poe, and so on. Some teachers brought in non-fiction books (e.g., *7 Habits of Highly Effective Teens*, by Sean

Covey, etc.), biographies (e.g., Charles Dickens, Rudyard Kipling, Jack London, Ray Bradbury, etc.), as well as assorted articles and handouts.

### **Alignment of Treatment to Control Textbooks**

Due to the diversity of textbooks used by the control group, it was impractical to align each of the control texts to the treatment textbook. Further adding to this complexity, both treatment and control groups were not exclusively using textbooks, but also incorporated novel-based lessons into their curriculum. Teachers within the treatment group were required by their districts to teach certain novels (e.g., *Animal Farm*, *To Kill a Mockingbird*, etc.). In addition, three teachers from the control group used what could be considered a novel-based curriculum (i.e., they included more reading selections from outside of their textbooks than reading selections from their textbooks over the course of the school year). These factors limited the number and type of comparisons that could be made between treatment and control teachers.

### **Comparing Coverage Across Treatment and Control Groups**

Two categories were used to compare coverage across treatment and control groups: reading selection coverage and skill area coverage. To make an accurate comparison of reading selection coverage, it was necessary to create an equivalency system. No manipulation was necessary in the case of the skill area coverage comparison.

#### **Reading Selections Coverage Across Treatment and Control Groups**

To create an equivalent summary that accounted for both reading selections covered within the textbook as well as outside the textbook, an equivalency system was devised. This system was based on two factors: (1) the average number of weeks it took teachers to cover a particular outside reading selection (oftentimes a novel); and (2) the length of the reading selection. For example, reading the novel *The Pearl*, by John Steinbeck (97 pages), took teachers approximately two weeks to cover. Researchers decided this novel would be worth the equivalent of two textbook reading selections. An equivalency system was also needed to control for the same reading selection (e.g., *Romeo and Juliet*) being presented differently across each textbook (which caused the reading selections to have different value in the synthesis of the data). For example, in the table of contents of the grade 9 treatment textbook, Shakespeare's play *Romeo and Juliet* was split into five separate reading selections, which corresponded to each of the five Acts. However, in the grade 9 control

textbooks *Romeo and Juliet* was treated as a single reading selection in each table of contents. Consequently, any teacher who covered *Romeo and Juliet* from their textbook was granted a value of five reading selections to make both treatment and control implementation equivalent.

Using the system outlined above, the total number of reading selections each teacher covered over the course of the school year, both within their textbook and outside of their textbook, was tallied. This analysis revealed that while both treatment and control groups read more selections from their textbooks than from outside their textbooks, treatment teachers were more likely than control teachers to read the majority of their reading selections from the textbook. In fact, after combining reading selections from within and outside their textbook, grade 7 treatment teachers covered significantly more reading selections ( $M = 49.83$ ) than grade 7 control teachers ( $M = 22.86$ ),  $t(11) = 3.72$ ,  $p < .01$  and grade 9 treatment teachers covered significantly more reading selections ( $M = 49.44$ ) than grade 9 control teachers ( $M = 28.00$ ),  $t(16) = 3.83$ ,  $p < .01$ . One possible explanation for these differences lie in the fact that treatment teachers were provided with implementation guidelines that outlined required and strongly recommended features of the curriculum, thus encouraging treatment teachers to move through the curriculum at a certain pace. Meanwhile, control teachers were instructed to do as they had done in the past, which may or may not have resembled the level of coverage required for the treatment teachers. Regardless, it was clear that students in the treatment teacher classrooms were reading more literature, on average, than students within the control classrooms.

To examine these differences by teacher, an implementation scale was created based on tally described above. Since the minimum number of total reading selections covered was 13 and the maximum number was 78, the scale was based on the following range: 1 = low implementation (0-29 reading selections covered), 2 = medium implementation (30-50 reading selections covered), and 3 = high implementation (51 or more reading selections covered). The results suggest that the treatment group fell into the categories of both medium (2) and high (3) implementation while the control group fell into the categories of low (1) and medium (2) implementation. This difference was the same for grade 7 and grade 9, so grade level differences were not reported. There was a noticeable difference between CA and non-CA treatment teachers in that only one CA treatment teacher fell into the high implementation category while five non-CA treatment teachers did.

### Skills Coverage Across Treatment and Control Groups

Control teachers were asked on the teacher implementation log to report the skills they covered in any given week. For instance, did they focus on “making predictions” or “drawing conclusions”. The skills listed on the control teacher log corresponded with the skills covered within each unit part in both the grade 7 and grade 9 treatment textbooks. The results of this comparison should be interpreted with caution to the difference in the data collection format. Control teachers were asked to report specifically whether they covered a particular skill area from a list of 15 skill areas, whereas treatment teachers were only considered to have covered a skill area if they reported covering a unit part that focused on that skill level. Therefore, the former is less precise than the latter.

The results of this analysis suggested that almost all control teachers reported covering each skill area. The exceptions include: compare and contrast (89% of grade 9 control teachers reported covering this), context clues (86% of grade 7 control teachers), fact and opinion (86% of grade 7 control teachers), and analyzing persuasive texts (71% of grade 7 control teachers and 67% of grade 9 control teachers). In comparison, the only skill areas all treatment teachers covered were prediction (grade 7 and 9), author’s purpose (grade 7 only), drawing conclusions (grade 7 only), and context clues (grade 7 only). The skill area covered by the fewest number of treatment teachers was compare and contrast at 33% coverage for grade 7 and 22% coverage for grade 9.

### **Comparing Teacher Quality Across Treatment and Control Groups**

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 teacher quality by observing over two occasions throughout the year, we believed that some measures of overall quality could be observed in this limited timeframe.

#### Teacher Quality

Each teacher within the treatment and control groups received a rating, from the researchers, on two dimensions (classroom management skill and teacher quality) using a scale of “1” (lowest) to “3” (highest). Examples of teacher behavior that typify a “1”, “2”, or “3” for classroom management and teacher quality are provided in Table 14.

**Table 14. Examples of Teacher Practices that Represent Scores on the Classroom Management and Teacher Quality Scales**

Scale	Examples of Classroom Management Behaviors	Examples of Teacher Quality Behaviors
1	<ul style="list-style-type: none"> <li>• Ignores/ reacts slowly to students who are disrupting the class.</li> <li>• Only addresses students who actively participate, while silent students are not encouraged to participate.</li> <li>• Allows disruptive students to dominate the class discussions while less disruptive students are prevented from participating.</li> <li>• Allows topic of conversation into irrelevant personal matters unrelated to course content with no apparent academic benefit.</li> </ul>	<ul style="list-style-type: none"> <li>• Does not provide structure/objectives for students to understand in the lesson.</li> <li>• Does not adequately enforce the completion of assigned reading.</li> <li>• “Steamrolls” through lessons without giving much energy or time to differentiated instruction, clarifying content, or answering questions.</li> <li>• Does not provide sensemaking or wrap up of lesson contents.</li> <li>• Misses obvious opportunities for learning with students.</li> <li>• Low expectations for student intellect or behavior.</li> </ul>
2	<ul style="list-style-type: none"> <li>• Generally keeps students on task while moving through the lesson; allows for minimal disruption.</li> <li>• Teacher sits behind a desk or at front of room more than they walk around the classroom.</li> <li>• May have occasional lapse of focus, but can redirect highly disruptive students.</li> </ul>	<ul style="list-style-type: none"> <li>• Learning objectives related to that day’s lesson are stated.</li> <li>• Encourages curiosity and creativity in students.</li> <li>• May allow time for sensemaking of lesson but may occasionally miss obvious opportunities for learning.</li> <li>• Moderate expectations for student intellect or behavior.</li> </ul>
3	<ul style="list-style-type: none"> <li>• A daily routine is clearly followed and this daily routine promotes student learning.</li> <li>• The teacher’s demeanor and day-to-day interactions with students have gained that teacher respect and a position of authority within the classroom.</li> <li>• Does not allow disruptive students to dominate conversation or misdirection of the lesson.</li> <li>• Engages attentive students in lesson and rewards student for positive behaviors.</li> </ul>	<ul style="list-style-type: none"> <li>• Learning objectives are not only stated, but also sufficiently accomplished.</li> <li>• Provides adequate time/structure for “sense-making” and wrap up.</li> <li>• Has high expectations for student intellect or behavior.</li> <li>• Teacher easily moves through lesson engaging all students in learning process.</li> <li>• Uses differentiation instruction techniques to engage all levels of learners.</li> </ul>

The two ratings were summed for each teacher to create an overall teacher quality variable. For example, a teacher who was rated a “1” for low level of classroom management, but was rated a “2” for medium level of teacher quality would receive a sum of “3” for overall teacher quality rating. These ratings were based on both the fall and spring classroom observations. Out of a possible score ranging from two to six, six teachers (19%) received an overall teacher quality rating of “6”, six teachers (19%) received a rating of “5”,

13 teachers (42%) received a rating of “4”, one teacher (3%) received a rating of “3”, and five teachers (16%) received the lowest rating, a “2”. The teacher quality ratings were fairly equivalent across treatment and control groups, except in CO where treatment teachers received an overall teacher quality rating of “6”, while the control teachers received overall ratings of “2” and “3” respectively.

### Teacher Influence

In addition to classroom observations, the student posttest survey included questions pertaining to students’ perceptions of their teachers. These data provide another vantage point from which to look at teacher quality. Mean ratings of “Teacher Influence” were calculated for each teacher. The mean “Teacher Influence” rating was created by averaging the mean response of the following four student posttest survey items: “My teacher expects me to do well in my ELA class,” “My teacher explains concepts clearly and correctly,” “My teacher helps me want to do my best in ELA class,” and “During my ELA class, it is always clear what I am supposed to be doing at all times.” In general, students rated their teachers in a relative positive light (no lower than a mean of 3.78) in relation to the likert scale of 1-5 (1 = strongly disagree with the statement; 5 = strongly agree with the statement) for the composite variable of Teacher Influence. The lowest mean in the composite, on average, was associated with the final question: “During my ELA class, it is always clear what I am supposed to be doing at all time.” On occasion, “My teacher helps me want to do my best in ELA class” would receive the lowest mean rating within the composite. Students’ perceptions of their teachers were comparable across treatment and control groups, although as will be discussed in Section Five, grade 7 students rated their teacher more favorably than grade 9 students in both the treatment and control groups.

### Differentiated Instruction

Both treatment and control teachers were given the opportunity to discuss whether or not they differentiated their instruction during the teacher interviews at the conclusion of the study and on the weekly teacher logs. The weekly teacher log data reflect similar information received from the teacher interviews.

The teacher interview data indicated that treatment teachers predominantly spoke about materials they used to adjust assignments and class work to students’ skills and abilities rather than specific teacher practices. The primary mode of differentiated instruction was through tailoring reading assignments for the students (i.e., using textbooks for some students and Reader’s Notebooks for others); adapting vocabulary lists and using the CDs provided were two other commonly reported strategies. Teachers did not use the paired

selections as a method of differentiating instruction, but rather chose the selection based on what teachers' thought their students would be most engaged. A number of treatment teachers reported that the curriculum was easy to adapt to meet students' needs. However, some treatment teachers said that they did not adjust the curriculum for any students, either because they did not see a need to do so, or because they did other activities or used other strategies. For example, when asked whether she used the curriculum in differentiated instruction, one teacher said, "Well, no... not really. I mean like... here's what the curriculum was for everybody and then from there, I just made adjustments because kids struggle with different parts. But I didn't use the Reader's Notebook on a differentiated level... overall I didn't see a big gaping hole..."

Control teachers reported giving students a range of reading options to account for varying student levels. Other techniques of differentiating instruction included adapting vocabulary assignments, group work that involved a combination of lower and higher-level students working together, and adjusting class assignments for lower and higher-level students. Some teachers created adapted assignments themselves, and others used variations available in their textbooks. The most-reported strategy used by control teachers involved giving personal attention to lower-level students. One teacher discussed having one-on-one conferences with students to discuss their writing, and many others gave a general description of helping students that need extra attention. Finally, there were five teachers who reported not using their textbooks/curricula for differentiated instruction. Some control teachers found using their textbook difficult or "cumbersome" and others reported that it was easier to create their own adaptations.

In general, teachers in both conditions differentiated instruction for their students, however, teachers in the treatment group were more likely to use the textbook and ancillary materials for differentiation than control group teachers. Control teachers, on the other hand, were more likely to adapt materials they created or change the dynamic of the class (specifically working with lower-level students, for example) when compared to treatment teachers.

## **SECTION FIVE: RESULTS RELATED TO STUDENTS' ATTITUDES AND ACHIEVEMENT IN ENGLISH LANGUAGE ARTS**

In the first part of Section Five , we report on three of the four research questions involving student outcomes: (1) How do student outcomes differ for students using the Prentice Hall *Literature (Penguin Edition)* 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 teacher characteristics (e.g., experience)?; and (3) What is the relationship between program implementation and student achievement in reading and language arts?

To answer the first three questions, it is important to note that we conducted all quantitative analyses separately for each of the two grade levels (7th and 9th) for two main reasons. First, our outcome measures were not entirely consistent across each grade level (e.g., Reading and Language in grade 7; Reading and Spelling in grade 9). Second, exploratory analyses suggested differences between these two grade levels in terms of missing patterns and the relationships between predictors and outcomes. Therefore, we will qualify the research question as it pertains to grade levels separately.

### **Differential Attrition**

Almost any experimental study has participant attrition, particularly in applied research settings (i.e., schools) where students may leave before the year is over due to circumstances outside of the control of the school and teacher. What is important to determine, however, is whether or not there was differential attrition such that students in one group (treatment or control) were more likely to exit the study than another group before completing the posttest measures. Section Two of this report reviewed the patterns of data comparing missing posttest data for treatment versus control groups and the results indicated that there were no major differences in student demographics that would be of concern to the researchers. However, it was also necessary to examine how attrition may have impacted students' achievement in ELA. It is important to note that while the overall study sample is 1,722 students, the number of students included in measuring each outcome variable will vary given that a student could potentially not complete one outcome measure (e.g., Spelling subtest) but yet complete all other outcome measures at pretest and posttest.

To examine how sample attrition might have impacted the treatment and control groups, we conducted a series of exploratory analyses involving students' pretest scores. Seventh graders with missing posttest reading scores (in both treatment and control groups) tended to have lower pretest reading scores than students who had posttest reading scores. This pattern held for the other three pretest measures (i.e., grade 7 language, grade 9 total reading and spelling). In addition, there was a significant interaction between missing posttest scores and group (i.e., treatment vs. control) for grade 7 ITBS language ( $p=.04$ ). Specifically, grade 7 students who did not have posttest scores in the control group, on average, had lower pretest scores than grade 7 students missing posttest scores in the treatment group. In contrast, grade 9 students missing posttest scores on the ITED reading total and spelling in the treatment group, on average, had lower pretest scores than grade 9 students who did not have posttest scores in the control group. In other words, students missing posttest scores tended to have lower prior achievement scores than those who did not; however, the missing pattern in the treatment vs. control groups differed somewhat for 7th and 9th graders.

To summarize, these exploratory analyses suggested differences between the treatment and control groups in terms of demographic background characteristics (proportion of girls and Caucasian) and prior reading achievement, which seemed to favor<sup>5</sup> the treatment groups at both grade levels.

### **Performance on Outcome Measures**

Tables 3 and 4 (found in Section Two of this report) indicated that the demographic characteristics of study participants were fairly comparable across treatment and control groups. While there were some differences between the grade levels (e.g., higher proportion of Caucasian students in grade 9 treatment group), we took this potential difference into account by including the ethnic minority indicator variables (African American and Latino) into the statistical models.

#### **ITBS and ITED Percentile Rank Scores**

Besides examining these key demographic variables, we also compared students on proxy indicators of prior achievement. These proxy indicator variables included: (1) pretest scores (ITBS language percentile rank and ITBS reading percentile rank for 7th graders;

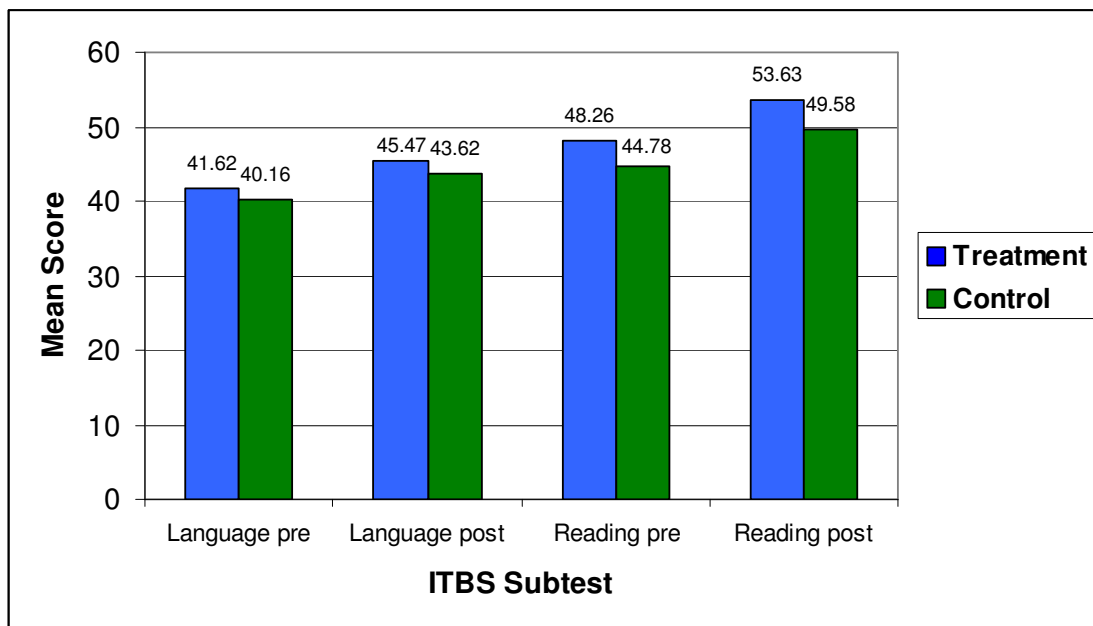
---

<sup>5</sup> Research has shown that girls tend to perform better than boys on language related subjects (Baker & Wigfield, 1999).

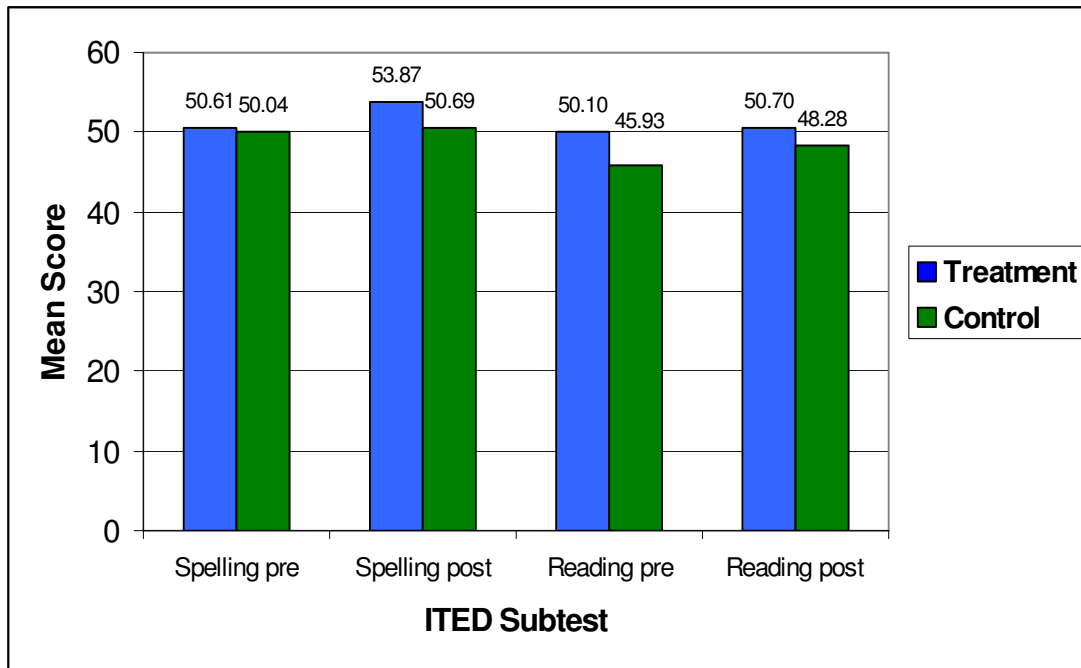
ITED spelling percentile rank and ITED reading total percentile rank for 9th graders, see Figures 1 and 2), and (2) writing assessment pretest measures (content, organization, voice, and convention, see Figures 3 and 4).

Figure 1 illustrates that grade 7 students in both treatment and control groups increased their achievement scores from pretest to posttest by approximately three to five percentile ranks. Grade 9 students (see Figure 2) demonstrated smaller growth in achievement, ranging from less than one percentile rank to approximately three percentile ranks across both treatment and control groups. In addition, the figures suggest that while the treatment and control groups were fairly similar with respect to language (for 7th graders) and spelling (for 9th graders), they differed on the reading assessments (for both 7th and 9th graders). Specifically, students in the treatment group at both 7th and 9th grade (48.26 and 50.10, respectively) scored higher than students in the control groups at corresponding grade levels the pretest (44.78 and 45.93, respectively). Therefore, the higher scores the treatment students demonstrated at posttest (in comparison to control students) can be explained by students' higher pretest scores.

**Figure 1. English Language Arts Achievement Pretest and Posttest Scores (Grade 7)**



**Figure 2. English Language Arts Achievement Pretest and Posttest Scores (Grade 9)**

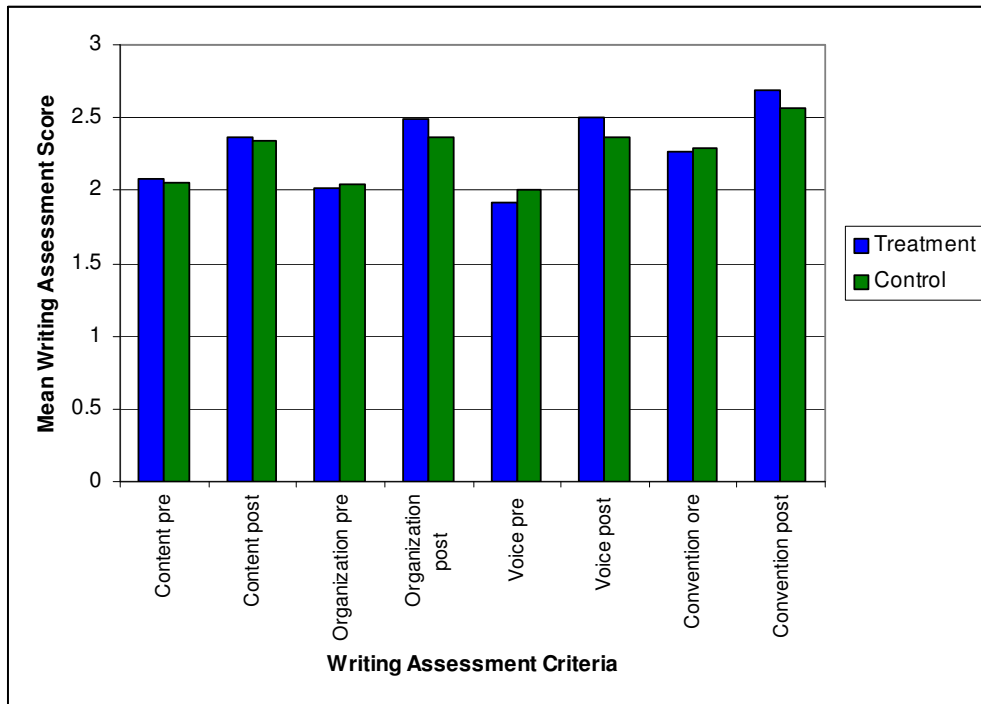


### Writing Assessment Scores

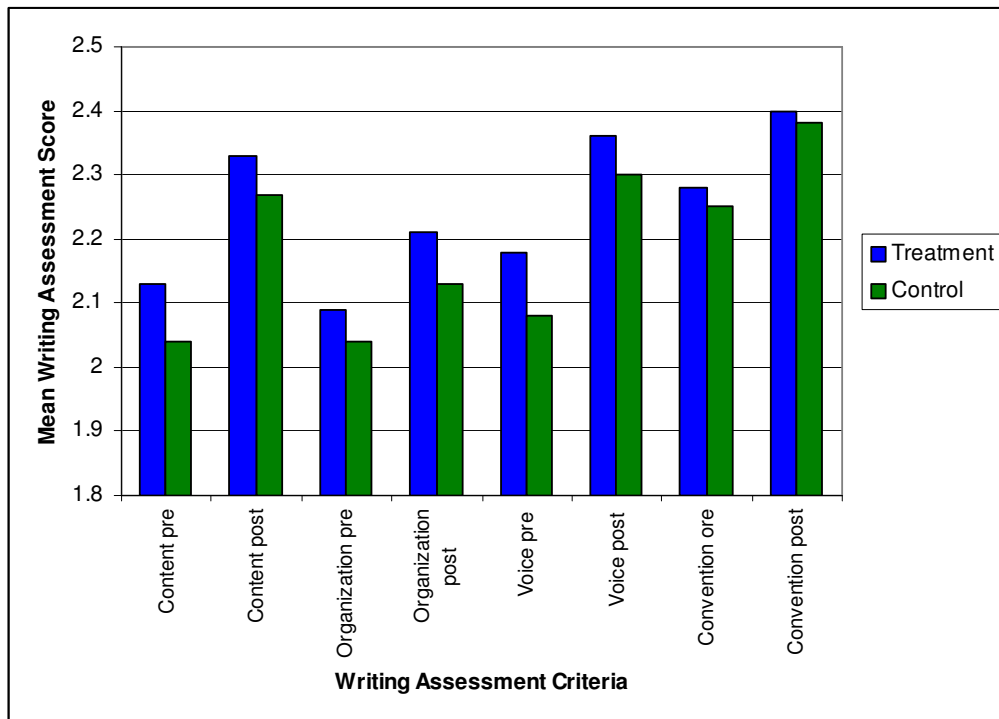
Another outcome measure that can be compared between treatment and control groups was the writing assessment. Participating students at both grade levels completed persuasive essays at both pre and posttest. To achieve sufficient reliability of scores, two raters coded each essay on four dimensions: content, organization, voice, and convention. Analysis of writing scores requires that inter-rater reliability is achieved as measured by Cohen's kappa (Stemler, 2004). Calculation of Cohen's kappa indicates a high level of inter-rater reliability, as most rater pairs were in agreement above .90. Figures 3 and 4 illustrate the writing assessment results for both grade levels and treatment groups. In general, treatment and control groups were similar in terms of their writing assessment scores. No differences were observed on any of the writing assessment measures between the treatment and control groups for 7th and 9th graders<sup>6</sup>. While both groups increased from pretest to posttest in all four dimensions of the writing assessment, treatment group students did not demonstrate greater increases on the writing assessment as compared with the control students at either grade level.

<sup>6</sup> To streamline our analysis focus, we therefore focused on the ITBS/ITED test scores as the key outcome measures in this study.

**Figure 3. Writing Assessment Pretest and Posttest Scores (Grade 7)**



**Figure 4. Writing Assessment Pretest and Posttest Scores (Grade 9)**



### Student Survey

In addition to demographic variables, we also incorporated student responses gathered from the posttest student survey. Based on an Exploratory Factor Analysis (EFA) on the posttest student survey, we derived four composite factors: intrinsic motivation, self-efficacy, teacher influence, and parent involvement (see Appendix N for a complete description of these factors). Except for parent involvement, the composite factors were significantly positively correlated to students' achievement in English Language Arts in grade 7 and grade 9 across each of the subtests of the ITBS and ITED. Therefore, these three factors were explored during the development of the final HLM model.

### **Differences in English Language Arts Achievement Between Treatment and Control Groups**

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 group (treatment and control) as well as student and teachers characteristics. HLM models were particularly appropriate for analyzing data of this kind (i.e., students within different classrooms) because they simultaneously examined the effect of student background variables (e.g., ethnicity) and teacher/instructional characteristics (e.g., teacher academic background) on students' language arts achievement. For a complete discussion of the rationale and theory underlying HLM models, please see Byrk and Raudenbush (1992).

Appendix C includes a list of variables and their operational definitions associated with student background characteristics and teacher/ classroom/ school characteristics that were explored prior to building the final HLM models (see Appendices V, W, X and Y). Descriptive statistics for the variables in the final HLM models can be found in their corresponding appendices. For the top section of Table 15, all level 1 and level 2 variables were dichotomously coded except for two level 1 continuous variables, which were student self rating of intrinsic motivation (MOTIVATION) and Reading pretest score (READPRE); and three level 2 continuous variables including years teaching English language arts (YRSELA), researcher ratings of teacher quality (TQUALITY) and level of fidelity to implementation (IMPLEMENT). There were similar values for the bottom half of Table 15, with the addition of three other variables. Specifically, socio-economic status, SES, (which was defined as the number of students eligible for free or reduced lunch), student rating of

self efficacy (EFFICACY), and Language pretest score (LANGPRE) were all continuous variables.

In an effort to build a parsimonious HLM model, univariate analyses were conducted comparing all level 1 and level 2 factors to the outcome variables (Language Arts Reading scores, and Language Arts Language scores). Some of these factors were not significantly related to the outcomes so they were excluded from the final HLM model. Additional variables, including the remaining student survey composite variables 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. However, more specific analyses of these additional survey variables can be found at the end of Section Five.

As Table 15 illustrates, controlling for pretest (which was a significant predictor of posttest) and other variables in the model, there was no statistically significant difference on the ITBS reading test between 7th graders in the treatment and those in the control group (coefficient, -0.98;  $p=.545$ ). Besides pretest scores, students' self-reported intrinsic motivation was a significant predictor of posttest scores (coefficient, 2.05;  $p=.034$ ).

Everything else being equal, students in classes where teachers' implementation (IMPLEMENT) was rated high did significantly better (a little less than five and half percentile ranks) than students in classes where teachers' implementation was rated medium or low (coefficient, 5.44;  $p=.047$ ). In a similar manner, students with teachers who had more years of experience teaching English language arts (ELA) had a slight advantage over students with teachers who had fewer years of ELA teaching experience (coefficient, 0.34;  $p=.018$ ). Finally, Latino students on average scored about 6 percentile ranks lower than other non-Latino students (coefficient, -6.31;  $p=.000$ ).

The next analysis involved using HLM to assess any differences in grade 7 Language test scores between treatment and control group students. Descriptive statistics related to the analysis in bottom part of Table 15 can also be found in Appendix E. As shown in Table 15, controlling for pretest (which was a significant predictor of posttest) and other variables in the model, there was no statistically significant difference on the ITBS language test between 7th graders in the treatment and those in the control group (coefficient, 1.26;  $p=.463$ ).

Table 15. HLM Results for Grade 7 Reading and Language

HLM Results for Grade 7 Reading						
Fixed Effect		Coefficient	Standard Error	Approx. T-Ratio	df	p-value
For INTRCPT1, $\beta_{0j}$	<i>INTRCPT2, <math>\gamma_{00}</math></i>	51.08	1.07	47.65	8	.000
	<i>CONDIT, <math>\gamma_{01}</math></i>	-0.98	1.56	-0.63	8	.545
	<i>YRSELA, <math>\gamma_{02}</math></i>	.34	.11	2.99	8	.018
	<i>TQUALITY, <math>\gamma_{03}</math></i>	1.06	.60	1.77	8	.115
	<i>IMPLEMENT, <math>\gamma_{04}</math></i>	5.44	2.32	2.35	8	.047
For READPRE slope, $\beta_{1j}$	<i>INTRCPT2, <math>\gamma_{10}</math></i>	.68	.02	24.67	564	.000
For MOTIVATION slope, $\beta_{2j}$	<i>INTRCPT2, <math>\gamma_{20}</math></i>	2.05	.96	2.13	564	.034
For AFRICAN-AMERICAN slope, $\beta_{3j}$	<i>INTRCPT2, <math>\gamma_{30}</math></i>	-2.28	2.61	-0.87	564	.383
For LATINO slope, $\beta_{4j}$	<i>INTRCPT2, <math>\gamma_{40}</math></i>	-6.31	1.58	-4.00	564	.000
HLM Results for Grade 7 Language						
Fixed Effect		Coefficient	Standard Error	Approx. T-Ratio	df	p-value
For INTRCPT1, $\beta_{0j}$	<i>INTRCPT2, <math>\gamma_{00}</math></i>	44.25	1.21	36.29	9	.000
	<i>SES, <math>\gamma_{01}</math></i>	-0.70	.17	-4.06	9	.003
	<i>CONDIT, <math>\gamma_{02}</math></i>	1.26	1.65	.77	9	.463
	<i>TQUALITY, <math>\gamma_{03}</math></i>	2.30	.58	3.97	9	.004
For LANGPRE slope, $\beta_{1j}$	<i>INTRCPT2, <math>\gamma_{10}</math></i>	.75	.02	25.83	554	.000
For EFFICACY slope, $\beta_{2j}$	<i>INTRCPT2, <math>\gamma_{20}</math></i>	2.34	.86	2.71	554	.007
For AFRICAN-AMERICAN slope, $\beta_{3j}$	<i>INTRCPT2, <math>\gamma_{30}</math></i>	-4.23	2.71	-1.56	554	.119
For LATINO slope, $\beta_{4j}$	<i>INTRCPT2, <math>\gamma_{40}</math></i>	-0.47	1.58	-0.30	554	.765

Besides pretest scores, students' self-efficacy was a significant predictor of posttest scores (coefficient, 2.34;  $p=.007$ ). Everything else being equal, students in classes where teachers scored higher on the overall ratings based on both classroom management and teacher quality did significantly better (a little less than two and half percentile ranks with one unit increase in teacher rating) than students in classes where teachers' overall ratings were lower (coefficient, 2.3;  $p=.004$ ).

Finally, we found one school level variable (i.e., socio-economic status) to be a significant predictor of 7th grade spelling percentile rank scores. Results indicated that schools with a higher percentage of students on free and reduced lunch had students with lower achievement in spelling.

The next analysis involved using HLM to assess any differences in grade 9 Reading test scores and Spelling test scores between treatment and control group students. A more specific description of the variables included in Table 16 can also be found in Appendices X and Y.

Similar variables for grade 9 factors were observed in the preceding grade 7 analyses, with the exception of one additional continuous variable (student attendance: ATTENDANCE) and one additional categorical variable, other ethnicity (OTHER). As Table 26 illustrates, controlling for pretest (which was a significant predictor of posttest) and other variables in the model, there was no statistically significant difference on the ITED reading total test between 9th graders in the treatment and those in the control group (coefficient, -2.13;  $p=.345$ ).

Besides pretest scores, students' self reported intrinsic motivation (coefficient, 1.72;  $p=.034$ ), self efficacy (coefficient, 1.54;  $p=.041$ ), and percent attendance of ELA classes (coefficient, 0.44;  $p=.002$ ) were significant predictors of posttest grade 9 total reading percentile rank scores. Results also showed that African American students on average scored about 4 percentile ranks lower than other non-African American students (coefficient, -3.97;  $p=.025$ ).

The next analysis involved using HLM to assess any differences in grade 9 Spelling test scores between treatment and control group students. When controlling for pretest score (which was a significant predictor of posttest) and other variables in the model, there was no statistically significant difference on the ITED spelling test between 9th graders in the treatment and those in the control group (coefficient, 1.38;  $p=.533$ ). Besides pretest scores, the only significant predictor of 9th graders' posttest ITED spelling percentile rank scores was students' self-reported intrinsic motivation (coefficient, 3.0;  $p=.001$ ).

**Table 16. HLM Results for Grade 9 Reading and Spelling**

<b>HLM Results for Grade 9 Reading</b>						
<b>Fixed Effect</b>		<b>Coefficient</b>	<b>Standard Error</b>	<b>Approx. T-Ratio</b>	<b>df</b>	<b>p-value</b>
For INTRCPT1, $\beta_{0j}$	<i>INTRCPT2, <math>\gamma_{00}</math></i>	<b>50.51</b>	<b>1.56</b>	<b>32.23</b>	<b>16</b>	<b>.000</b>
	<i>CONDIT, <math>\gamma_{01}</math></i>	-2.13	2.18	-0.97	16	.345
For READPRE slope, $\beta_{1j}$	<i>INTRCPT2, <math>\gamma_{10}</math></i>	<b>.79</b>	<b>.02</b>	<b>30.67</b>	<b>757</b>	<b>.000</b>
For ATTENDANCE slope, $\beta_{2j}$	<i>INTRCPT2, <math>\gamma_{20}</math></i>	<b>.44</b>	<b>.14</b>	<b>3.18</b>	<b>757</b>	<b>.002</b>
For MOTIVATION slope, $\beta_{3j}$	<i>INTRCPT2, <math>\gamma_{30}</math></i>	<b>1.72</b>	<b>.79</b>	<b>2.17</b>	<b>757</b>	<b>.030</b>
For EFFICACY slope, $\beta_{4j}$	<i>INTRCPT2, <math>\gamma_{40}</math></i>	<b>1.54</b>	<b>.75</b>	<b>2.04</b>	<b>757</b>	<b>.041</b>
For AFRICAN-AMERICAN slope, $\beta_{5j}$	<i>INTRCPT2, <math>\gamma_{50}</math></i>	<b>-3.97</b>	<b>1.77</b>	<b>-2.24</b>	<b>757</b>	<b>.025</b>
For LATINO slope, $\beta_{6j}$	<i>INTRCPT2, <math>\gamma_{60}</math></i>	.73	1.36	.53	757	.595
<b>HLM Results for Grade 9 Spelling</b>						
<b>Fixed Effect</b>		<b>Coefficient</b>	<b>Standard Error</b>	<b>Approx. T-Ratio</b>	<b>df</b>	<b>p-value</b>
For INTRCPT1, $\beta_{0j}$	<i>INTRCPT2, <math>\gamma_{00}</math></i>	<b>51.27</b>	<b>1.61</b>	<b>31.87</b>	<b>15</b>	<b>.000</b>
	<i>CONDIT, <math>\gamma_{01}</math></i>	1.38	2.17	.64	15	.533
For SPELPRE slope, $\beta_{1j}$	<i>INTRCPT2, <math>\gamma_{10}</math></i>	<b>.68</b>	<b>.03</b>	<b>25.6</b>	<b>732</b>	<b>.000</b>
For MOTIVATION slope, $\beta_{2j}$	<i>INTRCPT2, <math>\gamma_{20}</math></i>	<b>3.00</b>	<b>.86</b>	<b>3.45</b>	<b>732</b>	<b>.001</b>
For AFRICAN-AMERICAN slope, $\beta_{3j}$	<i>INTRCPT2, <math>\gamma_{30}</math></i>	-1.77	2.19	-.81	732	.420
For LATINO slope, $\beta_{4j}$	<i>INTRCPT2, <math>\gamma_{40}</math></i>	-1.27	1.77	-.72	732	.475
For OTHER slope, $\beta_{5j}$	<i>INTRCPT2, <math>\gamma_{50}</math></i>	2.75	1.85	1.48	732	.139

In summary, these results suggest that students in the treatment group did not outperform students in the control group, regardless of grade level or English Language Arts

achievement outcome. As anticipated, several student survey composites (e.g., self-efficacy, motivation, etc.) did significantly predict reading achievement, which is consistent with the research literature (see Baker & Wigfield, 1999, as an example). Student ethnicity also revealed expected relationships with students' achievement, such that African American grade 9 students and Latino grade 7 students had lower reading achievement than students across other ethnicities. Unfortunately, there were also no cross-level interactions between ethnicity and group; ethnic minorities in the treatment group did not outperform ethnic minorities in the control group at either grade level.

### Implementation and Achievement in English Language Arts

In this section, we expand one of the research questions addressed in the previous section, specifically: What is the relationship between program implementation and student achievement in reading and language arts? To address this, we collapsed across group and analyzed the relationship between teacher level of implementation (as defined in Section Four) and students' percentile rank posttest scores on the ITBS (grade 7) or the ITED (grade 9). These results are illustrated in Table 17.

**Table 17. Level of Program Implementation and Student Achievement Scores**

Implementation Level	Grade 7 ITBS Posttest Scores		Grade 9 ITED Posttest Scores	
	Reading Mean (SD)	Language Mean (SD)	Reading Mean (SD)	Spelling Mean (SD)
Low	49.35 (24.72)	43.79 (25.88)	47.42 (26.16)	50.81 (25.96)
Medium	46.04 (26.26)	39.89 (25.89)	48.73 (25.25)	54.16 (26.70)
High	62.94 (27.86)	50.97 (27.23)	53.69 (23.86)	48.83 (24.08)

This chart suggests that grade 7 teachers who were rated “high implementers” had students who performed better (by approximately 11-14 percentile ranks in Reading and Language) than students embedded within classes where teachers' level of implementation was rated “low” or “medium”. However, this pattern of results was not evident for grade 9 teachers. There was no difference in students' posttest scores on Reading or Spelling as a

function of teacher implementation level. This helps explain, in part, why this variable (IMPLEMENT) was significant in the HLM analyses for grade 7 students but not for grade 9 students.

### **Student Attitudes towards English Language Arts and Achievement**

In this section, we address the final research question: What is the relationship among students' attitudes toward reading and language arts and their achievement in reading and language arts (e.g., vocabulary, reading comprehension, spelling, language and writing skills)? Although the HLM results, in general, showed no significant differences between treatment and control groups across both grade 7 and grade 9 students, it was important to examine whether there were any treatment effects when analyzing the change in students' survey responses. Achievement gains are difficult to realize in one academic year, especially when students in the treatment and control are covering similar skills and reading selections. However, given the nature of the *Literature (Penguin Edition)* textbook series, it was possible that students' affective ratings across the year may have shown differential effects for treatment and control students.

To examine this issue, we conducted four Repeated Measures Analysis of Variance (ANOVA), one for each of four factor composites (i.e., intrinsic motivation, self-efficacy, teacher influence, and parent involvement) derived from the Exploratory Factor Analysis on student posttest responses (see Appendix N). Using two independent variables (condition and grade) and analyzing the change in students' affective ratings from pretest to posttest, several noteworthy findings emerged. First, for the intrinsic motivation composite, there was a main effect of group ( $F(1, 1,379) = 5.31, p < .05$ ) and grade ( $F(1, 1,379) = 6.51, p < .05$ ) as well as a within subjects interaction between grade and change scores on the intrinsic motivation composite ( $F(1, 1,379) = 8.60, p < .05$ ). These results indicated that treatment students were more intrinsically motivated than control students at both pretest and posttest (but the rate of change from pretest to posttest was the same for treatment and control students). The significant interaction indicates that grade 7 students increased their intrinsic motivation to a greater extent than grade 9 students. There was no three-way interaction between group, grade, and change in intrinsic motivation.

Second, there was a main effect of grade on teacher influence ( $F(1, 1,405) = 29.55, p < .0001$ ) and self-efficacy ( $F(1, 1,346) = 8.24, p < .01$ ). For teacher influence, grade 7 students rated their teacher higher than grade 9 students at both pretest and posttest.

However, on self-efficacy, grade 9 students had higher self-efficacy than grade 7 students at pretest and posttest.

Third, for the parent involvement composite, there was a main effect of group ( $F(1, 1,386)=6.24, p < .05$ ) and grade ( $F(1, 1,386) = 37.19, p <.0001$ ) as well as a within subjects interaction between grade and change scores on parent involvement ( $F(1, 1,386) = 5.13, p < .05$ ). These results indicated that treatment students had higher parent involvement than control students at both pretest and posttest (but the rate of change from pretest to posttest was the same for treatment and control students). Grade 7 students also had higher parent involvement than grade 9 students. The significant interaction indicates that grade 9 students rated parent involvement lower at posttest than at pretest, and to a greater extent than grade 7 students. There was no three-way interaction between group, grade, and change in intrinsic motivation.

Together, these results suggest that in general, the rate of change in students' affective ratings were similar across treatment and control students. That is, treatment students did not increase their affective ratings more so than students in the control group. Also, except for self-efficacy, grade 7 students generally had higher affective ratings than grade 9 students at both pretest and posttest.

## SECTION SIX: PRODUCT SATISFACTION

In this section, we examine the extent to which treatment teachers and students were satisfied with their *Literature (Penguin Edition)* textbooks. Four primary data sources, both qualitative and quantitative, were used to address this issue: (1) posttest student survey; (2) teacher interviews<sup>7</sup>; (3) teacher satisfaction survey (which included questions related to their own satisfaction as well as an assessment of their students' most and least favorite reading selection); and (4) evaluator observations. Although we chose not to compare treatment with control teacher responses in this section (primarily due to sample size constraints), we did compare students' product satisfaction by condition. At the conclusion of this section, we provide suggestions for improving the textbook and/or ancillary materials associated with the Prentice Hall *Literature (Penguin Edition)* curriculum.

### Teachers' Ratings of Product Satisfaction with Core Textbook Components

Teachers provided satisfaction ratings for 12 core components of the Prentice Hall *Literature (Penguin Edition)* textbook based on a five-point scale ranging from "Poor" to "Excellent". Since these core components appear in both 7<sup>th</sup> and 9<sup>th</sup> grade textbooks, ratings were collapsed across grade levels. As shown in Table 18, most teachers rated each component from "Average" to "Excellent". In fact, mean ratings across each component ranged from 3.66 to 4.33, indicating that most teachers generally rated each component "Good." The highest rated component was *From the Author's Desk* (a section which includes a narrative from the author concerning the proceeding story and a brief summary of the literary themes that apply to the story), given that 67% of teachers rated it as "Excellent". In fact, one teacher mentioned, "*The 'From the Author's Desk' DVD is absolutely wonderful. I think it is a great way to introduce the students into how an author writes and the ways they are inspired to write.*" However, the majority of the components were rated "Excellent" by only a handful of teachers. The lowest rated components were the *Writing Workshops* and *Spelling Workshops*. These workshops provide suggestions for student practice of concepts covered in the textbook. One teacher reported that he/she "*didn't believe that the textbook teaches the writing process very well. The topics of the workshops are fine, but the textbook doesn't really focus on writing skills. For example, there were three sentences on how to write a thesis statement. This is something that takes me at least a class period to teach, and then more time for them to identify and practice writing effective thesis statements. The*

---

<sup>7</sup> For complete transcripts of teacher interviews, please contact the principal investigator.

same would go for paragraph writing and organizing.” It is important to note that although these *Spelling* and *Writing Workshops* were rated lowest of the twelve, some teachers remained satisfied with them overall.

**Table 18. Overall Product Satisfaction of the Prentice Hall *Literature (Penguin Edition)* Core Curriculum Components**

Core Components of the Prentice Hall <i>Literature (Penguin Edition)</i> Curriculum	Number of Teachers Reporting Using Each Curriculum Component	Number of Teachers Responding “Excellent” (Percent “Excellent” of those who responded)	Mean* (SD)	Range
<i>From the Author’s Desk</i>	12	8 (67%)	4.33 (.99)	3-5
<i>Reading and Vocabulary Preview</i>	12	3 (25%)	3.83 (.84)	3-5
<i>Build Skills</i>	12	4 (33%)	4.17 (.72)	3-5
<i>Build Understanding</i>	12	4 (33%)	4.08 (.79)	3-5
<i>Build Language Skills</i>	12	4 (33%)	4.00 (.85)	3-5
<i>Apply the Skills</i>	12	4 (33%)	4.00 (.85)	3-5
<i>Reading and Vocabulary Skills Review</i>	12	3 (25%)	3.83 (.84)	3-5
<i>Monitor Your Progress: Assessment Practice</i>	9	2 (22%)	3.67 (.87)	3-5
<i>Writing Workshop</i>	11	2 (18%)	3.36 (1.29)	1-5
<i>Spelling Workshop</i>	9	1 (11%)	3.56 (.88)	2-5
<i>Reading Informational Materials</i>	12	3 (25%)	3.92 (.79)	3-5
<i>Comparing Literary Works</i>	12	4 (33%)	4.17 (.72)	3-5

\* Rated on a five-point scale from 1 “Poor” to 5 “Excellent.”

In response to informal feedback the researchers received during the fall observations, it was clear that there was some variability between the grade 7 and grade 9 teachers in relation to the depth and breadth of the curriculum. The teacher satisfaction survey was created in an effort to elicit formal feedback regarding these aspects of the curriculum.

### Grade 7 Teacher Perceptions of Breadth and Depth

The grade 7 teachers varied in their responses when asked whether the curriculum contains sufficient breadth in terms of reading comprehension, vocabulary, spelling, and grammar. Across each unit in which teachers worked, all teachers agreed there was sufficient breadth and depth in reading comprehension and most teachers thought there was sufficient breadth in vocabulary. For example, one teacher reported, *“There is academic vocabulary, content specific vocabulary and vocabulary for the specific pieces of literature. Skills are very covered.”* However, some teachers considered the breadth and depth insufficient in spelling, grammar, and some vocabulary exercises. For example, one teacher indicated that the *“[textbook] could do a better job of vocabulary exercises. There is just barely enough there to get the students started and then you have to start pulling from other sources to dig deeper. The same goes for spelling and grammar.”* Other teachers mentioned that the *“practices and questions and reading warm-ups were very helpful. There was not much mention or practice of the spelling and grammar”* and *“that grammar is a very small portion and somewhat unrelated to the sections.”*

### Grade 9 Teacher Perceptions of Breadth and Depth

Similarly, the grade 9 teachers varied in their responses when asked whether the curriculum contains sufficient breadth in terms of reading comprehension, vocabulary, spelling, and grammar. Contrary to the grade 7 teachers, the grade 9 teachers, in general, thought that there was sufficient depth and breadth of support in reading comprehension, spelling, and grammar. However, a few of the grade 9 treatment teachers in Colorado requested the supplemental Writing and Grammar textbook because they felt that the core textbook did not provide enough support or coverage of writing or grammar. Most teachers agreed that the textbook provided insufficient coverage of vocabulary. For example, one teacher reported that *“For about 15% of my students the vocabulary in the Reading Warm-up vocabulary lists is too easy. I have had to make up separate lists for these students.”* Another teacher reported that there was insufficient breadth in vocabulary *“because in most cases, there are only 5 or more examples. Some stories like the Cask of Amontillado have 20 or more words that need to be learned”* while another teacher said that *“[he/she] needed to add more vocabulary to Scarlet Ibis.”*

Together, these results suggest that although teachers were generally satisfied with the curriculum, specific steps could be taken to enhance the spelling and grammar support for the grade 7 curriculum and provide more (and sufficiently difficult) vocabulary words for the grade 9 curriculum.

## Students' Ratings of Product Satisfaction by Group

Students' overall product satisfaction was measured on the posttest student survey. To enable comparisons between treatment and control students, items were designed to apply to aspects of both the treatment and competitor texts.

Eight 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 and the large sample size, a stricter criterion for significance ( $p < .006$ , Bonferroni correction) was used to reduce the chance of a Type I error. As Table 19 indicates, treatment students rated seven out of the eight product satisfaction questions significantly more positively than control students. The only item that missed the significance cut-off (by .001) was whether students related to at least one author in their textbook. Regardless, in general, treatment students reported being much more satisfied with their textbook than control students. However, it is interesting to note that although these differences were statistically significant, both groups generally provided satisfaction ratings ranging from "Fair" to "Average."

**Table 19. Students' Textbook Product Satisfaction as a Function of Group**

Overall Textbook Product Satisfaction	Treatment Mean (SD) n = 886-901	Control Mean (SD) n = 725-751	p-value
I enjoy reading my ELA book.	2.75 (1.02)	2.47 (1.01)	p<.001
I like the reading selections in my ELA textbook.	2.89 (1.04)	2.64 (1.08)	p<.001
The overall layout of my ELA book makes reading the stories easier.	3.22 (.99)	3.01 (.95)	p<.001
My ELA book was boring to read.	3.08 (1.13)	3.30 (1.20)	p<.001
I have learned a great deal from my ELA book.	3.21 (.99)	3.03 (1.01)	p=.001
I personally related to at least one author in my ELA book.	2.98 (1.19)	2.82 (1.17)	p=.007
I would like to read more stories from the authors I learned about in my ELA book.	2.90 (1.09)	2.74 (1.12)	p=.003
I enjoyed learning personal info about the authors included in the textbook.	2.86 (1.13)	2.69 (1.14)	p=.003

\* Rated on a five-point scale from 1 "Poor" to 5 "Excellent."

Given that there were overall differences between the treatment and control groups in terms of product satisfaction, we examined how these product satisfaction ratings differed by grade level. We conducted another series of independent samples t-tests to compare ratings of grade 7 and grade 9 students. These results are illustrated in Table 20 below. We found significant differences between the groups such that grade 7 students rated the quality of their textbook more favorably than grade 9 students. All comparisons were statistically significant, which suggests that these results did not occur by chance. To explore whether there was a significant interaction between grade level and group, we conducted a Multivariate Analysis of Variance (MANOVA). Across each textbook quality rating, grade 7 and grade 9 students in the treatment and control groups provided similar ratings of their textbooks.

**Table 20. Students' Textbook Product Satisfaction for Treatment Students**

<b>Overall Textbook Product Satisfaction</b>	<b>Grade 7 Mean (SD) n = 401-409</b>	<b>Grade 9 Mean (SD) n = 485-492</b>	<b>p-value</b>
I enjoy reading my ELA book.	2.95 (.98)	2.58 (1.01)	p<.001
I like the reading selections in my ELA textbook.	3.09 (1.02)	2.73 (1.02)	p<.001
The overall layout of my ELA book makes reading the stories easier.	3.35 (.97)	3.11 (1.0)	p<.001
My ELA book was boring to read.	2.89 (1.14)	3.24 (1.09)	p<.001
I have learned a great deal from my ELA book.	3.41 (.98)	3.03 (.96)	p<.001
I personally related to at least one author in my ELA book.	3.10 (1.24)	2.89 (1.14)	p=.008
I would like to read more stories from the authors I learned about in my ELA book.	3.17 (1.09)	2.68 (1.04)	p<.001
I enjoyed learning personal information about the authors included in the textbook.	3.09 (1.14)	2.67 (1.08)	p<.001

\* Rated on a five-point scale from 1 "Poor" to 5 "Excellent."

### **Favorite and Least Favorite Reading Selections**

On the teacher satisfaction survey, treatment teachers were asked to report their perceptions of the reading selections students liked and disliked the most. Table 21 (grade 7) and Table 22 (grade 9) display these reading selections across each unit. It is important to note that only one or two teachers may have listed each story that is presented in the

table. This suggests that although a few teachers reported their students particularly liking (or disliking) a certain reading selection, it was clear that there was no reading selection that all students either unanimously liked or disliked.

**Table 21. Teachers Report of Students' Favorite and Least Favorite Grade 7 Reading Selections**

<b>Unit</b>	<b>Favorite Reading Selection (# of Teachers Who Reported This)</b>	<b>Least Favorite Reading Selection (# of Teachers Who Reported This)</b>
<b>Unit 1</b>	<ul style="list-style-type: none"> <li>∨ "All Summer in a Day," by Ray Bradbury (2)</li> <li>∨ "Was Tarzan a Three-Bandage Man," by Bill Cosby (1)</li> <li>∨ "Papa's Parrot," by Cynthia Rylant (1)</li> </ul>	<ul style="list-style-type: none"> <li>∨ "My Furthest Back Person," by Alex Haley (2)</li> <li>∨ "Luckiest Time of All," by Lucille Clifton (1)</li> <li>∨ "mk," by Jean Fritz (1)</li> <li>∨ "My First Free Summer," by Julia Alvarez (1)</li> </ul>
<b>Unit 2</b>	<ul style="list-style-type: none"> <li>∨ "Rikki-tikki-tavi," by Rudyard Kipling (2)</li> <li>∨ "The Treasure of Lemon Brown," by Walter Dean Myers (2)</li> <li>∨ "Seventh Grade," by Gary Soto (1)</li> <li>∨ "Zoo," by Edward D. Hoch (1)</li> </ul>	<ul style="list-style-type: none"> <li>∨ "The Bear Boy," by Joseph Bruchac (1)</li> <li>∨ "He-y, Come on O-ut," by Shinichi Hoshi (1)</li> <li>∨ From "Letters from Rifka," by Karen Hesse (1)</li> <li>∨ "Ribbons," by Laurence Yep (1)</li> <li>∨ "After Twenty Years," by O. Henry (1)</li> <li>∨ "Stolen Day," by Sherwood Anderson (1)</li> </ul>
<b>Unit 3</b>	<ul style="list-style-type: none"> <li>∨ "I Am A Native of North America," by Chief Dan George (1)</li> <li>∨ "The Night the Bed Fell," by James Thurber (1)</li> <li>∨ "What makes a Rembrandt a Rembrandt?," by Richard Muhlberger (1)</li> <li>∨ "Life Without Gravity," by Robert Zimmerman (1)</li> <li>∨ "Bernie Williams: Yankee Doodle Dandy," by Joel Poiley (1)</li> </ul>	<ul style="list-style-type: none"> <li>∨ From "In Search of Our Mothers' Gardens," by Alice Walker (1)</li> <li>∨ "Rattlesnake Hunt," by Marjorie Kinnan Rawlings (1)</li> <li>∨ "The Real Story of a Cowboy's Life," by Geoffrey C. Ward (1)</li> <li>∨ "No Gumption," by Russell Baker (1)</li> <li>∨ "I Am A Native of North America," by Chief Dan George (1)</li> </ul>
<b>Unit 4</b>	<ul style="list-style-type: none"> <li>∨ "Stopping by Woods on a Snowy Evening," by Robert Frost (1)</li> <li>∨ "Sarah Cynthia Sylvia Stout," by Shel Silverstein (1)</li> <li>∨ "The Desert is My Mother," by Pat Mora (1)</li> <li>∨ "The Highwayman," by Alfred Noyes (1)</li> </ul>	<ul style="list-style-type: none"> <li>∨ "Father William," by Lewis Carroll (1)</li> <li>∨ "Loo-Wit," by Wendy Rose (1)</li> </ul>
<b>Unit 5</b>	<ul style="list-style-type: none"> <li>∨ "The Monsters Are Due on Maple Street," by Rod Serling (1)</li> </ul>	<ul style="list-style-type: none"> <li>∨ None listed</li> </ul>
<b>Unit 6</b>	<ul style="list-style-type: none"> <li>∨ "The Other Frog Prince," by Jon Scieszka (1)</li> </ul>	<ul style="list-style-type: none"> <li>∨ "The Seasons on Earth," from Science Explorer (1)</li> </ul>

**Table 22. Favorite and Least Favorite Grade 9 Reading Selections**

<b>Unit</b>	<b>Favorite Reading Selection (# of Teachers Who Reported This)</b>	<b>Least Favorite Reading Selection (# of Teachers Who Reported This)</b>
<b>Unit 1</b>	<ul style="list-style-type: none"> <li>∇ “The Cask of Amontillado,” by Edgar Allan Poe (3)</li> <li>∇ “The Secret Life of Walter Mitty,” by James Thurber (1)</li> <li>∇ “New Directions,” by Maya Angelou (1)</li> <li>∇ “Checkouts,” by Cynthia Rylant (1)</li> <li>∇ “If I forget Thee, Oh Earth,” by Arthur C. Clarke (1)</li> <li>∇ “The Washwoman,” by Isaac Bashevis Singer (1)</li> </ul>	<ul style="list-style-type: none"> <li>∇ “The Cask of Amontillado,” by Edgar Allan Poe (1)</li> <li>∇ From “The Giant’s House,” by Elizabeth McCracken (1)</li> <li>∇ “Uncle Marcos,” by Isabel Allende (1)</li> <li>∇ From “A White House Diary,” by Lady Bird Johnson (2)</li> <li>∇ “The Washwoman,” by Isaac Bashevis Singer (1)</li> </ul>
<b>Unit 2</b>	<ul style="list-style-type: none"> <li>∇ “The Most Dangerous Game,” by Richard Connell (2)</li> <li>∇ “The Scarlet Ibis,” by James Hurst (1)</li> <li>∇ “The Necklace,” by Guy de Maupassant (1)</li> </ul>	<ul style="list-style-type: none"> <li>∇ “Blues Ain’t No Mockin Bird,” by Toni Cade Bambara (1)</li> <li>∇ “The Necklace,” by Guy de Maupassant (1)</li> <li>∇ “The Golden Kite, the Silver Wind,” by Ray Bradbury (1)</li> </ul>
<b>Unit 3</b>	<ul style="list-style-type: none"> <li>∇ “I Have a Dream,” by Martin Luther King (2)</li> <li>∇ “Before Hip-Hop Was Hip-Hop,” by Rebecca Walker (1)</li> </ul>	<ul style="list-style-type: none"> <li>∇ “First Inaugural Address,” by Franklin Delano Roosevelt (1)</li> <li>∇ “A Celebration of Grandfathers,” by Rudolfo Anaya (1)</li> </ul>
<b>Unit 4</b>	<ul style="list-style-type: none"> <li>∇ “Casey at the Bat,” by Ernest Lawrence Thayer (2)</li> <li>∇ “Macavity: The Mystery Cat,” by T. S. Eliot (1)</li> <li>∇ “I Wandered Lonely as a Cloud,” by William Wordsworth (1)</li> </ul>	<ul style="list-style-type: none"> <li>∇ “The Seven Ages of Man,” by William Shakespeare (1)</li> <li>∇ “The Writer,” by Richard Wilbur (1)</li> <li>∇ “Rocking,” by Gabriela Mistral (1)</li> </ul>
<b>Unit 5</b>	<ul style="list-style-type: none"> <li>∇ “The Tragedy of Romeo and Juliet,” by William Shakespeare (2)</li> </ul>	<ul style="list-style-type: none"> <li>∇ “The Tragedy of Romeo and Juliet,” by William Shakespeare (1)</li> <li>∇ None (1)</li> </ul>
<b>Unit 6</b>	<ul style="list-style-type: none"> <li>∇ From the “Odyssey,” by Homer (3)</li> <li>∇ From the “Odyssey: Part 2,” by Homer (1)</li> </ul>	<ul style="list-style-type: none"> <li>∇ From the “Odyssey: Part 1,” by Homer (1)</li> </ul>

### **Topics Noted by Teachers and Students as “Needing Improvement”**

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

#### *Pacing and Flexibility*

Several teachers felt frustrated that the intended pacing was too fast. General consensus across grade levels was that it was too difficult to get through the entire textbook in one school year. Also, the lack of flexibility in the curriculum (i.e., feeling “rushed” to cover

all of the model selections and multiple paired reading selections) impacted the time available to practice the skills embedded within the text.

### Organizational Structure of Curriculum

Grade 9 teachers consistently complained about the organizational structure of the curriculum. Many teachers thought it was too structured to engage high school students (i.e., not enough opportunities for student creativity and expression), too repetitive (with the same structure embedded within each unit and with the type of “drill and kill” worksheets contained in the ancillaries), and lacked a coherent organization of elements within the stories. One grade 9 teacher commented, *“the books we’d used in the past had a little bit more smooth organization of the elements of a short story. We could go through an entire unit and they could learn what plot was, and character and setting and those kinds of things in kind of a sequential manner, and it seems like all Prentice Hall wanted at the beginning as to hammer the fiction and nonfiction and not develop an awareness, really, of either one...”*.

Although more than two-thirds of grade 9 teachers expressed one or more of these concerns, no grade 7 teachers voiced any of these concerns. In fact, one grade 7 teacher commented that *“What I liked most about this version of Prentice Hall is that a skill would be introduced and then you would practice it, everything was right together, rather than scattered throughout the textbook. So I can tell the students this year really learned the skills that were being taught in each of the units.”*

### Writing Workshops

A commonly avoided aspect of the textbook was the writing workshops. This was because either there was not enough time to devote to the workshop, or because the teacher did not feel that it “tied in” with the unit or readings surrounding it. Plus, some teachers thought that the writing workshops assumed that the students were already proficient with editing and revision, which was not the case for many students.

### Depth of Curriculum

Some treatment teachers at the seventh and ninth grade level have voiced displeasure with the depth and breadth of support for grammar, spelling, writing instruction, and vocabulary development. Although only one treatment teacher reported this in every domain (i.e., grammar, writing, and vocabulary), most treatment teachers reported this inadequacy in at least one domain. In particular, teachers wanted more depth in relation to vocabulary and grammar. A couple teachers thought that the vocabulary words helped students understand the story, but were not as useful in terms of vocabulary they would use in every day life or would come across the following school year. Teachers desired more

coverage of prefixes, suffixes, root words, diagramming sentences, etc. that would help add depth to student learning of vocabulary and grammar. Also, grade 9 treatment teachers in Colorado requested the supplemental Writing and Grammar text in early fall 2007, since they did not feel the core curriculum covered, in enough depth, writing and grammar within the core text.

### Thematic Organization of the Units

One middle school teacher reversed the order of the last two units, so that they could teach drama last. Drama was not part of their state testing in grade 7 and the students enjoyed doing the plays at the very end of the year. Two other middle school teachers thought that the drama unit should come earlier in the book because they liked to teach *The Christmas Carol* around the holidays and by December they were not yet in unit 5. The same argument applied to *The Monsters Are Due on Maple Street*, which they liked to teach at Halloween.

Another middle school teacher taught unit 1 and then part 2 of unit 5, followed by part 1 of unit 5 and then returned to units 2-4, and ultimately finished with unit 6. This made more thematic sense to the teacher than the order the units were currently in. Also, one grade 9 teacher suggested that in future years he/she might skip unit 1 entirely and just cover fiction and non-fiction in the separate units given the overlap of concepts across the three units (unit 1: fiction and non-fiction, unit 2: fiction, unit 3: non-fiction).

### Curriculum Materials

The word “overwhelmed” was used by several teachers to describe their reaction to the number of supplementary materials available to them. This feeling of being overwhelmed usually led to the non-use of materials, even after initial and follow-up product training. It was difficult for teachers to determine which resources were for what purpose. One teacher did not realize for the first eight weeks of the semester that the Skills Development Workbook and Unit Resource Books were nearly identical. Once teachers realized the similarity of those two resources, they would use the one book and ignore the other. Overall, teachers appreciated having support materials (and not having to create their own tests, quizzes, etc.), but the amount could be scaled back and more guidance would be necessary on when and how to use them.

Two high school teachers also reported loose bindings that resulted in pages falling out of the student textbook.

### Curriculum Support for Required Novels

Given that teachers in both grade levels were mandated by their district to read certain novels throughout the year, it was clear that some content in the textbook had to be sacrificed to allow time for the novels. Some teachers suggested that having units or supporting material embedded within the text that correlated to the novels they read most frequently (e.g., *To Kill a Mockingbird*) would help connect the textbook (and students' relevant skill development) to outside readings.

## SECTION SEVEN: DISCUSSION

The purpose of the current study was to conduct a scientifically rigorous RCT study on the effects of the Prentice Hall *Literature (Penguin Edition)* curriculum on student achievement in English language arts (ELA) as well as student attitudes towards ELA. Teachers were randomly assigned at each study site to either treatment or control groups in grades 7 and 9, over seven different schools (four middle schools and three high schools) across three states (California, Colorado, and Illinois). Multiple process measures were collected throughout the year (online implementation logs, classroom observations, teacher interviews, and teacher surveys) and three outcome measures (ELA achievement assessment, writing assessment, and student survey addressing attitudes towards ELA) were collected during fall 2006 and, again, in spring 2007. Our primary research questions allowed us to investigate the impact of the curriculum on student achievement (and as a function of student and teacher characteristics), understand the relationship between implementation and achievement, and finally, assess product satisfaction among treatment teachers and students.

### Major Findings

According to the implementation analyses, there were substantive differences across grade level and group in relation to curriculum implementation. First, treatment teachers at both grade 7 and grade 9 covered significantly more reading selections (from the text as well as from outside sources) than control teachers at both grade levels. Second, grade 7 treatment teachers had slightly higher fidelity of implementation than grade 9 treatment teachers, meaning that grade 7 teachers covered more units and required sections than grade 9 teachers. Third, although teacher quality was fairly comparable across groups, we observed higher quality teachers at grade 7 than in grade 9 (as well as higher quality teachers in the treatment group than in the control group). These findings suggest that differences in implementation favored the treatment group over the control group and grade 7 teachers over grade 9 teachers. The group with the highest teacher quality and level of curriculum implementation was the grade 7 treatment teachers.

These implementation analyses were essential for interpreting the results of the HLM analyses on students' achievement in ELA. According to the HLM results, there were no overall differences in student achievement across treatment and control groups, after controlling for students' demographics and their prior academic background characteristics.

Students in both conditions increased approximately the same amount (a few percentile rank points) from pretest to posttest across each of the subtests of the ITBS or ITED (Reading and Language in grade 7; Reading and Spelling in grade 9). However, there were significant student-level predictors of achievement outcomes across grade levels, including students' ethnicity and self-report affective ratings of motivation and self-efficacy. School/teacher level predictors of students' achievement were also significant for grade 7 students, but not for grade 9 students. One interesting finding, commensurate with our implementation analyses, was that teachers with higher levels of implementation in grade 7 had students who scored better on one of the standardized assessments (i.e., Language subtest of the ITBS) at posttest in comparison to teachers who had lower levels of implementation. This gives us some reason to believe that if a program is implemented with high fidelity at grade 7, there is a greater likelihood that student scores will increase compared to those who implement with lower fidelity. However, given that teacher quality also significantly predicted grade 7 students' reading scores, it is difficult to isolate the effect of teacher quality from teacher implementation, especially since those two constructs are conceptually related. That is, the greater gains achieved for grade 7 students in relation to teacher quality and teacher implementation may exclusively be a function of a teacher's ability to move through the curriculum efficiently and reinforce student skills as opposed to the actual curriculum.

Determining why there were not larger effects in the treatment group than in the control group could be explained by a couple of factors. First, the content of readings selections of treatment text (*Literature, Penguin Edition*) may have been too similar to each of the control texts to produce observable differences in student achievement across condition. All students read very similar selections in the treatment and control group (e.g., *Romeo and Juliet, the Odyssey, a Christmas Carol, the Necklace*, etc.), although there were typically many more unique selections available in the treatment textbook than in the control textbooks. Second, the success of any curriculum is largely dependent on the quality with which it is implemented. By randomly assigning at the teacher level, we theoretically equated the groups in relation to teacher characteristics, including teacher quality. However, when examining the relation between implementation and outcomes, we did observe that the teachers with the highest implementation and quality did predict students' achievement outcomes. This provides support for the fact that teacher factors may play a larger role in impacting students rather than choice of curriculum. Third, separating the analysis by grade level may have impacted our ability to detect significant effects. However,

with effect sizes between treatment and control were so small that even with much larger sample size, statistical significance probably would not have been revealed. Furthermore, the variables included in this study generally favored the treatment group over the control group, including on student demographic characteristics (i.e., more Caucasian and primary English speakers in the treatment group), teacher implementation, and teacher experience. Given these inherent advantages for the treatment group, we would have expected to find significant effects if they were present.

### **Product Satisfaction**

Compared to control teachers' satisfaction ratings of the control curricula, treatment teachers were much more satisfied with the *Penguin Edition* curriculum. Reasons for teachers' increased satisfaction include the ease of differentiating instruction, the number and diversity of readings within the text, and the author model selections, to name a few. Similar satisfaction ratings were apparent when examining students' perceptions on the quality of their textbook. Based on students' posttest survey responses, treatment students rated each aspect of their textbook (e.g., "enjoy reading ELA textbook", "layout makes the stories easier to read", "enjoyed learning personal information about the authors in the textbook") significantly more positively than control students (although overall ratings were generally considered "Fair" to "Average"). Taken together, these findings suggest that both teachers and students clearly felt more satisfied with the *Penguin Edition* curriculum than control teachers and students were with their ELA curricula.

Examining these satisfaction ratings by grade level also revealed pronounced grade level effects. Grade 7 treatment teachers responded much more favorably to the *Penguin Edition* curriculum than grade 9 treatment teachers. Positive comments about the organizational structure, conceptual layout, and the quality of the content were much more likely to be reported by grade 7 than grade 9 teachers. Furthermore, grade 7 students echoed similar levels of satisfaction with their textbook when compared to grade 9 students. Across each aspect of the textbook, grade 7 students consistently reported higher satisfaction ratings than grade 9 students. One explanation for this grade level difference is supported by our analysis of curriculum implementation and coverage across units. The *Penguin Edition* curriculum appears to be more conducive to lower grade levels than higher grade levels. For example, in this study, we found that grade 9 treatment teachers desired more flexibility the structure, content, and activities contained within the text. While most grade 7 teachers perceived the regimented nature of the units as welcome, the curriculum

did not allow for in-depth writing, vocabulary development, and creativity desired by many grade 9 teachers.

However, across grade level, there were specific elements of the curriculum that teachers used often and reported favorably. For example, *From the Author's Desk* sections in the unit introductions were used by almost all teachers (and rated very positively). By an overwhelming majority, teachers reported that this element of the program was their favorite component (for themselves and their students). In fact, they often supplemented this section with the *From the Author's Desk DVD* (which was also highly rated). The lowest rated components were the Writer's Workshops, which teachers reported were either (1) not connected to the text or (2) assumed that students already knew the material.

Satisfaction ratings in the treatment group may also have been impacted by teachers' participation in the research study. That is, treatment teachers were given implementation guidelines that outlined what aspects of the text had to be implemented for the researchers to consider the curriculum "implemented as intended" by the publishers. The guidelines prescribed that units should be covered in sequence from beginning to end, and that classes should not deviate from the sequence unless absolutely necessary. The guidelines also suggested that all participating treatment teachers and their classes should cover the entire textbook in one 36-week school year. In response to these guidelines, we found several noteworthy findings.

First, the implementation pacing guidelines were not realistic given that only two treatment teachers (one in grade 7 and one in grade 9) out of the 15 were able to cover every "unit part" of every single unit. Most teachers were forced to skip entire units or parts of units in response to district mandates, district pacing calendars, standardized testing schedules, etc. Second, the implementation guidelines were generally not appropriate given the considerations that many ELA teachers needed to address throughout the year. For example, many treatment teachers did not cover the units in sequence because (1) the need to coordinate the units with their state testing (e.g., California grade 9 teachers needed to cover unit 6 early in the year to ensure that their students had a solid foundation in mythology); (2) the content of the reading selections may have been more appropriate at certain times of the year (e.g., *A Christmas Carol* at Christmas time); and/or (3) the need to coordinate with district pacing schedules. Third, although the guidelines presumed that no outside reading would be required for participation in the study, teachers at every school were mandated to read at least one outside novel. This was especially true for grade 9 teachers, who were often mandated by their district to read up to three outside novels. This

provides one plausible reason why grade 7 teachers were more likely to faithfully implement the curriculum as intended. Fourth, treatment teachers did not use the paired selections in the manner prescribed by the curriculum (i.e., selecting the reading based on the diagnostic assessment and the lexile level of the selection). Rather, teachers chose a paired reading selection based on student interest or what story teachers' thought would engage most students, regardless of difficulty level.

### **Study Limitations**

Although this study provided the strongest empirical test of the impact of the Prentice Hall *Literature (Penguin Edition)* curriculum on student achievement, there were three 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 seven schools were the only ones who were interested in participating and who agreed to abide by all research protocols. The results of this study theoretically 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 this is a new product line, it does not appear that the program has been sufficiently tested prior to implementation of this RCT. 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. Conversely, control teachers had the flexibility to supplement their curriculum 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 used and/or prepared for the first time). Although most treatment teachers reported feeling confident and prepared to implement the curriculum in fall 2006, it was fundamentally unfair to compare curriculum being implemented the first time to curriculum that has been used in previous years.

The third limitation concerned our need to separate grade 7 and grade 9 students, which limited our statistical power in the HLM analyses. Given the similarity of the curriculum and structure of the text for grades 7 and 9, as well as the pedagogy related to the curriculum across both grade levels, we anticipated combining the groups in all analyses. However, based on students' pattern of responses for the achievement results and the student survey responses, it was not appropriate, nor ethical to combine grade 7 and grade 9 students. However, we feel confident in these research findings given the general small effect sizes revealed in this study and that when analyzing the data using a more liberal test than HLM (e.g., Analysis of Covariance), similar results were revealed.

### **Future Research**

Future research related to the Literature Penguin Edition curriculum should focus on investigating the connections between program implementation and student achievement. It is clear that less experienced teachers, in particular, may require additional time to acclimate themselves and their students to the program. It is also clear that there are particular features of the program that teachers found appealing (e.g., *From the Author's Desk*) that should be investigated more fully. Limitations with conducting research across multiple states with varying content standards also seeks to complicate unnecessarily some research questions, so we would recommend more in-depth study in a few key states with the integration of state standardized test scores as an additional outcome variable to track student achievement.

Our study also revealed that those teachers who were considered "high implementers" and those with higher "teacher quality" ratings had students with higher test scores. There are multiple ways to follow up on these factors to fully test the Penguin edition curriculum. For example, teachers are much more likely to fully implement the curriculum if they are familiar with the product. Consequently, having them use the product for an entire school year prior to tracking teacher behavior and student scores will allow many teachers sufficient time to build familiarity and increase the probability of implementing the curriculum as intended. Another alternative consideration would be to alter the research design to systematically match pairs of teachers who are more or less experienced or demonstrate higher or lower levels of quality ratings and track to what extent this impacts use of the curriculum and student scores. To ensure that all teachers begin the study in a comparable manner, a third alternative research design would be to randomly assign teachers to two or

three different conditions where new curricula are being used in all groups, and then systematically compare based on this condition.

## REFERENCES

- Baker, L. & Wigfield, A. (1999). Dimensions of children's motivation for reading and their relations to reading activity and reading achievement. *Reading Research Quarterly, 34*(4), 452-477.
- Byrk, A., & Raudenbush, S. (1992). *Hierarchical Linear Models*. Sage Publications: Thousand Oaks, CA.
- Datnow, A., Borman, G, & Stringfield, S. (2000). School reform through a highly specified curriculum: Implementation and effects of the core knowledge sequence. *The Elementary School Journal, 101, No. 2*, 167-192.
- 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.
- 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.
- No Child Left Behind Act (2001). Public Law No. 107-110. 107th Congress, 110 Congressional Record 1425, 115 Stat.
- 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.
- Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation, 9*(4). Retrieved September 27, 2007 from <http://PAREonline.net/getvn.asp?v=9&n=4>

## Appendix A. Skill Areas Covered in Treatment Textbooks

**Table 1. Skill Areas Covered in the Treatment Textbooks by Grade Level.**

<b>Unit</b>	<b>Part</b>	<b>Grade 7 Skill Areas</b>	<b>Grade 9 Skill Areas</b>
1	1	Make Predictions	Context Clues*
	2	Author's Purpose	Author's Purpose
2	1	Make Inferences	Predicting
	2	Cause and Effect	Making Inferences
3	1	Main Idea and Supporting Details	Main Idea
	2	Analyze Persuasive Texts*	Fact and Opinion*
4	1	Read Fluently*	Drawing Conclusions
	2	Paraphrase	Paraphrasing
5	1	Summarize	Purpose for Reading*
	2	Draw Conclusions	Summarizing
6	1	Cultural and Historical Context*	Case and Effect
	2	Compare and Contrast	Comparison and Contrast

\* A skill area that is exclusive to that particular grade level (all other skill areas are in common across grade levels).

## Appendix B. Information for Interested Schools

### Study of the Prentice Hall *Literature (Penguin Edition)* Curriculum 2006-07: Information for Interested Schools

#### Study Description

Claremont Graduate University (CGU) will use a Randomized Control Trial (RCT) to test the Prentice Hall *Literature (Penguin Edition)* curriculum in grades seven & nine. This comprehensive literature program features new advances in author collaboration, differentiated instruction, scope and sequence of reading skills, and benchmark pacing. The study will incorporate student and teacher data to assess product efficacy. Students will complete standardized assessments and student surveys, while teachers will complete teacher surveys and online teacher implementation logs. Classroom observations and teacher interviews will also be conducted.

CGU is seeking approximately eight schools (four high schools and four middle schools) to participate in this study. Ideally, one middle school and high school would be from the same district so that four districts are represented in the study. Interested districts should be either urban or suburban and schools should have at least 750 students. Schools must agree to have participating teachers randomly assigned to “treatment” and “control” groups. Ideally, each participating school should have a minimum of four participating teachers (two “treatment” teachers and two “control” teachers). Treatment group teachers will be asked to use the Prentice Hall *Literature (Penguin Edition)* curriculum with their classes during the 2006-07 school year. Control group teachers will be asked to continue using the literature curriculum currently in place at the school. Unfortunately, schools will be excluded from this study if they are using the previous version of the Prentice Hall Literature Series (*Timeless Voices, Timeless Themes*).

#### Incentives and Training

All participating teachers will be required to complete initial study training in the summer of 2006. This study orientation and training session will take no longer than 90 minutes. All participating teachers will be paid a stipend of \$300.00 for their participation in the study. The stipend will be paid in two increments: half (\$150.00) after the initial training session and half (\$150.00) after completion of the study. One teacher participating in the study (or other designated curriculum supervisor) will serve as the study liaison and is responsible for coordinating the following: teacher trainings, receipt of materials, classroom observations, student assessments, student surveys, and teacher surveys. The teacher who serves as the study liaison will be paid an additional \$300.00. In addition to the initial study orientation and training session, treatment group teachers will participate in professional development training for the Prentice Hall *Literature* materials that will last approximately 3-4 hours. Schools will receive FREE Prentice Hall *Literature* materials worth \$7100.00 per classroom. All participating treatment group teachers will receive enough student editions of Prentice Hall *Literature (Penguin Edition)* textbooks to use during the study period and will retain them after completion of the study. All participating treatment group teachers will also receive teacher’s editions of textbooks and any necessary supplemental Prentice Hall materials. A full list of the supplemental materials is available upon request. All participating control group teachers will receive an agreed upon number of student and teacher’s edition textbooks at the completion of the study. All participating teachers will also receive a product voucher in the amount of \$500 to purchase ancillary Prentice Hall *Literature* materials. All participating schools will also receive one professional development

session from a Penguin Author whose work is included in the *Penguin Edition*. This session may be used toward professional development credits at your school.

Table 1. General Study Tasks, Time Spent, and Responsibilities

	<b>Data Collection Instruments</b>	<b>Timeline</b>	<b>Protocol Administration</b>	<b>Location of Completed Data</b>	<b>Estimated Time to Complete Task</b>	<b>Total Estimated Time Spent on Task Over Entire Year</b>
<b>Students</b>						
	ITBS (grade 7) ITED (grade 9)	Once Fall 2006 & Once Spring 2007	All Treatment and Control Teachers	Return Completed Tests to Study Liaison	ITBS=60 minutes ITED=65 minutes	ITBS=120 minutes ITED=130 minutes
	Iowa Writing Assessment	Once Fall 2006 & Once Spring 2007	All Treatment and Control Teachers	Return Completed Tests to Study Liaison	50 minutes	100 minutes
	Student Survey	Once Fall 2006 & Once Spring 2007	All Treatment and Control Teachers	Return Completed Surveys to Study Liaison	20 minutes	40 minutes
<b>Teachers</b>						
	Weekly Implementation Logs	On-going September 2006 – May 2007	All Treatment and Control Teachers	Online via the Internet	10 minutes per week	350 minutes
	Teacher Survey	Once Fall 2006 & Once Spring 2007	All Treatment and Control Teachers	Online via the Internet	15 minutes	30 minutes
	Interviews	Spring 2007 Only	CGU Researchers	CGU	30 minutes	30 minutes
<b>Environment</b>						
	Classroom Observations	Once Fall 2006 & Once Spring 2007	CGU Researchers	CGU	50 minutes per classroom	100 minutes per classroom

### Data Collection

Several research questions motivate this study, including:

1. How do student outcomes differ for students using the Prentice Hall *Literature (Penguin Edition)* program compared to other literature programs?
2. How do student outcomes differ for treatment and control students as a function of their background characteristics (e.g., gender, ethnicity, etc.) and teacher characteristics (e.g., experience)?
3. What is the relationship between program implementation and student achievement in reading and language arts?
4. What is the relationship among students' attitudes toward reading and language arts and their achievement in reading and language arts (e.g., vocabulary, reading comprehension, spelling, language and writing skills)?

To address these research questions, the study uses a combination of pre/post surveys, standardized assessments, and program implementation and monitoring. Please refer to Table 1 for time allocation for each component that will be used in the study.

### Timeline of Study Activities

#### Visit One—Summer 2006 (July/August 2006)

- CGU researchers will present study overview, expectations, and timelines for participation for all teachers (lasting 1 hour)
- CGU researchers will train all teachers how to administer student surveys and the reading and language arts assessment (lasting ½ hours)
- CGU researchers will train all teachers in using the online system to track weekly implementation (lasting 1 hour)
- A Prentice Hall consultant will conduct initial product training with treatment group teachers (lasting 3-4 hours)

#### Visit Two—Fall 2006 (September/October 2006)

- Prentice Hall consultant will conduct follow-up of product training with treatment group teachers (lasting approximately 2 hours)

#### Visit Three—Fall 2006 (November 2006)

- CGU researchers will observe each participating (treatment and control) classroom over one to two days
- CGU researchers will conduct informal interviews with all participating teachers to assess study implementation for the first semester

#### Visit Four—Spring 2007 (May 2007)

- CGU researchers will observe each participating (treatment and control) classroom over one to two days
- CGU researchers will conduct informal interviews with all participating teachers to gain additional insight into the curriculum implementation for the year

### Additional Tasks:

The study liaison will be responsible for collecting assessments administered in fall and spring and shipping them to CGU researchers (student surveys) and test publishers (reading/ language arts assessments). CGU will pay all shipping costs. The liaison is also responsible for ensuring that all teachers complete weekly implementation logs and will follow-up with individual teachers if logs are not completed on-time or as instructed.

### Researcher Background and Experience

The study will be directed by Dr. Tiffany Berry and Dr. Rebecca Eddy of the School of Behavioral and Organizational Sciences at CGU. Both primary researchers have vast experience in conducting program evaluation in the context of schools and other educational programs. Currently they are in the process of conducting a randomized control trial of the Miller and Levine (2006) Prentice Hall *Biology* curriculum. Their research has primarily focused on conditions in low-performing and generally urban schools; however, they are currently working with suburban and rural school settings in their evaluation work. The researchers are fully committed to meeting any district research requirements and protocols.

For questions regarding research design and study procedures, please contact:

Tiffany Berry, Ph.D.  
Research Assistant Professor  
Claremont Graduate University School of Behavioral and Organizational Sciences  
123 East 8th Street  
Claremont, CA 91711  
(909) 607-1540  
email: [tiffany.berry@cgu.edu](mailto:tiffany.berry@cgu.edu)  
website: <http://www.cgu.edu/pages/976.asp>

Rebecca M. Eddy, Ph.D.,  
Research Assistant Professor  
Claremont Graduate University School of Behavioral and Organizational Sciences  
123 East 8th Street  
Claremont, CA 91711  
(909) 607-1549  
email: [Rebecca.eddy@cgu.edu](mailto:Rebecca.eddy@cgu.edu)  
website: <http://www.cgu.edu/pages/1828.asp>

## Appendix C: Variables Explored during Data Analysis

### Outcome variables:

1. 7th grade ITBS reading percentile rank scores posttest
2. 7th grade ITBS language percentile rank scores posttest
3. 9th grade ITBD reading total percentile rank scores posttest
4. 9th grade ITBD spelling percentile rank scores posttest

### Student background characteristics variables:

1. proxies for prior academic achievement
  - 7th grade ITBS reading percentile rank scores pretest
  - 7th grade ITBS language percentile rank scores pretest
  - 9th grade ITBD reading total percentile rank scores pretest
  - 9th grade ITBS spelling percentile rank scores pretest
2. Writing assessment pre-and-post measures<sup>8</sup>
  - a. Content
  - b. Organization
  - c. Voice
  - d. Convention
3. Moved: did not move (0), moved from treatment to control (1), moved from control to treatment (2)
4. Student status: new student (0) and old student (1)
5. Gender: female (0) and male (1)
6. African American: yes (1) no (0)
7. Latino: yes (1) no (0)
8. Primary language: English (1) other (0)
9. Percentage English language arts classes attended
10. Pre-and-post survey composite variables: intrinsic motivation, parent involvement, teacher influence, and self-efficacy. To construct these affective composite variables, we first conducted exploratory factor analysis to examine the factor loading patterns. We then took the average of the items loaded on a factor as the composite variable for the factor (see Appendix N for lists of items).

### Teacher/classroom/school characteristics variables:

1. Group: indicator of treatment (1) or control (0) group
2. Years of teaching experience
3. Years of teaching English language arts
4. Percentage of on-line teacher logs completed
5. Total number of reading sections covered
6. Implementation scale: high (1) vs. medium/low (0)
7. Highest level of education
8. Classroom management
9. Teacher quality
10. Summation of classroom management and teacher quality
11. School size
12. Percent Latino

---

<sup>8</sup> As mentioned previously, exploratory analyses indicated that students in the treatment and control groups at both 7th and 9th grade did equally well on all of the four writing assessment measures (i.e., content, organization, voice, and convention) at both pre-and-post occasions. To streamline our analysis focus, we therefore did not run HLM on these writing assessment measures.

13. Percent African American
14. Percent other ethnicity
15. Percent free/reduced lunch
16. Percent age 25+ with college degrees
17. Percent reading proficient/advanced
18. Percent math proficient/advanced
19. Percent science proficient/advanced
20. Log median household income

## Appendix D: HLM Final Model-Grade 7 ITBS Reading

	Level 1 Descriptive Statistics				
Variable Name	N	Mean	SD	Minimum	Maximum
Pretest Reading Score (READPRE)	726	46.76	27.10	1.00	99.00
Intrinsic Motivation Indicator (MOTIVATION)	680	3.33	.78	1.00	5.00
African-American Indicator (AFRICAN-AMERICAN)	747	.10	.30	0.00	1.00
Latino Indicator (LATINO)	747	.27	.45	0.00	1.00
	Level 2 Descriptive Statistics				
Treatment Condition Indicator (CONDIT)	13	.46	.52	0.00	1.00
Years Teaching English Language Arts (YRSELA)	13	7.19	6.52	1.00	19.00
Teacher Quality Rating (TQUALITY)	13	4.46	1.45	2.00	6.00
Teacher Implementation of Curriculum (IMPLEMENT)	13	.15	.38	0.00	1.00

### Level-1 Model

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

Where

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

$\beta_{0j}$  is predicted posttest scores for students in control class with teachers who have average years of teaching ELA, average sum teacher quality rating, and whose implementation ratings were low or medium;

$\beta_{1j}$  to  $\beta_{4j}$  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  $\sigma^2$ .

## Level-2 Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(CONDIT)_j + \gamma_{02}*(YRSELA)_j + \gamma_{03}*(TQUALITY)_j + \gamma_{04}*(IMPLEMENT)_j + U_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

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

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

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

Where

$\beta_{0j}$  to  $\beta_{4j}$  are the intercept ( $\beta_{0j}$ ) and slopes ( $\beta_{1j}$  to  $\beta_{4j}$ ) from the level-1 model;

$\gamma_{00}$  to  $\gamma_{40}$  represent the mean of intercept ( $\gamma_{00}$ ) or slopes ( $\gamma_{10}$  to  $\gamma_{40}$ );

$\gamma_{01}$  to  $\gamma_{04}$  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  $\beta_{kj}$  after taking classroom characteristic variables (i.e., condition and teacher level characteristics) into consideration.

## Appendix E: HLM Final Model-Grade 7 ITBS Language

	Level 1 Descriptive Statistics				
Variable Name	N	Mean	SD	Minimum	Maximum
Pretest Language Score (LANGPRE)	722	40.99	24.63	1.0	99.00
Intrinsic Motivation Indicator (MOTIVATION)	680	3.33	.78	1.00	5.00
African-American Indicator (AFRICAN-AMERICAN)	747	.10	.30	0.00	1.00
Latino Indicator (LATINO)	747	.27	.45	0.00	1.00
	Level 2 Descriptive Statistics				
Socioeconomic status indicator (SES)	13	61.85	5.24	55.00	69.00
Treatment Condition Indicator (CONDIT)	13	.46	.52	0.00	1.00
Teacher Quality Rating (TQUALITY)	13	4.46	1.45	2.00	6.00

### Level-1 Model

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

Where

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

$\beta_{0j}$  is predicted posttest scores for students in control class with teachers who have average teacher summary quality ratings in schools with average percent students participating in free/reduced lunch program (SES);

$\beta_{1j}$  to  $\beta_{4j}$  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  $\sigma^2$ .

## Level-2 Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01} * (\text{SES})_j + \gamma_{02} * (\text{CONDIT})_j + \gamma_{03} * (\text{TQUALITY})_j + U_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

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

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

$$\beta_{4j} = \gamma_{40}$$

Where

$\beta_{0j}$  to  $\beta_{4j}$  are the intercept ( $\beta_{0j}$ ) and slopes ( $\beta_{1j}$  to  $\beta_{5j}$ ) from the level-1 model;

$\gamma_{00}$  to  $\gamma_{40}$  represent the mean of intercept ( $\gamma_{00}$ ) or slopes ( $\gamma_{10}$  to  $\gamma_{50}$ );

$\gamma_{01}$  to  $\gamma_{03}$  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  $\beta_{kj}$  after taking classroom characteristic variables (i.e., condition and teacher/school level characteristics) into consideration.

## Appendix F: HLM Final Model-Grade 9 ITED Reading Total

	Level 1 Descriptive Statistics				
Variable Name	N	Mean	SD	Minimum	Maximum
Pretest Reading Score (READPRE)	901	48.13	24.01	1.00	99.0
Attendance Indicator (ATTENDANCE)	881	94.78	4.85	52.00	100.00
Intrinsic Motivation Indicator (MOTIVATION)	906	3.24	.80	1.00	4.89
Self efficacy Indicator (EFFICACY)	905	3.52	.82	1.00	5.00
African-American Indicator (AFRICAN-AMERICAN)	941	.12	.32	0.00	1.00
Latino Indicator (LATINO)	941	.22	.42	0.00	1.00
	Level 2 Descriptive Statistics				
Treatment Condition Indicator (CONDIT)	18	.50	.51	0.00	1.00

### Level-1 Model

$$\begin{aligned}
 Y_{ij} = & \beta_{0j} + \beta_{1j} * (\text{READPRE} - \text{READPRE..})_{ij} + \beta_{2j} * (\text{ATTENDANCE} - \text{ATTENDANCE..})_{ij} \\
 & + \beta_{3j} * (\text{MOTIVATION} - \text{MOTIVATION..})_{ij} + \beta_{4j} * (\text{EFFICACY} - \text{EFFICACY..})_{ij} + \\
 & \beta_{5j} * (\text{AA-AA..})_{ij} + \beta_{6j} * (\text{LATINO-LATINO..})_{ij} + r_{ij}
 \end{aligned}$$

Where

- $Y_{ij}$  is the outcome for student  $i$  in classroom  $j$ ;
- $\beta_{0j}$  is predicted posttest scores for students in control class;
- $\beta_{1j}$  to  $\beta_{6j}$  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  $\sigma^2$ .

## Level-2 Model

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

$$\beta_{1j} = \gamma_{10}$$

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

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

$$\beta_{4j} = \gamma_{40}$$

$$\beta_{5j} = \gamma_{50}$$

$$\beta_{6j} = \gamma_{60}$$

Where

$\beta_{0j}$  to  $\beta_{6j}$  are the intercept ( $\beta_{0j}$ ) and slopes ( $\beta_{1j}$  to  $\beta_{5j}$ ) from the level-1 model;

$\gamma_{00}$  to  $\gamma_{60}$  represent the mean of intercept ( $\gamma_{00}$ ) or slopes ( $\gamma_{10}$  to  $\gamma_{50}$ );

$\gamma_{01}$  is level 2 regression coefficient that captures the effects of classroom-level variables (i.e., treatment vs. control condition in this case) on the within-classroom relationships between individual student background and student outcomes; and

$u_{0j}$  represent the variability in  $\beta_{kj}$  after taking classroom characteristic variables (i.e., condition) into consideration.

## Appendix G: HLM Final Model-Grade 9 ITED Spelling

	Level 1 Descriptive Statistics				
Variable Name	N	Mean	SD	Minimum	Maximum
Pretest Spelling Score (SPELPRE)	904	50.34	25.70	1.00	99.00
Intrinsic Motivation Indicator (MOTIVATION)	906	3.24	.80	1.00	4.89
African-American Indicator (AFRICAN-AMERICAN)	941	.12	.32	0.00	1.00
Latino Indicator (LATINO)	941	.22	.42	0.00	1.00
Other Ethnicity Indicator (OTHER)	941	.20	.40	0.00	1.00
	Level 2 Descriptive Statistics				
Treatment Condition Indicator (CONDIT)	18	.50	.51	0.00	1.00

### Level-1 Model

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

Where

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

$\beta_{0j}$  is predicted posttest scores for students in control class;

$\beta_{1j}$  to  $\beta_{5j}$  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  $\sigma^2$ .

## Level-2 Model

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

$$\beta_{1j} = \gamma_{10}$$

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

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

$$\beta_{4j} = \gamma_{40}$$

$$\beta_{5j} = \gamma_{50}$$

Where

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

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

$\gamma_{01}$  is level 2 regression coefficient that captures the effects of classroom-level variables (i.e., treatment vs. control condition in this case) on the within-classroom relationships between individual student background and student outcomes; and

$u_{0j}$  represent the variability in  $\beta_{kj}$  after taking classroom characteristic variables (i.e., condition) into consideration.