THE EFFECTS OF PEARSON *PRENTICE HALL LITERATURE* (2010) ON STUDENT PERFORMANCE: EFFICACY STUDY

Final Report

October 31, 2010

Submitted by: Rebecca M. Eddy, Ph.D.

H. Todd Ruitman, M.A., Nancy Hankel, Michelle Sloper
and special acknowledgement to Xiaoxia Newton, Ph.D.

COBBLESTONE
APPLIED RESEARCH & EVALUATION, INC.

Acknowledgements

Table of Contents

2

Tables

Figures

Executive Summary

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

Study Description, Design, and Measures

The study design was a randomized control trial (RCT) in which teachers and their corresponding classes were randomly assigned to either the treatment condition (using the *Prentice Hall Literature* program) or the control condition (using the existing language arts program at their school). The study required that treatment teachers and their students use a prescribed amount of the curricuum (including ancillary materials) to be considered appropriately implemented in classrooms during the 2009-10 school year.

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

Study Sample

Twenty-nine teachers across eight schools in four states (California, Oregon, Arizona, and Ohio) from a combination of suburban and rural areas taught using either the *Prentice Hall Literature* (2010) program (treatment) or their existing language arts program (control) in their classrooms during the study. The study included three grade levels (seventh, eighth, and tenth). Data were analyzed for 2,729 participating students in 91 separate class groups. The study sample was primarily Hispanic students and had an almost even distribution of male and female students. Teachers taught language arts for 9 years, on average. Of the 26 teachers who those completed the teacher survey,reporting, ten teachers had teaching credentials, four teachers held Bachelor of Arts degrees, and twelve had Master of Arts degrees.

Outcome Measures	
Gates MacGinitie Reading Test (GMRT)	A norm-referenced general high school and middle school reading assessment that included sub-tests measuring vocabulary knowledge (45 items) and reading comprehension (48 items) which was also combined for an overall reading score.
Metropolitan Achievement Test, 8th edition (MAT8)	A norm-referenced measure of writing achievement. Students composed written responses to picture prompts which varied with grade level and received an overall score ranging from 0 to 6 from one of two coders.
Student attitude survey	Included questions related to students' interest and enjoyment of reading, teacher's influence of learning, and self efficacy of language arts. An additional section on the posttest asked students to rate their satisfaction with elements of the *Prentice Hall Literature* (2010) program.
Implementation Measures	
Online Teacher Logs	Completed by all participating teachers weekly to report the content covered and specific program components used in their classrooms.
Classroom Observations	All study teachers and their students were observed by the research team during the study period, specifically treatment (*Prentice Hall Literature*) teachers were observed twice (fall and spring) while control teachers were observed once (fall or spring).
Teacher Interviews/ Focus Groups	Completed at the end of the study, most teachers participated in individual interviews or focus groups to discuss the program implementation and product satisfaction and usage over the duration of the school year.

Student Sample	Response Options	All Sites (*n* ~ 2729)
Gender	Male	51.9%
	Female	48.0%
Ethnicity	Caucasian	21.8%
	African American	15.0%
	Hispanic	55.1%
	American Indian	3.1%
	Asian	1.3%
	Multiple Ethnicity/Other	3.4%
Primary Language	English	86.3%
	Other	13.7%

Program Implementation

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

> **Most common Prentice Hall Literature (2010) Components Implemented:**
> Introduce the Unit *Big Question*
> Introduce the Unit Author & Forms
> Model Selections
> Reading Selections 1 & 2

Study Results

Research Question 1:

Are teachers able to successfully integrate pedagogical elements of the Understanding by Design *model using the* Prentice Hall Literature *(2010) program?*

Answer: Teachers were comfortable starting off a new unit using the *Big Question* and reported that it was one of their favorite components. Teachers were able to easily integrate this in lessons throughout the unit. Teachers

and students had extensive conversations around the *Big Question* during the unit and were less likely to cover the Applying the *Big Question* sections to close the unit. While teachers appreciated the continuity the *Applying the Big Question* provided during a unit, seventh grade students rated it as one of their least favorite elements and eighth and tenth grade students gave it low ratings as well.

> *"I like the fact that … [The* Big Question*] is thought-provoking and there's not necessarily one right answer to it."*
> **—Middle school teacher using the** *Prentice Hall Literature* **(2010) program**

Research Question 2:

How does student achievement differ for those using the Prentice Hall Literature *(2010) program compared with those using another language arts program at three specific grade levels (seventh, eighth, and tenth)?*

Answer: With regard to differences between students using the *Prentice Hall Literature* (2010) program and students using another literature program, the data analysis model suggested that the quality of implementation was a significant predictor of students' scores on the achievement measures. Where significantly positive treatment effect was observed, it was for students in the treatment groups where the implementation was rated medium or high. Where significant negative effect was observed, it was for students in the treatment where the implementation was low. Where positive effect was observed, the higher the implementation, the better the results; and where negative effect was observed, the higher the implementation, the less negative the results.

Specifically, the treatment students in medium- and high-implementing classrooms significantly outperformed the students in the control group on the GMRT vocabulary subtest. On the GMRT comprehension subtest, the low and medium implementation treatment groups were significantly outperformed by the control group while the high implementation group showed no significant difference as compared to the control group. The only significant differences detected on the overall GMRT score were the control group outperformed only the low implementing treatment group. The results of the MAT8 writing test indicated the high implementation treatment group significantly outperformed the control group.

Answer: The results of the data analysis model showed only ethnicity to be a significant predictor of achievement on the GMRT while gender and primary language were significant predictors on the MAT8. As such, only differences for these specific student characteristics were examined for each outcome achievement measure. Specifically, in the treatment group, Caucasians outperformed other ethnic groups (i.e., Latino, African American, and other ethnicity) on GMRT vocabulary subtest, GMRT comprehension subtest and GMRT overall score. The students using *Prentice Hall Literature* (2010) all showed consistent growth from pretest to posttest except on the GMRT comprehension subtest for African American students.

Unique to the MAT8, the HLM analysis showed gender and primary language as important characteristics overall (i.e., ignoring which program the student used). In order to further explore this finding, we analyzed if this was consistent when we looked at the treatment and control groups independently. Specifically, female students' outperformed male students overall; however, males still showed similar growth from pretest to posttest to the female students only in the treatment group. The control group showed no growth from pretest to posttest for both male and female students. On the other hand, non-English speaking students showed no growth from pretest to posttest on the MAT8 while English speaking students showed pretest to posttest growth. This trend was similar in both the treatment and control groups.

Answer: The HLM analysis showed that the *student's self efficacy in language arts* was a significant predictor on performance for the GMRT comprehension subtest, GMRT total score, and MAT8 writing assessment. Additionally, *students' interest and enjoyment of reading* was a significant predictor of the GMRT comprehension subtest. The results suggest that some student attitudes can have a significant positive effect on student achievement in language arts such that high ratings for *self efficacy* and *interest and enjoyment of reading* would predict higher scores. For example, the results showed that students' high *self efficacy* in language arts could impact their scores as much as about five points on the GMRT comprehension subtest.

Answer: For the treatment group, the analysis of attitudes measured through the student survey indicated *students' interest and enjoyment of reading and self efficacy in language arts* showed significant increases from pretest to posttest while *teacher's influence of learning* showed a significant decrease from pretest to posttest. Students in the control group showed similar resulted except when measuring students' *interest and enjoyment of reading* such that control students failed to show a significant increase from pretest to posttest in this area.

Product Satisfaction

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

Overall, students rated the *Prentice Hall Literature* (2010) program more favorably than students using a different language arts program.

Favorite	Least Favorite
Aesthetics: – Students rated the pictures and artwork as their favorite elements of the textbook *Online:* – The interactive vocabulary games were rated as the favorite electronic resource across all grades – Middle school students consistently enjoyed the BQ Tunes *Stories:* – Students "liked the stories…they seem maybe a little more relevant to them."	*Writing Workshop:* – Students in all grades rated the *Writing Workshops* lowest *Online:* – High school students strongly disliked the BQ Tunes One teacher reported "My kids [said] 'Oh this is horrible. The music is horrible.'" – Interactive journals and online worksheets were the electronic resources rated the lowest

Textbook Aesthetics, Design, and Layout

Students rated the artwork and pictures as their favorite elements of the textbook. Multiple teachers, however, commented either on their weekly teacher logs or during interviews or focus groups regarding the size of the book. Students disliked the size of the book, and many noted it was extremely heavy. Many teachers also commented that the book was laid out "logically as far as the units went" and was extremely easy to navigate. One teacher mentioned they "liked the way it's grouped as far as the standards are concerned…and that is effective."

Reading Selections

Many teachers commented on the improved selection and relevance of the stories and poems in the *Prentice Hall Literature* (2010) textbook and students rated the selections as their second favorite element of the textbook. Seventh and eighth grade teachers noted their students' extreme liking of and interest in the *Reality Central* element of the program. Middle school students rated it very highly on the student survey. High school teachers, however, noted many of their students found *Reality Central* to be slightly juvenile or below their level.

Teacher Instruction Components

Teachers had many complimentary comments regarding the book and the program. Some elements that were frequently mentioned were the organization of the book, especially with regard to standards, the vocabulary was realistically challenging, the grammar lessons were good (though some teachers commented they wanted more grammar, especially because the grammar was frequently tested on the Unit Review), and there were good suggestions for writing assignments. However, one consistent comment regarding the program was the pacing guide was unrealistic. Many teachers commented on the ease with which they could differentiate instruction for their students. Teachers said that having the paired selections and *Reality Central* selections and being able to choose which one would be better for each class, was extremely helpful.

Classroom Components

The audio resources were also mentioned as a favorite by both middle school and high school teachers in multiple interviews and weekly logs. One teacher noted the audio resources, and the ability to use English subtitles on the videos, were particularly helpful for students whose first language was not English. In the weekly logs, teachers commented they enjoyed the author videos, but would have liked more video resources.

Online Components

Teachers gave high remarks on the website, especially the interactive vocabulary games. Many teachers simply said their students loved having the online components. Also, many teachers and students appreciated having the full textbook available online and that it included the ability to search the pages. However, teachers expressed frustration from the beginning that the website logged them out extremely quickly and teachers could not assign homework using the website components because if a student did not have extremely high speed internet at their house they were unable to access the resources in a timely manner.

Study Conclusions

While teachers reported liking many of the *Prentice Hall Literature* (2010) components such as paired reading selections and the *Big Question*, they were not always used in practice as prescribed. Results that indicated significantly higher vocabulary scores for medium and high implementers in the treatment group were most likely due to the specific emphasis on vocabulary in the *Prentice Hall Literature* program. Treatment teachers using the *Prentice Hall Literature* program emphasized vocabulary more than in control classrooms. Reading comprehension scores increased similarly for treatment and control students which are also supported by similar student attributions regarding the influence of the textbook on their reading ability. This is contrasted with treatment students' attributions that the textbook had a more positive influence on their writing as compared with control students, which was partially supported by significantly higher student scores in the higher implementation treatment group. Enjoyment of reading increased significantly for treatment students but remained stable for control students despite the fact that both groups' ratings of their teachers influence declined. Given that fidelity to implementation played a strong role in the research findings, a stronger test of findings would be to artificially manipulate implementation level in future studies to make stronger causal statements and rule out other alternative explanations.

* Statistically significant difference
Treatment · Control
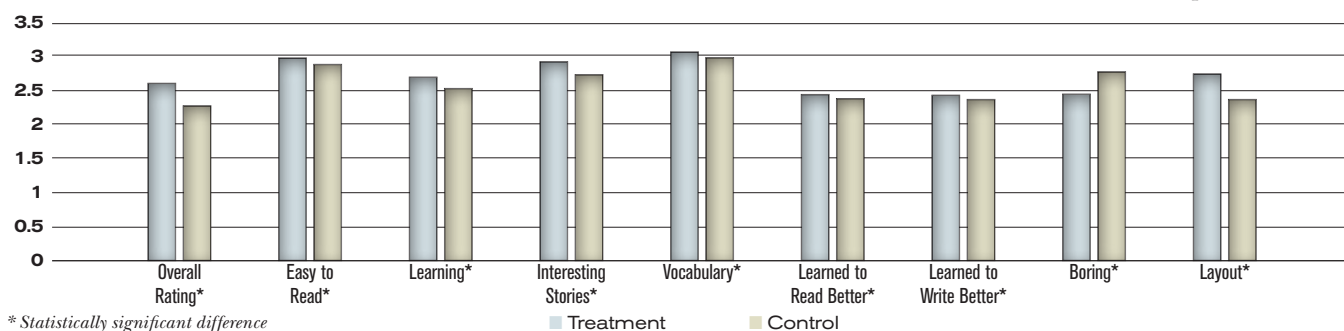
Section One: Efficacy Study Background, Study Purpose, and Program Description

Efficacy Study Background

English language arts instruction has been a mainstay of American education since the twentieth century. However, American students' reading scores have remained essentially stagnant since 1992 (U.S. Department of Education, 2009). Additionally, a recent National Endowment for the Arts study (2007) indicated that students were reading less often and less well. These facts, coupled with the general decline in book sales and reading in society as a whole, indicate a need for updated and relevant curricula in language arts classrooms. Therefore, it is important to investigate how the materials used in students' classrooms may impact students' reading abilities and interest, especially considering that young adolescent readers are not necessarily apathetic toward reading as a whole but rather toward the selections available in traditional language arts curriculum (Ivey & Broaddus, 2001). Especially in response to No Child Left Behind, language arts curricular materials have focused primarily on meeting state and national standards and secondarily targeting multiple students learning levels, as well as engaging students substantially in the thinking process and acquisition of knowledge of specific language arts concepts (Center on Education Policy, 2008).

The Pearson *Prentice Hall Literature* (2010) program is designed to address all of these priorities, which include; wide coverage of state and national standards, strategies for differentiated instruction, and engaging new features and activities. In addition, the *Prentice Hall Literature* (2010) program incorporates the *Understanding by Design* (UbD) pedagogical model (Wiggins & McTighe, 1998). Ultimately, it is important to know the extent to which the teachers integrated the UbD model into their practice while using *Prentice Hall Literature* and to determine if the *Prentice Hall Literature* (2010) program positively impacts student achievement and attitudes towards language arts.

Given the requirements of U.S. Department of Education's *What Works Clearinghouse*[1], research designs must include the use of experimental controls (among other features), usually referred to as Randomized Controlled Trials (RCTs), which are considered efficacy studies. The current report summarizes findings from the efficacy study of the *Prentice Hall Literature* (2010) program.

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

Study Purpose

An efficacy study of the *Prentice Hall Literature* (2010) program was conducted in four states (Arizona, California, Ohio, and Oregon) during the 2009-10 school year. Seventh, eighth, and tenth grade teachers and their classrooms were recruited to participate. During the study, teachers' implementation of the *Prentice Hall Literature* curriculum as well as a diverse set of student outcomes was explored. Teachers used multiple units within the textbook for the study at each site. This study focused on systematically tracking curriculum implementation, measuring students' achievement in language arts, and investigating the relationship between these elements with an assessment of the students' attitudes towards reading/motivation and product satisfaction of the *Prentice Hall Literature* (2010) program. These data provide insight into how the *Prentice Hall Literature* (2010) curriculum may affect students' attitudes and achievement in language arts during seventh, eighth, and tenth grades. The main purpose for conducting the efficacy study was to answer the following research questions:

Research Question 1:
Are teachers able to successfully integrate pedagogical elements of the Understanding by Design *model using the* Prentice Hall Literature *(2010) program?*

Research Question 2:
How does student achievement differ for those using the Prentice Hall Literature *(2010) program compared with those using another language arts program at three specific grade levels (seventh, eighth, and tenth)?*

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

Research Question 4:
What is the relationship among students' engagement and motivation in language arts and language arts achievement?

Research Question 5:
How do students using the Prentice Hall Literature *(2010) program perform from pretesting to posttesting on assessments related to student engagement and motivation in reading and achievement in language arts?*

Program Description

The *Prentice Hall Literature* (2010) program is available from sixth to twelfth grade with each grade offering unique and tailored content. It includes six units focused on a specific genre for each grade level, for example, fiction, nonfiction, poetry, etc. The design for each grade level is similar and allows us to provide a general description of the entire program noting the important differences between the seventh, eighth, and tenth grade programs that were used in the study.

The 2010 version is an update of the *Prentice Hall Literature Penguin Edition* (2007) program. Like the 2007 edition, the Literature (2010) version is designed to provide "fresh contemporary selections with classic favorites"[2] In addition, one new distinctive feature of the *Prentice Hall Literature* (2010) program includes a focus on the *Big Questions* and the *Understanding by Design* Model (UbD) (Wiggins & McTighe, 1998). The *Big Questions* are introduced at the beginning of each unit and are "integrated into the unit with activities, videos and assessments" (Personal communication, Pearson Education). The *Big Question* is designed to be referenced throughout the unit to tie together the various program components. Each

_____
[2]http://www.pearsonschool.com/index.cfm?locator=PSZl4w

unit in the different grade level texts was organized similarly with a combination of literature selections, vocabulary support and assessments embedded throughout.

> ### *The* Big Question*:*
> What is the best way to find the *truth*?
> – 7th grade, Unit 1
> Is *truth* the same for everyone?
> -8th grade, Unit 1
> Is there a difference between *reality* and *truth*?
> - 10th grade, Unit 1

Other elements of the *Prentice Hall Literature* (2010) curriculum include the integration of paired reading selections, which is based on reading difficulty so that both higher and lower level readers have the support they need. This differentiated instruction model allows the same skills to be taught to students with different levels of reading ability. Other ancillary materials are available to teachers to further enhance the teachers' ability to teach to varying levels of students. For example, the *Reality Central* textbook and accompanying writing journal were designed to provide students with additional non-fiction,

Table 1. *Prentice Hall Literature* (2010) Unit Components

Units 1-4, 6	Unit 5
Introduce the Unit *Big Question*	
Introduce the Unit author and the Unit forms	Introduce the Unit author and the Unit genre
Model Selections (2)	Model Selection
Featured Selection Pairing 1	Drama Selection 1
Featured Selection Pairing 2	
Test Practice: Reading	
Informational Texts	
Comparing Literary Works	
Writing Workshop	
Featured Selection Pairing 3	Drama Selection 2
Featured Selection Pairing 4	
Test Practice: Reading	
Informational Texts	
Comparing Literary Works	
Writing Workshop	
Applying the *Big Question*	
Vocabulary Workshop	
Communications Workshop	
Test Practice: Unit Review	

and contemporary stories that are related to paired selections within each unit. *Reality Central* texts were written below grade level to provide additional support to supplement the core *Prentice Hall Literature* (2010) text. In addition, the program's online components are used to merge the traditional classroom to the tech-savvy students of today. The program's website contains access to the same content of the student edition textbook as well as videos, audio content, and interactive activities that support reading selections and other skills taught in each of the units.

Participating treatment teachers were instructed to implement Units 1-6 throughout the school year. Table 1 shows how units are organized for seventh, eighth, and tenth grade curricula. Every unit, except Unit 5, was organized in the same manner. The table shows the main components of the unit and the order in which each component should be taught. Some components share the same name; however, the content in each component is unique in that it supports specific skills to be taught throughout the units. Also, the main difference in the organization

of Unit 1 and Unit 5 is the number of featured reading selections. Units 1 through 4 and Unit 6 have four paired reading selections while Unit 5 has two selections without the option of choosing between paired selections.

Background, Study Purpose and Program Description Summary

An efficacy study of the *Prentice Hall Literature* (2010) program was conducted during the 2009-10 school year. The program includes numerous features designed to engage students' engagement and motivation in language arts including *Understanding by Design* features such as the *Big Question*, paired reading selections targeted to lower or higher level readers, and various online components. Teacher resources are also available along with numerous ancillary materials such as the *Reality Central* textbook and journal and Study Workbooks. The study was designed to assess implementation of the curriculum in classrooms, answer research questions related to student achievement and attitudes, and to assess product satisfaction from teachers and students.

Section Two: Description of Study Design, Setting, and Sample

Study Design

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

Site Selection

Site selection began in February 2009 and continued through the summer of 2009. Initially, Pearson Education provided references to schools and districts to Cobblestone researchers who were interested in participating in the study. In addition, Cobblestone researchers identified potential sites throughout the United States by selecting specific criteria from districts listed in the National Center for Education Statistics (http://nces.ed.gov/ccd/schoolsearch/ and http://nces.ed.gov/ccd/districtsearch/). Several hundred school districts were contacted through phone and email. It is important to note that schools with diverse student ethnicity and lower-socio economic status individuals were targeted specifically for inclusion in the study to determine the impact of the program in a variety of settings. Ultimately, a majority of districts that had the most diverse group of students declined to participate in the study. This was not unexpected, as the most diverse districts tend to be concentrated in urban areas where students typically have high mobility, district research protocols are particularly stringent, and numerous competing district initiatives does not allow participation in a research study to be a priority.

Site selection began in February 2009. Recruitment focused on schools with at least two teachers with multiple sections of language arts or English classes. Of the schools that met the inclusion criteria, securing their participation occurred through initial contact with teachers or district supervisors. In total, eight sites were confirmed for participation in the study. Two of the eight participating sites, one in California and one in Arizona, were identified through Pearson. The remaining sites were identified and recruited by Cobblestone researchers. All participating teachers, site liaisons, district personnel, and Cobblestone researchers signed a Memorandum of Understanding (MOU) document to formally secure each school's participation. Most schools solicited for participation were unable to participate in the study. The most common reasons provided for declining participating included lack of interest or resources at schools to participate in an experimental study or satisfaction with a current language arts program.

Site Demographic Characteristics

As indicated earlier, a total of eight schools across four states participated in the study. Table 2 provides information about each site. The eight sites consisted of six suburban schools, with at least 1,200 students in each school and two rural sites with 700 students in each school. The rural sites and one suburban site were in primarily Caucasian communities while the remaining suburban sites were located in communities with high ethnic minority populations. With the exception of one site, all schools had at least 35% of students who were eligible for free or reduced-price lunch (as a measure of socio-economic status). Most communities had a median household income between $30,000 and $60,000.

Student Participants

Table 3 summarizes the demographic characteristics of all participating students from the eight sites. Though some attrition occurred during the school year, 2,729 students completed at least one pretest or posttest during the study and these are the students that are considered "participating." A summary of student participant demographic characteristics (including gender, ethnicity, and primary language) can be found in Appendix A.

Table 2. School Level Demographic Characteristics for Participating Sites

State		Arizona	California						Ohio	Oregon
School Site		Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8
Location*		Rural	Sub	Sub	Sub	Sub	Sub	Sub	Rural
School Size*		700	2800	1400	1200	1400	3800	1800	700
Ethnicity*	% Caucasian	71%	10%	3%	6%	9%	16%	78%	75%
	% Hispanic/Latino	27%	59%	73%	67%	60%	53%	2%	9%
	% African American	N/A	27%	20%	23%	26%	23%	14%	1%
	% Other Ethnicity	2%	2%	2%	3%	2%	7%	6%	15%
Economic Measure*	% Free & Reduced Lunch	37%	48%	81%	71%	61%	55%	9%	42%
Community Measure**	% Age 25+ With College Degree	21%	14%	7%	14%	14%	12%	39%	22%
	Median Household Income	$35,000	$58,000	$38,000	$58,000	$58,000	$38,000	$76,000	$38,000

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

**US Census 2000.

There were approximately equal numbers of male and female students at all eight of the sites. Consistent with the ethnic distributions within the eight communities included in this study, students were primarily of Caucasian descent and spoke English as their primary language in Ohio and Oregon but were more likely to report being a racial minority in the California and Arizona schools.

Table 4 summarizes parent education level for participating students, as reported on student pretest surveys. Students most often reported that they did not know their parents' education level. Those students who did report their mother and father's education levels, the largest

Table 3. Demographic Characteristics
of Student Participants

	Response Options	All Sites (*n* ~ 2729)
Gender	Male	51.9%
	Female	48.0%
Ethnicity	Caucasian	21.8%
	African American	15.0%
	Hispanic	55.1%
	American Indian	3.1%
	Asian	1.3%
	Multiple Ethnicity/Other	3.4%
Primary Language	English	86.3%
	Other	13.7%

* Note. Approximately 28% of students had missing data for primary language.

Table 4. Parent Education Level
of Student Participants

Response Options	Mother's Reported Education Level	Father's Reported Education Level
Not a High School Graduate	12%	11%
High School Graduate	20%	20%
Some College	16%	12%
Bachelor's Degree	5%	4%
Master's Degree	4%	4%
Doctoral/ Professional Degree	1%	0.5%
Don't Know/Not Stated	42%	49%

Table 5. Summary of Teacher Characteristics

	Highest Degree Attained			Teaching Experience	
	Bachelor of Arts/ Science	Teaching Credential/ Certificate	Master of Arts/ Science	Number of years teaching (average)	Number of years teaching English (average)
Treatment	2 (14%)	5 (36%)	7 (50%)	8.5	9.1
Control	2 (16%)	5 (42%)	5 (42%)	9.7	9.4
Overall	4 (15%)	10 (38%)	12 (46%)	9.2 years	9.2 years

number of students reported that both their mother and father had completed high school, and these rates were identical for mothers and fathers. Parent education levels were also fairly equivalent between treatment and control groups. See Table 4 for a parent education levels reported by student participants.

Teacher Participants

There were a total of 29 teachers who participated in the study, 4 of whom taught at multiple grade levels. On average, teachers had just over nine years' teaching experience overall and teaching English/language arts, and nearly half held master's degrees. Table 5 summarizes teacher characteristics for those who reported experience and education level (three teachers did not report this information).

Design, Setting, and Sample Summary

Twenty-nine teachers across eight schools in four states from a combination of suburban and rural areas taught using either the *Prentice Hall Literature* (2010) program (treatment) or their existing language arts program (control) in their classrooms during this efficacy study. Data were analyzed for 2,729 participating students in 91 separate class groups. The study sample was primarily Hispanic/Latino students. Teachers taught English/ language arts 9.2 years, on average, and nearly half (46%) possessed a Master's level degree.

Section Three: Description of Study Procedures and Measures

Teacher Compensation

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

Training Activities

A summary of study activities and corresponding dates can be found in Table 6. Treatment teachers used the *Prentice Hall Literature* (2010) program during the course of the 2009-10 school year. Noting that not all schools began and completed the school year at the same time, the sequence of study activities was similar across all sites. A detailed description of teacher training is documented in more detail in Appendix B.

Data Collection Measures: Curriculum Implementation

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

Teacher Implementation Logs. Each week both control and treatment teachers were required to complete online logs that described the Unit components they covered that week and resources used to teach the different components (student edition textbook, All-in-One Workbook, PHLitOnline.com, etc.). In addition, teacher logs were useful as a source of teacher reflection on their own practice or for providing informal feedback regarding use of the products or other issues with program implementation. The primary goals in developing an online teacher log system were:

Table 6. Schedule of Study Activities

	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun
Study Orientation & Product Training	→	→	→									
Student Pre-testing		→	→									
Follow-Up Training			→	→	→							
Use of *Prentice Hall Literature* (2010) Program												
Classroom Observations & Teacher Interviews				→	→	→				→	→	→
Student Post-testing										→	→	→

- Capture as accurately as possible both the *Prentice Hall Literature* (2010) content covered in classes (e.g., *Big Question*) as well as supplemental material utilized by treatment teachers.

- Allow teachers to report any activities or events whether at school or within the classroom that might have impacted their teaching or student learning.

- Reduce the strain on teachers by making the process user-friendly and efficient.

- Collect data in a way that was meaningful to researchers and could be reported back easily.

Classroom Observations. All participating classrooms were observed by at least one member of the research team on one or two occasions. The final classroom observation protocols were adapted from materials used in the pilot study. The instruments used included descriptive information such as instructional variable, classroom activities, materials used, the extent to which students were "engaged" in the lesson, etc. Separate protocols were created for the treatment and control group classrooms. Together, these data sources allowed us to understand the activities that occurred in participating classrooms throughout the efficacy study. The most useful part of tracking implementation allowed us to provide context for the quantitative results.

Teacher Focus Groups and Individual Interviews. As part of the debriefing process, a teacher interview protocol was developed for all participating (treatment and control) teachers. Questions were similar for both groups with the exception of product satisfaction questions specifically geared toward the *Prentice Hall Literature* (2010) program for treatment teachers. Twenty-eight of the twenty-nine teachers either participated in a focus group or an individual interview. Focus group and interview questions inquired about several facets of the research study, including specific questions about satisfaction with the textbook, appropriateness to their specific levels of student, and the like. Teacher interview protocols can be found in Appendix C.

Data Collection Measures: Outcomes

Participation in the study required students to complete three measures at pretest and posttest. The student outcome measures in this study were the Gates MacGinitie Reading Test (GMRT) which included two subtests—Vocabulary and Reading Comprehension; the Metropolitan Achievement Test, 8th Edition (MAT8)—a writing test; and a student attitude survey. The goal of the GMRT and MAT8 assessments was to obtain objective measures of student achievement in language arts skills to compare across schools in multiple states. These instruments were intended to measure the impact of the *Prentice*

Hall Literature (2010) curriculum in comparison to the control curriculum. The following includes a description of outcome measures used in the current study.

Standards-based Reading Assessment. A standards-based, nationally recognized reading assessment was identified to measure student learning in language arts class. The Gates-MacGinitie Reading Tests (GMRT), fourth edition, currently distributed by Riverside Publishing, is a standardized test used to determine reading ability for students by combining scores on two sub-tests: vocabulary and reading comprehension. Two levels of this test were used that were appropriate for seventh/eighth and tenth grades. National percentile ranks and grade-equivalence scores were computed from this assessment. The vocabulary section of the assessment included 45 multiple-choice questions. The reading comprehension section contained 48 multiple-choice questions.

Norm-Referenced Writing Assessment. The Metropolitan Achievement Test, 8th Edition Writing Test (MAT8) provided a norm-referenced measurement of students' writing achievement. Students composed written responses to picture prompts, which varied with grade level. Members of the research team scored the responses on a six-point holistic scale. Scores were determined by passage length, creativity, content, organization, and mechanics (grammar, spelling, etc).

Student Survey. Student surveys were administered as both a pretest and posttest to assess attitude change over the duration of the study. All students participating in the study were required to complete a self-report survey that addressed attitudes towards language arts and product satisfaction. A factor analysis was conducted by the Cobblestone research team on posttest responses to assess the reliability of underlying constructs. A more specific description of the subscales is provided next. Please see Appendix D for a list of all language arts attitude questions used on the student survey with associated reliabilities obtained for our sample.

Student Interest and Enjoyment of Reading: The survey addressed students' attitudes and motivation in literature and language arts. These factors must be measured so that we understand which factors contribute to variability in language arts achievement scores. Also, since the *Prentice Hall Literature* (2010) texts offer innovative technologies to engage students, it is important to measure whether these students' attitudes towards/interest in reading literature change for reasons attributable to factors other than textbook content.

Self Efficacy in Language Arts: The survey asked students questions that assessed their confidence in reading and writing. Student confidence in reading and writing can be directly linked to their achievement, and

therefore we were interested in determining if students' confidence increased over the duration of the study when using the *Prentice Hall Literature* (2010) program.

Teacher's Influence of Learning: Prior research has demonstrated that despite administration of the same content across all study sites, the classroom environment and interaction between students and teachers can significantly impact student achievement. In fact, Hattie (2009) in a meta-analysis of over 800 studies related to student achievement, found that teacher practices in the classroom can have a substantial effect on student learning. We investigated students' perceptions of their classrooms within these survey questions.

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

Study Procedures and Measures Summary

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

Outcome Measures	
Gates MacGinitie Reading Test (GMRT)	A norm-referenced general high school and middle school reading assessment that included sub-tests measuring vocabulary knowledge (45 items) and reading comprehension (48 items) which was also combined for an overall reading score.
Metropolitan Acheivement Test, 8th edition (MAT8)	A norm-referenced measure of writing achievement. Students composed written responses to picture prompts which varied with grade level and received an overall score ranging from 0 to 6 from one of two coders.
Student attitude survey	Included questions related to students' interest and enjoyment of reading, teacher's influence of learning, and self efficacy of language arts. An additional section on the posttest asked students to rate their satisfaction with elements of the *Prentice Hall Literature* (2010) program.
Implementation Measures	
Online Teacher Logs	Completed by all participating teachers weekly to report the content covered and specific program components used in their classrooms.
Classroom Observations	Observed by researchers, all teachers and their students participated including two times for treatment classrooms (fall and spring) and one time for control classrooms (fall or spring).
Teacher Interviews/Focus Groups	Completed at the end of the study, most teachers participated in individual interviews or focus groups to discuss the program implementation and product satisfaction and usage over the duration of the school year.

Section Four: Assessment of Curriculum Implementation

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

Treatment Curriculum Implementation

Teachers were required to adhere to specific implementation guidelines requiring the integration of specific components of the *Prentice Hall Literature* (2010) program into their classes. Guidelines for using the *Prentice Hall Literature* (2010) curriculum were reviewed during the study orientation sessions. Appendix E includes an example of implementation guidelines for seventh grade, which was identical in structure to guidelines for eighth and tenth grades. The purpose of the implementation guidelines was to ensure that treatment teachers would fully implement the *Prentice Hall Literature* (2010) curriculum as intended by the developers. These guidelines were developed with the cooperation of the research team and Pearson editorial/product management team.

Based on the established implementation guidelines, we tracked the extent to which treatment teachers followed these guidelines throughout the year by assigning specific weights to each element (e.g., required components were worth one full point, and highly recommended elements were worth one half point, with ratings established for every section of the program covered over the year). Data on teacher level of adherence was retrieved from weekly implementation logs provided by teachers for the entire school year. Teachers were assigned ratings of high, medium and low for overall implementation. (To protect teacher confidentiality, we have not provided individual ratings of teacher implementation). For level of adherence to implementation, a "High" rating corresponds to teachers covering the majority of elements of the *Prentice Hall Literature* (2010) program as required in the implementation guidelines. A "High" rating was also achieved by covering more textbook content; a "Low" rating indicates that a teacher did not fulfill a majority of the curriculum component requirements or covered fewer sections of the book. Therefore, the implementation rating was based on both the quantity of required program components used as well as the coverage of number of sections in the book. These ratings were used in later analyses to compare level of implementation to student performance in these classrooms.

Coverage of the *Prentice Hall Literature* (2010) Program

Participating teachers were required to complete weekly online logs that detailed classroom activities and textbook usage. Appendix F outlines the program components that each teacher reported on when completing in the online teacher logs, separated by grade level. For each unit, a similar format is followed in the seventh, eighth, and tenth grades, with the exception of Unit 5: Drama. Unit 5 did not contain paired reading selections or *Comparing Literary Works*. These were replaced with Drama 1 and Drama 2.

An analysis of reported activities shows that fifteen of the sixteen treatment teachers were able to implement at least part of three, four, or five units throughout the year. Additionally, the analysis shows that teachers generally implemented the first five components of each Unit: 1) *Introduce the Unit Big Question,* 2) *Introduce the Unit author and the Unit forms,* 3) *Model Selections,* 4) *Selection from Pairing 1/Drama 1 (Unit 5), and* 5) *Selection from Pairing 2 /Drama 2.* Implementation of the other Unit components was sporadic. For example, while most teachers completed *Introduce the Unit Big Question* for all units, fewer than half of teachers completed *Test Practice: Unit Review* for any units covered and even fewer completed the *Communications workshop.*

***Most common* Prentice Hall Literature (2010) Components Implemented:**

Introduce the Unit *Big Question*

Introduce the Unit Author & Forms

Model Selections

Reading Selections 1 & 2

Teachers implemented many online components as well. Every component was implemented by at least one teacher in almost all units. On average, teachers implemented about six distinct features of PHLitOnline.com throughout the year. The most frequently implemented components were the *Big Question* Video, Vocabulary Central illustrated Vocabulary Words, BQ Tunes, Vocabulary Worksheets, and the Get Connected Video. Most of the teachers in the study mentioned in interviews or in teacher logs that their classrooms were not set up to use online resources. This may explain the reason why teachers did not utilize more of the online resources. In general, middle school teachers and high school teachers utilized the same amount of resources.

> ### *Most common* Prentice Hall Literature (2010) *Online Components Implemented:*
>
> *Big Question* Video
>
> Vocabulary Central
>
> BQ Tunes
>
> Vocabulary Worksheets
>
> Get Connected Video

Implementation of *Understanding by Design* Model

> ### Research Question 1:
>
> *Are teachers able to successfully integrate pedagogical elements of the* Understanding by Design *model using the* Prentice Hall Literature *(2010) program?*

To answer to what extent teachers were able to implement the *Understanding by Design* pedagogical model, we referenced coverage reported in teacher logs, teacher interviews, and classroom observations. Appendix F provides a summary of coverage of key components of the *Prentice Hall Literature* (2010) program. Generally speaking, teachers were comfortable starting off a new unit while using the *Big Question*. We often observed teachers referring back to the *Big Question* in the middle of a unit (during observations) and students appeared familiar with the concept throughout the year. It seems that teachers were able to easily integrate this in lessons. More specific information about how much teachers and students liked the *Big Question* can be found in Section Six of this report (Product Satisfaction). However, teachers were less likely to wrap up a unit using Applying the *Big Question*. It appears that if teachers and students had extensive conversations around the *Big Question* during the unit, than it

became less necessary to integrate *Applying the Big Question* sections to close the unit.

Control Curricula Implementation

The curriculum used in control classrooms varied by state; however, there were some similarities that were observed through observations and weekly implementation logs. A clear distinction between teachers in the control group was the source of teaching material. Most teachers ($n = 10$) used a textbook to guide their instruction. These teachers reported that their districts had established pacing guides to follow. These pacing guides were generally designed to cover certain material from the textbook at certain times in the year depending on what students needed to know for state testing. To supplement the textbook, many teachers had designed their own writing and vocabulary activities. Also, these teachers read with their students at least two novels in the school year.

The other three control teachers did not use a textbook as their main source of teaching material. They generally read several novels and short stories accompanied with activities that the teachers either created themselves or found on the internet. One district in particular did not have a textbook for students and worked from a district created curriculum based on teacher created activities that had been used for several years.

Classroom Observations

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

Observation Summary. Overall, treatment teachers and their students used a wide variety of *Prentice Hall Literature* (2010) resources during classroom observations. Teachers and students routinely referred to the *Big Question* topic throughout the observations. No two classroom observations were the same, in the implementation of similar components of the *Prentice Hall Literature* (2010) program varied. For example, some teachers were observed having long discussions about the *Big Question* topic with their students, while other had the question posted in the room. Some teachers used online program components during the observations, while others assigned student

workbook exercises or had the class read a story aloud. Although students often used workbooks, teachers varied the use and presentation of these and other materials to fit their own routines. Some teachers had students read paired selections aloud, while some played the audio selection while students followed along. There was often discussion regarding the content of the story (sometimes in reference to margin notes in the text), that appeared very similar to teaching in control classrooms. Control classrooms were much more likely to have students read from novels or other sources other than a textbook, occasionally integrating other project-based lessons on reading or writing.

Comparing Classroom Environments Across Treatment and Control Groups

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

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

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

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

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

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

Qualities of Good Classroom Management:

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

---

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

Implementation Summary

To establish construct validity of our implementation fidelity measures, we assessed teachers in a variety of ways including self-reported online teacher logs of coverage of problems, interviews, and classroom observations. The level and quality of implementation varied throughout the study, but treatment teachers generally used the *Prentice Hall Literature* (2010) program in their classes during observations, which were also reported in weekly logs. However, there were a few teachers that did not consistently use the *Prentice Hall Literature* (2010) program for a variety of reasons. Most common reasons for not adhering to implementation guidelines were 1) obligations to cover other material in class such as novels or standardized test preparation; 2) dissatisfaction with some elements of the program (i.e., Writer's workshop); and 3) desire to integrate other sources of literature or language arts instruction that was not textbook based. Treatment teachers generally used many required elements of the programs in the units covered; however, very few teachers covered more than four units during the year and most reported that the recommended pacing guide was unrealistic in terms of how much time should be allotted to complete a unit.

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

In this section, we answer the major research questions involving student outcomes in achievement and attitudes. Each research question addressed in this section is listed and followed by a detailed explanation of the results obtained from the achievement and attitudes data obtained by the outcome measures (i.e., GMRT, MAT8, and student survey).

Analysis of Outcome Measures

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

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

Research Question 2:

How does student achievement differ for those using the Prentice Hall Literature *(2010) program compared with those using another language arts program at three specific grade levels (seventh, eighth, and tenth)?*

Student achievement was measured using the GMRT and MAT8. The GMRT was used to assess comprehension, vocabulary, and overall reading ability. The results from the GMRT were converted into scaled scores in order to combine the results from grades seven, eight, and ten. The MAT8 was used to assess writing ability. The raw scores for the MAT8 were also converted to scaled scores in order to combine the results for each grade level. The following sections address each component of the GMRT and the MAT8 individually. While there were no overall differences between treatment and control students, the following analyses show the implementation of the *Prentice Hall Literature* (2010) had an important impact on the results.

Gates MacGinitie Reading Tests: Vocabulary Subtest

As shown in Table 7, controlling for various student and teacher characteristics, we found two significant treatment effects (see coefficients associated with "Treatment Implementation-med" and "Treatment Implementation-high" respectively). Specifically, students in the treatment condition where the implementation of the curriculum was rated medium outperformed students in the control group on the GMRT vocabulary subtest. Similarly, students in the treatment condition where the implementation was rated high outperformed students in the control group. Furthermore, this significant difference was more pronounced than the difference observed for the medium-level implementation. In other words, the difference between the treatment and the control groups was larger for the high implementation treatment group than for the medium implementation treatment group. This indicates that the higher the implementation, the more pronounced the observed difference in the outcome as measured by the GMRT vocabulary subtest scores.

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

after identifying the student and teacher characteristics that are contained therein. With regard to student characteristics, the following covariates that were significantly associated with students' GMRT vocabulary subtest scores:

- Pretest GMRT Vocabulary score: higher pretest scores predicted higher posttest scores
- Ethnicity
  o Other ethnicity: predicted a lower score than the reference group (Caucasian)
  o African American: predicted a lower score than the reference group (Caucasian)
- Grade Level
  o Grade 7: predicted a lower score than the reference group (Grade 10)
  o Grade 8: predicted a lower score than the reference group (Grade 10)

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

To further explain and explore the results of the HLM analysis, Figure 1 shows the pretest and posttest scaled scores of the GMRT vocabulary subtest broken down by control and implementation level (i.e., low, medium, high) for the treatment group. While each of these groups increased their scores from pretest to posttest, the control group's increase was 9.8 points compared to the treatment groups' increases of 11.1 and 13.6 points for the medium and high implementation treatment groups, respectively. In terms of grade equivalents, these scores showed an average increase of less than a full grade level for the control group and an average increase of *more* than a grade level for the medium and high implementation treatment groups (see Table 8).

Table 7. HLM Results for GMRT Vocabulary Scaled Scores (*n* = 1,617)

Fixed Effect	Coefficient	Standard Error	Approx. T-Ratio	*p*-value
Pretest GMRT Vocabulary Score	0.69	0.03	21.29	**< 0.01**
Treatment Implementation-low	-0.25	3.15	-0.08	0.94
Treatment Implementation-med	5.34	2.86	1.87	**0.06**
Treatment Implementation-high	8.66	3.22	2.69	**0.01**
Male	2.71	1.65	1.65	0.10
Latino	-3.67	2.42	-1.52	0.13
African American	-6.31	3.04	-2.08	**0.04**
Other ethnicity	-11.56	3.64	-3.18	**< 0.01**
English is not primary language	0.93	2.44	0.38	0.70
Mother's education	0.74	0.76	0.97	0.33
Grade 7	-8.27	2.92	-2.84	**0.01**
Grade 8	-7.28	2.99	-2.44	**0.02**
Enjoyment and Interest in Reading	1.37	1.19	1.15	0.25
Self efficacy	1.64	1.26	1.31	0.19
Teacher classroom management	-1.38	0.95	-1.45	0.15
Teacher years of teaching experience	0.21	0.15	1.45	0.15
Intercept	166.23	17.68	9.40	**< 0.01**

Figure 1. Pretest and Posttest GMRT Vocabulary Scaled Scores: Control versus Treatment Implementation Levels
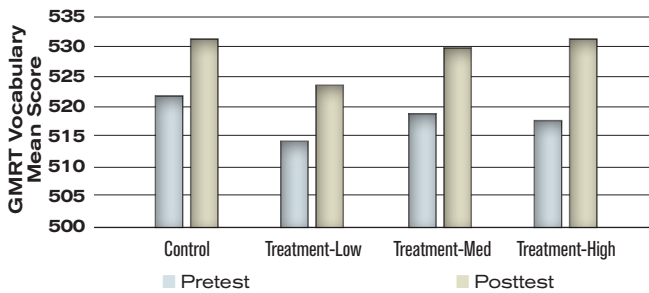

Table 8. Grade Equivalents for GMRT Vocabulary Subtest

GMRT Vocabulary Subtest			
	Pretest	Posttest	Difference
Control (*n* = 838)	6.9	7.8	0.9
Treatment-Low (*n* = 215)	6.2	7.0	0.8
Treatment-Med (*n* = 341)	6.7	7.7	1.0
Treatment-High (*n* = 223)	6.5	7.8	1.3

Gates MacGinitie Reading Tests: Reading Comprehension Subtest

A similar analysis was conducted for the Reading Comprehension subtest of the GMRT. As shown in Table 9, controlling for various student and teacher characteristics, we found that students in the treatment group where the curriculum implementation was low or medium did not perform as well as students in the control group on the GMRT comprehension subtest (see coefficients associated with "Treatment Implementation-low" and "Treatment Implementation-medium" in Table 9). In addition, the difference between the treatment and control was more pronounced for the low implementation than for the medium implementation. In other words, the results suggest that implementation did seem to matter; the lower the implementation, the worse the student performance on the GMRT Reading Comprehension subtest.

With regard to student and teacher characteristics, we found the following covariates were significantly associated with students' GMRT comprehension scores:
- Pretest GMRT Comprehension score: higher pretest scores predicted higher posttest scores
- Ethnicity
  - Other ethnicity: predicted a lower score than the reference group (Caucasian)
  - African American: predicted a lower score than the reference group (Caucasian)
- Grade Level
  - Grade 7: predicted a lower score than the reference group (Grade 10)
  - Grade 8: predicted a lower score than the reference group (Grade 10)
- Language other than English as primary language: predicted lower scores than reference group (English as primary language)
- Students' enjoyment and interest in reading: a higher rating on the self-efficacy construct of the student survey predicted a high score
- Students' self-efficacy: a higher rating on the self-efficacy construct of the student survey predicted a high score.

In addition, we found the following teacher level variable that was significantly related to students' SAT9 scores:
- Teachers' classroom management (continuous scale; 1 = low, 5 = high): better classroom management predicted higher student scores.

Again, the results from the GRMT Reading Comprehension subtest also highlight the importance of how the *Prentice Hall Literature* (2010) program was implemented. Figure 2 shows that scores from students in low implementing treatment classrooms actually dropped from pretest to posttest. Our interactions and observations of the teachers and students in these low implementation treatment classrooms would suggest that there may be other reasons for this finding besides suggesting that low implementation of the *Prentice Hall Literature* program leads to lower test scores. The more likely explanation for the lower test scores is that students in these classrooms did not take the posttest assessments seriously which led to lower test scores. Nevertheless, we can say with confidence that students in classrooms where implementation of the *Prentice Hall Literature* (2010) program was low had a different classroom experience than the students whose teachers had higher fidelity to program implementation.

Figure 2. Pretest and Posttest GMRT Reading Comprehension Scaled Scores: Control versus Treatment Implementation Levels
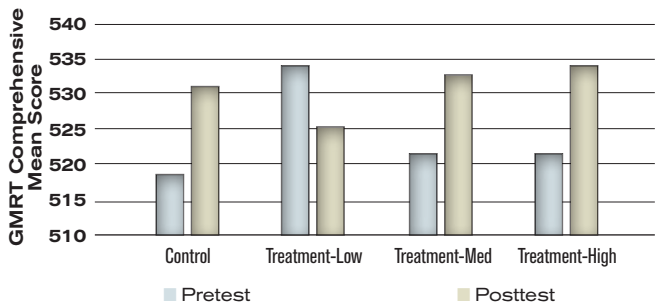
Table 9. HLM Results for GMRT Reading Comprehension Scaled Scores (n = 1,611)

Fixed Effect	Coefficient	Standard Error	Approx. T-Ratio	p-value
Pretest GMRT Comprehension	0.38	0.03	11.32	**< 0.01**
Treatment Implementation-low	-19.11	5.13	-3.73	**< 0.01**
Treatment Implementation-med	-8.43	4.73	-1.78	**0.08**
Treatment Implementation-high	4.53	5.22	0.87	0.39
Male	-2.55	2.16	-1.18	0.24
Latino	-1.98	3.28	-0.60	0.55
African American	-12.92	4.16	-3.11	**< 0.01**
other ethnicity	-8.02	4.84	-1.66	0.10
English is not primary language	-6.01	3.28	-1.83	**0.07**
Mother's education	0.75	1.02	0.74	0.46
Grade 7	-18.34	4.73	-3.88	**< 0.01**
Grade 8	-8.85	4.93	-1.79	**0.07**
Enjoyment and Interest in Reading	3.28	1.58	2.08	**0.04**
Self efficacy	3.81	1.64	2.33	**0.02**
Teacher classroom management	3.59	1.58	2.27	**0.02**
Teacher years of teaching experience	-0.21	0.24	-0.86	0.39
Intercept	316.13	19.05	16.59	**< 0.01**

Table 10 shows the breakdown of GMRT Reading Comprehension scaled scores after being converted to grade equivalents. For the Reading Comprehension subtest, the results show a similar average increase of about one grade level for all groups except the low implementation treatment group. This again supports the HLM results above.

Table 10. Grade Equivalents for GMRT Reading Comprehension Subtest

GMRT Comprehension Subtest			
	Pretest	Posttest	Difference
Control (n = 802)	6.5	7.7	1.2
Treatment-Low (n = 264)	8.0	7.1	-0.9
Treatment-Med (n = 328)	6.8	7.8	1.0
Treatment-High (n = 217)	6.7	8.0	1.3

Gates MacGinitie Reading Tests: Total Score

As shown in Table 5, controlling for various student and teacher characteristics, we found one significant difference in student performance on the GMRT total score. Specifically, students in the treatment group where the implementation level was low did not perform as well as students in the control group (see the coefficient associated with "Treatment Implementation-low" in Table 5). There were no significant differences in students' GMRT total scaled scores between students in the treatment where the implementation was medium or high and students in the control group (see the coefficients associated with "Treatment Implementation-med" and "Treatment Implementation-high" in Table 5).

Among various student and teacher variables, we observed the following as significantly related to the outcome as measured by GMRT total scaled score:

• Pretest GMRT total scaled score: higher pretest scores predicted higher posttest scores

Table 11. HLM Results for GMRT Total Scaled Scores (*n* = 1,518)

Fixed Effect	Coefficient	Standard Error	Approx. T-Ratio	*p*-value
Pretest GMRT	0.63	0.03	20.22	< 0.01
Treatment Implementation-low	-8.79	3.18	-2.76	0.01
Treatment Implementation-med	-1.29	2.89	-0.45	0.66
Treatment Implementation-high	4.33	3.24	1.34	0.18
Male	0.81	1.49	0.54	0.59
Latino	-2.52	2.26	-1.12	0.26
African American	-10.26	2.84	-3.61	< 0.01
other ethnicity	-9.20	3.30	-2.79	0.01
English is not primary language	-3.76	2.24	-1.68	0.09
Mother's education	0.91	0.70	1.30	0.19
Grade 7	-8.29	3.00	-2.76	0.01
Grade 8	-4.93	3.07	-1.60	0.11
Enjoyment and Interest in Reading	1.68	1.08	1.56	0.12
Self efficacy	2.37	1.13	2.11	0.04
Teacher classroom management	-0.18	0.98	-0.18	0.85
Teacher years of teaching experience	-0.08	0.15	-0.52	0.61
Intercept	197.84	16.97	11.66	< 0.01

- Ethnicity
  o Other ethnicity: predicted a lower score than the reference group (Caucasian)
  o African American: predicted a lower score than the reference group (Caucasian)
- Grade Level
  o Grade 7: predicted a lower score than the reference group (Grade 10)
- Students' self-efficacy: a higher rating on the self-efficacy construct of the student survey predicted a high score.

With the combination of the two GMRT subtests (i.e., Vocabulary and Reading Comprehension) to form the GMRT total score, we expected the results that are displayed in Figure 3. There were noticeable gains for control group and the medium and high implementing treatment groups; however, students in the low implementing treatment group failed to show any real improvement on the GMRT total reading score.

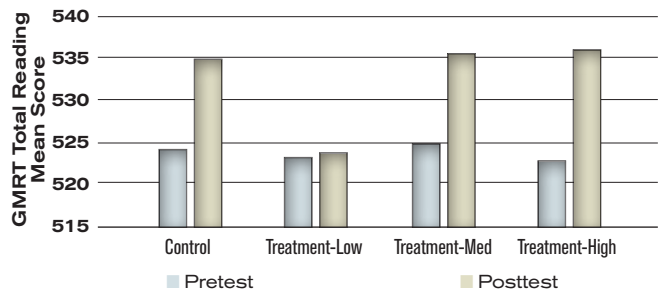Figure 3. Pretest and Posttest GMRT Total Scaled Scores: Control versus Treatment Implementation Levels

Table 12 shows the breakdown of GMRT total scaled scores after being converted to grade equivalents. Overall, students in the high implementation treatment group

showed the largest pretest to posttest gains. The medium implementation treatment group performed about the same at the control group while the low implementation group showed no improvement from pretest to posttest on the GMRT total score.

Table 12. Grade Equivalents for GMRT Total Score

Grade Equivalents for GMRT Total Score			
	Pretest	Posttest	Difference
Control (n = 774)	7.0	8.2	1.2
Treatment-Low (n = 204)	7.0	7.0	-
Treatment-Med (n = 324)	7.1	8.2	1.1
Treatment-High (n = 216)	6.9	8.3	1.4

Metropolitan 8: Writing Test

As shown in Table 13, controlling for various student and teacher variables, we found a significant treatment effect for the high implementation treatment group. Students in the treatment group where the implementation was high outperformed students in the control group on the MAT8 writing assessment (see coefficient associated with "Treatment Implementation-high" in Table 13). No significant difference was observed between the treatment and control where the implementation was low or medium on the MAT8 writing assessment.

We also found the following covariates that were statistically significant predictors of performance:

- Pretest MAT8 scaled score: higher pretest scores predicted higher posttest scores
- Gender: males predicted lower scores than the reference group (females)
- Language other than English as primary language:

Table 13. HLM Results for MAT8 Writing Assessment Scores (n = 1,536)

Fixed Effect	Coefficient	Standard Error	Approx. T-Ratio	p-value
MAT8 Pretest	0.30	0.04	8.17	< 0.01
Treatment Implementation-low	-2.63	4.61	-0.57	0.57
Treatment Implementation-med	0.39	4.02	0.10	0.92
Treatment Implementation-high	10.70	4.37	2.45	0.01
Male	-9.93	1.80	-5.51	< 0.01
Latino	-0.06	2.66	-0.02	0.98
African American	-3.11	3.41	-0.91	0.36
Other ethnicity	-4.37	3.90	-1.12	0.26
English is not primary language	-5.80	2.68	-2.17	0.03
Mother's education	0.56	0.83	0.68	0.50
Grade 7	0.84	4.00	0.21	0.83
Grade 8	-1.34	4.18	-0.32	0.75
Enjoyment and Interest in Reading	-1.00	1.29	-0.77	0.44
Self efficacy	3.48	1.35	2.57	0.01
Teacher classroom management	2.89	1.49	1.93	0.05
Teacher years of teaching experience	-0.63	0.20	-3.09	< 0.01
Intercept	417.65	24.06	17.36	< 0.01

predicted lower scores than reference group (English as primary language)

- Students' self-efficacy: a higher rating on the self-efficacy construct of the student survey predicted a high score

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

- Teachers' classroom management (continuous scale; 1 = low, 5 = high): better classroom management predicted higher student scores
- Years of teaching experience: more teaching experience by classroom teachers predicted higher student scores.

Figure 4. Pretest and Posttest MAT8 Writing Scaled Scores: Control versus Treatment Implementation Levels
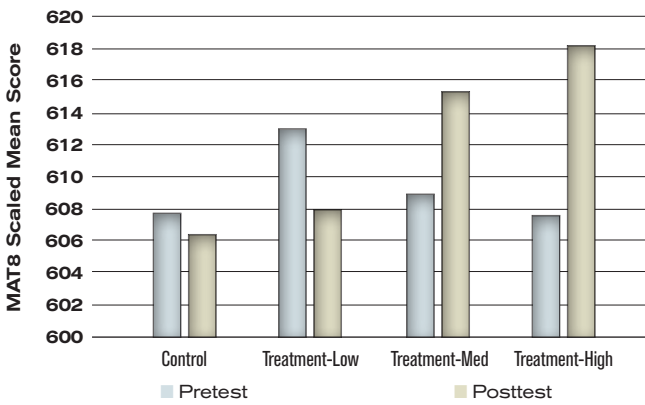
Table 14. Scaled Scores for MAT8 Writing Assessment

Scaled Scores for MAT8 Writing			
	Pretest	Posttest	Difference
Control (*n* = 713)	607.7	606.2	-1.5
Treatment-Low (*n* = 257)	612.8	608.0	-4.8
Treatment-Med (*n* = 338)	608.9	615.0	6.1
Treatment-High (*n* = 228)	607.5	618.1	10.6

The MAT8 writing assessment is the first of the outcome assessments where we see the control group's score fall from pretest to posttest; however, the difference between pretest and posttest is not a significant drop, $t(712)$ = 1.237, *ns.* Again, consistent with the GMRT results, the low implementing treatment group delivered poor results. However, we see increases from the medium and high implementation groups that used the *Prentice Hall Literature* (2010) program where the high implementation

group significantly outperformed the control group according to the HLM analysis. These results are displayed in Figure 4. Grade Equivalents were not available for the MAT8 so overall differences of the scaled scores from pretest to posttest are shown in Table 14.

Research Question 3:

How do students with different characteristics (e.g. English Learners, various ethnicities) using the Prentice Hall Literature *(2010) program perform on student-related outcomes?*

Research question three assessed how treatment students with different demographic characteristics performed on the GMRT and the MAT8. The HLM analyses showed that there were differences between ethnicities on the GMRT (see Tables 7, 9, and 11 above) and the MAT8 showed differences between gender and students primary spoken language (see Table 13 above). Our analyses found no other significant differences based on student characteristics.

Specifically for the GMRT (total score and its subtests), results showed that overall (i.e., treatment and control) African American students and students classified as *other ethnicities* performed significantly lower than Caucasian students (see Tables 7, 9, and 11 above). The "other ethnicities" variable was created for analysis to combine ethnicities (i.e., Native American, Asian, Multiethnic, and Other) because the number of students in each of these ethnic groups was too small for individual analysis. Figure 5 shows the treatment group's pretest and posttest scaled scores on the GMRT for these ethnic groups. Although differences were found in the HLM analysis between ethnic groups, the students using *Prentice Hall Literature* (2010) all showed consistent growth from pretest to posttest except on the GMRT comprehension subtest for African American students. Results for students in the control group were consistent to the treatment group (i.e., showing consistent growth from pretest to posttest); however, African American students in the control group did show growth from pretest to posttest on the GMRT comprehension subtest.

For the MAT8, we observed differences in the analysis between students' gender as well as differences between students' primary spoken language (see Table 13). Figure 6 displays the results for these groups of students in the treatment group on the MAT8 writing test. The results for the treatment group support the overall findings of the HLM analysis. While female students outperformed male students in the treatment group, males still showed similar growth from pretest to posttest to the female students. On

Figure 5. Pretest and Posttest Treatment GMRT Scaled Scores by Ethnicity
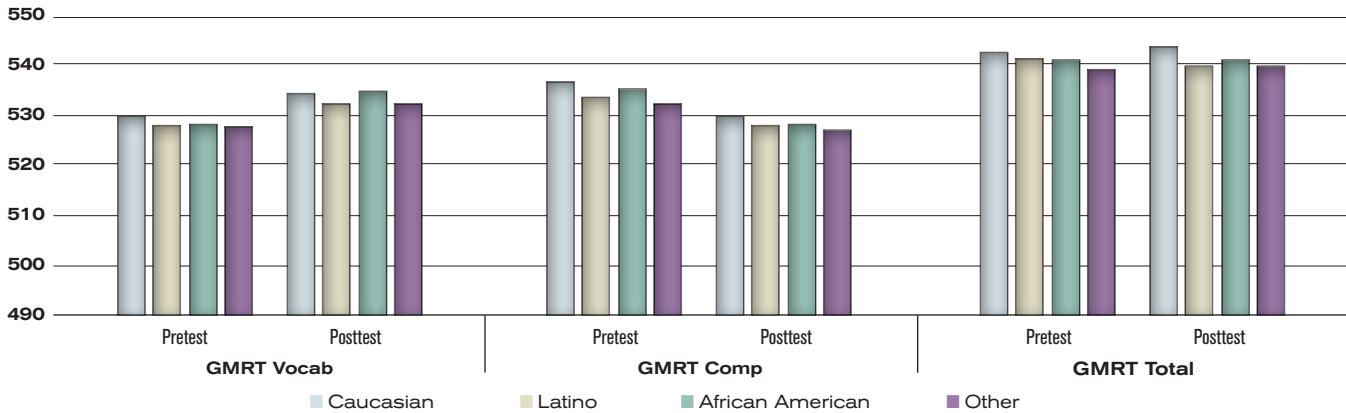
the other hand, non-English speaking students showed no growth from pretest to posttest while English speaking students showed pretest to posttest growth.

Figure 6. Pretest and Posttest Treatment MAT8 Writing Scaled Scores by Gender and Primary Language
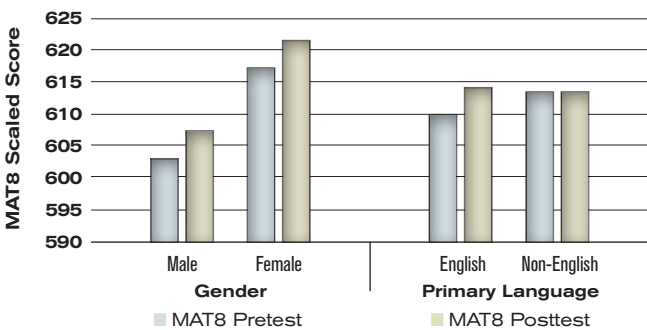
Research Question 4:

What is the relationship among students' engagement and motivation and language arts achievement?

Research question four examines the relationship between student attitudes and achievement in language arts class. Specifically, the student survey measured the overall constructs of student interest and enjoyment of reading, the classroom teacher's influence of learning, and the student's self efficacy in language arts (see Section Three of this report for a broader explanation of the student survey components). The HLM analysis showed that the student's self efficacy was a significant predictor on performance for the GMRT comprehension, GMRT total score, and MAT8 writing assessment (see Tables 9, 11, and 13). Additionally, Table 9 showed that the students' interest and enjoyment of reading was a significant predictor of the GMRT Reading Comprehension subtest. These finding suggest that the more positive certain student

attitudes are towards language arts the better students will perform on reading and writing assessments.

Research Question 5:

How do students using the Prentice Hall Literature *(2010) program perform from pretesting to posttesting on assessments related to student engagement and motivation in reading and achievement in language arts?*

Research question five examines the specific results of only those using the *Prentice Hall Literature* (2010) program from pretest to posttest for the major outcome variable of the student survey and achievement measures (i.e., GMRT and MAT8).

Treatment Student Survey Results

The following figures display the pretest to posttest results of the constructs obtained from the student survey. The results of the analyses done on each construct are briefly discussed before each figure. The results of the *t* tests performed on each construct are displayed in Table 15.

Figure 7. Student Interest and Enjoyment of Reading: Pretest and Posttest Overall and by Grade Level
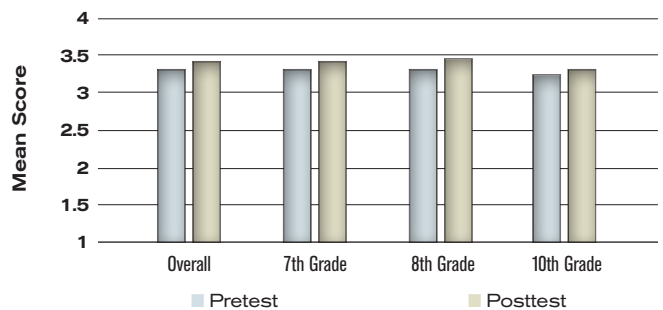

Figure 7 shows the pretest and posttest scores for student interest and enjoyment of reading in the treatment group. Overall, students' attitudes significantly increased from pretest to posttest; however, when looking

at each grade level, only the eighth grade students showed a significant increase (see Table 15). These results suggest that students overall students in the treatment group were positively influenced by the *Prentice Hall Literature* (2010) program. This statement is further bolstered by the control groups results where there was not a significant growth from pretest to posttest overall, $t(707) = .983$, *ns.*

Figure 8 shows the pretest and posttest scores for teacher's influence of learning in the treatment group. Students' attitudes significantly decreased from pretest to posttest overall and in each grade level except tenth grade where the results remain relatively unchanged from pretest to posttest (see Table 15). This result suggests that students overall rated their teachers poorer on the posttest than their initial ratings reflected on the pretest. This finding was consistent to the control group as well. This finding also suggests that student attitudes toward the textbook were not greatly influenced by their teachers.

Figure 8. Teacher's Influence of Learning: Pretest and Posttest Scores Overall and by Grade Level
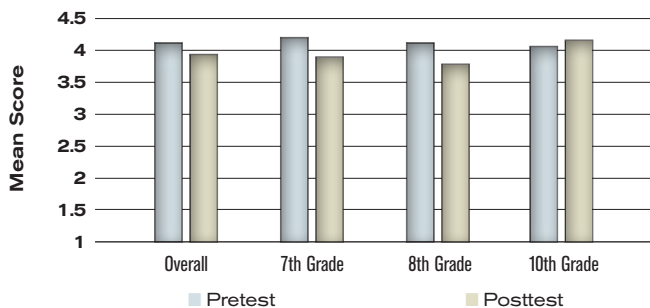
Figure 9 shows the pretest and posttest scores for self efficacy in language arts in the treatment group. This construct ultimately measured how well students rated themselves in the skills needed to succeed in language arts (i.e., reading and writing). The seventh grade remained statistically unchanged from pretest to posttest while grades eight and ten showed significant increases along with the overall combined results in the treatment group (see Table 15).

Figure 9. Self Efficacy in Language Arts: Pretest and Posttest Scores Overall and by Grade Level
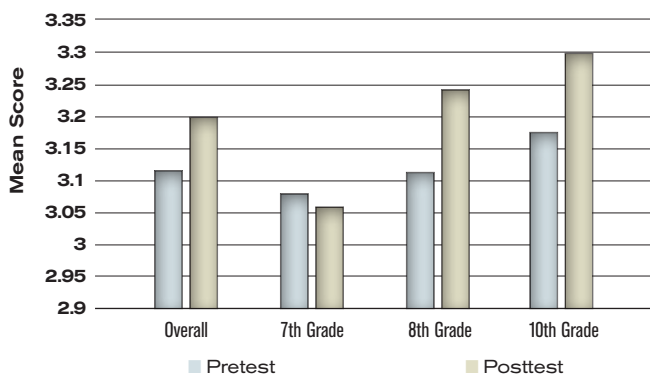
Table 15 shows the pretest and posttest scores of each construct obtained from the student survey overall and separated by grade level for the treatment group.

Treatment Student Achievement Measures

Those students using the *Prentice Hall Literature* (2010) program significantly increased their achievement scores from pretest to posttest, specifically on the GMRT and the MAT8 writing test. The following figures show the overall and grade level results for each of the achievement measures. Specifically, the overall results of the GMRT in addition to the Reading Comprehension and Vocabulary subtests are discussed followed by the results of the MAT8 writing assessment. The overall results from the t tests are show in Table 16.

Figure 10 shows the pretest and posttest scores of the GMRT Vocabulary subtest. The increases in scores from pretest to posttest are highly significant for the overall analysis and each grade level. Figure 11 shows the pretest and posttest scores for the GMRT Reading Comprehension subtest. Tenth Grade is the only area where a statistical increase is not seen from pretest to posttest. The overall analysis and grades seven and eight show significant increases. Figure 12 shows the pretest and posttest GMRT Total Reading score. The increases from pretest to posttest are highly significant for all grade levels and the overall analysis (see Table 16).

Figure 10. Treatment Students Pretest and Posttest Results for the GMRT Vocabulary Subtest
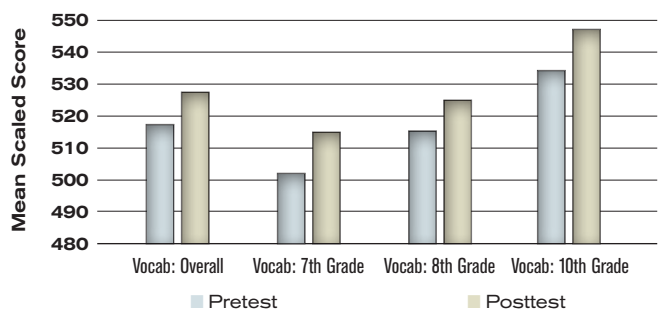
Figure 11. Treatment Students Pretest and Posttest Results for the GMRT Reading Comprehension Subtest
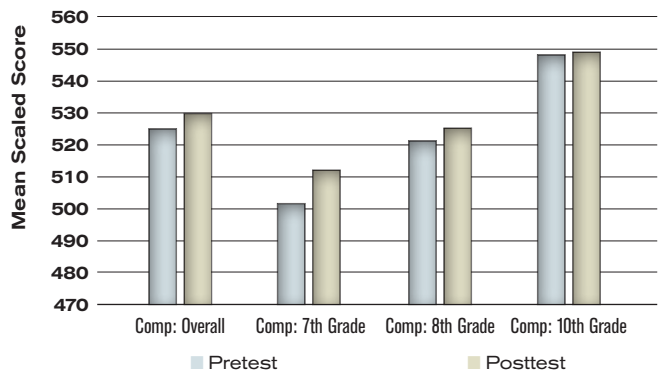
Table 15. Pretest and Posttest Student Survey Results: Overall and by Grade Level for Treatment Group

Student Interest and Enjoyment of Reading						
	Pretest Mean	**Posttest Mean**	**Difference**	***t***	***df***	***SD***
Overall	3.29	3.37	.07	3.02**	638	.65
7th Grade	3.31	3.38	.07	1.48	202	.68
8th Grade	3.32	3.42	.10	2.22*	234	.71
10th Grade	3.24	3.29	.05	1.46	200	.53
Teacher's Influence of Learning						
Overall	4.12	3.94	-.18	-5.93***	675	.79
7th Grade	4.20	3.92	-.28	-5.25***	212	.77
8th Grade	4.12	3.84	-.28	-5.45***	248	.80
10th Grade	4.06	4.09	.03	0.51	213	.05
Self Efficacy in Language Arts						
Overall	3.12	3.20	.08	2.96**	666	.72
7th Grade	3.08	3.06	-.02	-.39	213	.79
8th Grade	3.11	3.24	.13	3.20**	243	.67
10th Grade	3.17	3.30	.13	2.60*	208	.69

Note. Scale range from 1 to 5. 1 = strongly disagree, 5 = strongly agree.

* p < .05; ** p < .01; *** p < .001

Figure 12. Treatment Students Pretest and Posttest Results for the GMRT Total Scaled Score

Figure 13. Treatment Students Pretest and Posttest Results for the MAT8 Writing Assessment
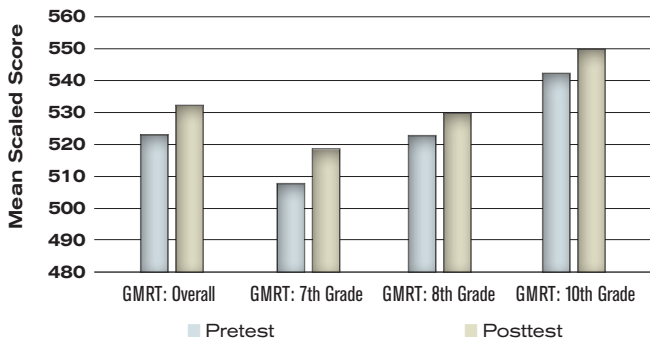
The results of the MAT8 writing assessment are displayed in Figure 13. The overall analysis shows a significant increase from pretest to posttest. Grades seven and ten also show significant increases with seventh grade showing the biggest increase. Eighth grade shows a significant decrease from pretest to posttest for this writing assessment (see Table 16).

Table 16 shows results of the paired *t* tests that were used to analyze the achievement measures. These data were also referenced to support the statistical comments of Figure 10 through Figure 13.

Attrition

We conducted attrition analyses based on a comparison of the student sample obtained from pretesting to posttesting. A full description of the study attrition and differential attrition (comparing treatment and control groups) can be found in Appendix I. Attrition analyses suggest some minor differences between treatment and control in

Table 16. Pretest and Posttest Achievement Measures: Overall and by Grade Level for Treatment Group

GMRT Vocabulary Subtest					
	Pretest Mean	Posttest Mean	Difference	t	df	SD
Overall	516.87	528.32	11.45	15.72***	778	20.33
7th Grade	502.22	516.06	13.83	9.82***	259	22.70
8th Grade	515.72	524.97	9.25	8.40***	286	18.65
10th Grade	534.72	546.22	11.51	9.09***	231	19.29
GMRT Comprehension Subtest					
Overall	525.83	530.32	4.49	3.65***	808	34.94
7th Grade	502.82	511.36	8.55	3.02**	243	44.17
8th Grade	521.63	525.96	4.33	2.12*	271	33.64
10th Grade	548.90	550.15	1.24	0.82	292	26.08
GMRT Total Score					
Overall	523.68	532.15	8.47	11.12***	743	20.78
7th Grade	506.55	518.20	11.64	7.53***	240	24.02
8th Grade	522.58	529.45	6.87	6.16***	269	18.34
10th Grade	542.68	549.73	7.05	5.51***	232	19.52
MAT8 Writing Assessment					
Overall	609.75	613.66	3.91	3.79***	822	29.55
7th Grade	592.75	606.15	13.4	7.91***	246	26.63
8th Grade	613.86	609.61	-4.25	-2.20*	277	32.11
10th Grade	620.02	623.66	3.64	2.33*	297	27.01

Note. Pretest and posttest scores are scaled.

* p < .05; ** p < .01; *** p < .001

terms of demographic characteristics that did not appear to affect the results obtained in Section Five. The analyses of the MAT8 writing test showed no differences between control and treatment for the attrition students, $t(356) = 1.12$, *ns.* There were significant differences found between students that left the study before completing a posttest GMRT (i.e., students defined as attrition). Overall, the results of the attrition analysis would favor the control group because the scores of the students who left the study in the treatment group ($M = 529.14$) were significantly higher than the students who left the study in the control group ($M = 519.08$), $t(389) = 3.33$, $p = .001$. (The results of the GMRT subtests were virtually identical so they are not discussed further.) This finding favors the control group

because students scoring higher on the pretest tended to score higher on the posttests. However, further analysis of where these differences occurred revealed that the difference between control and treatment was isolated to Latino students, $t(168) = 2.19$, $p < .05$. This result combined with the fact that Latino students made up 44% of the attrition group and were not a significant predictor of the GMRT or its subtests in the HLM analyses (see Tables 7, 9 and 11) gives us confidence that the results of the attrition analyses are not a threat to the overall results of Section Five.

Summary of Major Findings

Research Question 1: Treatment participants started units with the *Big Question* and had extensive conversations

around it during the unit, based on logs, observations and interviews. However, they were less like to cover *Applying the Big Question* sections to close the unit.

Research Question 2: HLM results suggested that the quality of implementation was a significant predictor of students' scores on the achievement measures. Where significantly positive treatment effect was observed, it was for students in the treatment groups where the implementation was rated medium or high. Where significant negative effect was observed, it was for students in the treatment where the implementation was low. Where positive effect was observed, the higher the implementation, the better the results; and where negative effect was observed, the higher the implementation, the less negative the results.

Research Question 3: The results of the HLM showed only ethnicity to be a significant predictor of achievement on the GMRT while gender and primary language were significant predictors on the MAT8. As such, only differences for these specific student characteristics were examined for each outcome achievement measure. Specifically, in the treatment group, Caucasians outperformed other ethnic groups (i.e., Latino, African American, and other ethnicity) on GMRT vocabulary subtest, GMRT comprehension subtest and GMRT overall score. The students using *Prentice Hall Literature* (2010) all showed consistent growth from pretest to posttest except on the GMRT comprehension subtest for African American students.

For the MAT8, female students outperformed male students in the treatment group; however, males still showed similar growth from pretest to posttest to the female students. On the other hand, non-English speaking students showed no growth from pretest to posttest while English speaking students showed pretest to posttest growth.

Research Question 4: The HLM analysis showed that the *student's self efficacy in language arts* was a significant predictor on performance for the GMRT comprehension, GMRT total score, and MAT8 writing assessment. Additionally, *students' interest and enjoyment of reading* was a significant predictor of the GMRT comprehension subtest.

Research Question 5: For the treatment group, the analysis of attitudes measured through the student survey indicated *students' interest and enjoyment of reading* and *self efficacy in language arts* showed significant increases from pretest to posttest while *teacher's influence of learning* showed a significant decrease from pretest to posttest.

The student achievement measures (i.e., GMRT and MAT8) showed highly significant increases from pretest to posttest when combining all grade levels for one analysis; however, the increases were less pronounced when each grade level was analyzed individually for GMRT and its subtests and the MAT8 writing assessment. Tenth grade failed to show an increase for the GMRT comprehension subtest, and the eighth grade students performed significantly worse on the posttest for the MAT8 writing assessment.

Section Six: Product Satisfaction

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

Student Edition Textbook

We wanted to understand how the *Prentice Hall Literature* (2010) program was viewed by students in comparison to other programs used by the control group. We asked all students to rate how much they liked various aspects of their textbook (see Table 13) from 1 = Strongly Disagree to 4 = Strongly Agree. Table 17 summarizes these results. We conducted paired samples *t* tests to compare control students' ratings of their textbooks and treatment students' ratings of their *Prentice Hall Literature* (2010) text. Table 17 shows that overall, students preferred all aspects of the *Prentice Hall Literature* (2010) text and rated the *Prentice Hall Literature* (2010) text significantly higher in comparison

with control students' ratings. The only exception was the area of increased reading ability, which was still rated higher by treatment students than control students, but not significantly.

Program Components

Students using the *Prentice Hall Literature* (2010) (including *Reality Central*) materials rated their level of satisfaction for each chapter component. Students were asked to rate how much they *liked* or *disliked* each part of the textbook on a scale from 1 = Strongly Dislike to 4 = Strongly Like, with an option for them to report "N/A" or "Did not use". Please see Appendix J for a summary of results related to textbook satisfaction. Students across all grades rated almost every element of the textbook in a similar manner, rendering a breakdown of results by grade redundant. Figure 1 summarizes students' favorite and least favorite elements of the overall program, as derived from student survey responses and teacher feedback from interviews and focus groups.

Overall, results show that seventh, eighth, and tenth grade students rated the pictures and artwork highest with ratings of 3.39 and 3.25 out of 4. Seventh graders rated

Table 17. Comparison of Student Ratings for Control and Treatment Textbooks

Question	Mean (Std. Deviation) for Control Textbooks	Mean (Std. Deviation) for *Prentice Hall Literature* (2010) Textbook	Difference
I like my English Language arts [*PH Literature*] textbook.	2.22 (.835)	2.64 (.826)	-.420***
My English Language arts [*PH Literature*] textbook is easy to read.	2.87 (.770)	2.96 (.674)	-.091*
I have learned a great deal from my English Language arts [*PH Literature*] textbook.	2.53 (.793)	2.72 (.774)	-.192***
My English Language arts [*PH Literature*] textbook has interesting stories.	2.69 (.865)	2.93 (.784)	-.238***
I learned several useful vocabulary words from my English Language arts [*PH Literature*] textbook.	2.92 (.819)	3.03 (.749)	-.107**
I learned how to read better from my English Language arts [*PH Literature*] textbook.	2.34 (.865)	2.42 (.829)	-.080
My English Language arts [*PH Literature*] textbook helped me learn how to write better than before I took this class.	2.29 (.898)	2.42 (.829)	-.123**
My English Language arts [*PH Literature*] textbook is boring to read.	2.73 (.990)	2.44 (.944)	.296***
I like the layout of my English Language arts [*PH Literature*] textbook.	2.32 (.910)	2.71 (.833)	-.384***

*Significant at the p< .05 level **Significant at the p< .01 level ***Significant at the p< .001 level*

Figure 14. Student Favorite and Least Favorite Textbook and Program Elements

Favorite	Least Favorite
– Students rated the *pictures* and *artwork* as their favorite elements of the textbook – The interactive *vocabulary games* were rated as the favorite electronic resource across all grades – Teachers reported that students "liked the stories… they seem maybe a little more relevant to them." – Middle school students consistently enjoyed the BQ Tunes	– Students in all grades rated the *Writing Workshops* lowest – High school students strongly disliked the BQ Tunes One teacher reported "My kids [said] 'Oh this is horrible. The music is horrible.'" – Interactive journals and online worksheets were the electronic resources rated the lowest

both the *Big Questions* and *Writing Workshops* lowest (2.28 out of 4) while eighth and tenth grade students rated just *Writing Workshops* lowest (2.30 and 2.31 out of 4). Students also rated their level of satisfaction regarding electronic resources used. Again, seventh, eighth, and tenth graders agreed on the element they liked the most, *Interactive Vocabulary Games*, with ratings of 3.09, 2.97, and 2.94 out of 4. Seventh and tenth grade students rated Interactive Journals lowest (2.25 and 2.39 out of 4) and eighth grade students rated *Online Worksheets* the lowest (2.48 out of 4). Full results of ratings of online elements are presented in Appendix K.

Understanding By Design: The Big Question

Each unit was centered on the *Big Question*—a question designed to encourage students to make connections throughout the unit based around an original inquiry. As opposed to encouraging students to learn "the right answer", the *Big Question* was designed to teach students learn how to learn holistically. Teachers and students mostly agreed on the favorite and least favorite elements of the textbook, with the exception of the *Big Question*. While teachers appreciated the continuity it provided during a unit, seventh grade students rated it as one of their least favorite elements and eighth and tenth grade students gave it low ratings as well. However, one teacher noted during an interview that for their seventh grade students, making the *Big Question* connection was difficult without a concrete example because "they're so literal."

> "I like the Big Question. And I like the way it starts out each unit."
> **–A middle school teacher**
>
> "I like the fact that … [The Big Question]'s thought-provoking and there's not necessarily one right answer to it."
> **–A middle school teacher**

Textbook Aesthetics, Design, and Layout

Students rated the artwork and pictures as their favorite elements of the textbook. Multiple teachers, however,

commented either on their weekly teacher logs or during interviews or focus groups regarding the size of the book. Students disliked the size of the book, and many noted it was extremely heavy. Additionally, a teacher mentioned there were multiple typos in the book and "it seems like it was really poorly edited."

> "The book is too heavy. They don't want to carry it. Some students did not complete a homework assignment because they did not want to take the book home…"
> **–A middle school teacher**
>
> "It was pretty easy [to navigate] compared to some of the other textbooks I've seen because everything is so well chunked."
> **–A high school teacher**

Many teachers also commented, though, that the book was laid out "logically as far as the units went" and was extremely easy to navigate. One teacher mentioned they "liked the way it's grouped as far as the standards are concerned. And that is effective."

> "[T]his really provides a lot of different varieties of literature for them to get into and…that was really, really excellent."
> **–A high school teacher**
>
> "I have students who would never read in class, who… wanted to read out loud that Cisneros story."
> **–A high school teacher**

Reading Selections

Many teachers commented on the improved selection and relevance of the stories and poems in the *Prentice Hall Literature* (2010) textbook and students rated the selections as their second favorite element of the textbook. Many teachers specifically mentioned how previous editions of textbooks included much material that students simply could not relate to, but this textbook did an excellent job of including more relatable stories.

Favorite Stories and Poems	Least Favorite Stories and Poems
7th Grade: "*Rikki Tikki Tavi*" and "*Amigo Brothers*" 8th Grade: "*Anne Frank*" 10th Grade: "*The Bridegroom*" and "*Julius Caesar*"	7th Grade: "*Ribbons*" 8th Grade: "*An American Childhood*" 10th Grade: "*Contents of a Dead Man's Pockets*" and "*Garden of Stubborn Cats*"

Reality Central

Seventh and eighth grade teachers noted their students' extreme liking of and interest in the *Reality Central* element of the program. High school teachers, however, noted many of their students found *Reality Central* to be slightly juvenile or below their level. Middle school students, though, rated it very highly on the student survey. Teachers mentioned that *Reality Central* was an easy resource to use because the selections were "short and sweet" and they could easily get through them.

Teacher Instructional Components

Teachers praised the ease of use of the *Prentice Hall Literature* (2010) program. Many commented that the book was "great" and they would use it again. However, one consistent comment regarding the program was the pacing guide. Almost every teacher mentioned, either in an interview or focus group or in weekly logs that the pacing guide was simply unrealistic, and much too ambitious given the amount of testing, testing preparation, and other school activities that occur during the year. Many teachers also reported that though having ancillary materials was helpful, the amount of materials and expectations of coverage were "overwhelming."

However, teachers had many complimentary comments regarding the book and the program. Some elements that were frequently mentioned were the organization of the book, especially with regard to standards; the vocabulary (particularly the *Big Question*/academic vocabulary) was realistically challenging; the grammar lessons were good (though some teachers commented they wanted more grammar, especially because the grammar was frequently tested on the Unit Review); and there were good suggestions for writing assignments.

> "It's a really good book for literature and…it has some good suggestions as far as writing assignments…
>
> "It had good academic vocabulary."
>
> "There's not nearly enough of the grammar."
>
> "It's set up very easily to differentiate instruction."

Many teachers commented on the ease with which they could differentiate instruction for their students. Teachers said that having the paired selections and *Reality Central*

selections and being able to choose which one would be better for each class, was extremely helpful.

Paired Selections

Teachers often noted that they really liked the idea of having paired selections to choose from in each unit—specifically one that was more accessible for lower-level readers. While the Diagnostic test and lexile information for each paired selection was intended as a tool for assigning the correct level for students, most teachers admitted that their choice of reading selection was often based on what their students would find most appealing. Teachers also reported that they would not necessarily choose only one selection, but instead read both of the first set of paired selections and perhaps skip an entire set later on in the unit. So, although the program was not always used as intended, the teachers gave high marks for the paired selections in concept and in practice.

> "I liked the fact that you have the two different [selections], the A and the B. I like that because it gives us a choice of which one to read..,"
>
> "It was nice to have one that was little bit more accessible and one that was a little bit more advanced…And that we could decide which one was better for each class."
>
> "There was always advice for things to do to make it more accessible…"

Technology Components

A primary component of the *Prentice Hall Literature* (2010) program was the inclusion of multiple technological elements, which were used by teachers and students alike. The program included videos, audio selections, interactive games, and CD-ROMs that could be used in the classroom in addition to a complete accompanying website. Teachers commented on each element used during focus groups and interviews and in weekly logs. The following is a summary of their comments.

Classroom components

Many teachers liked having so many elements to use in the classroom such as the author videos, interactive vocabulary games, and selection audio to accompany

readings. However, almost all teachers noted that the lack of advanced technology in their classrooms prevented them from using everything that they wanted to use. One teacher commented "I didn't get a chance to use any of the web resources because [we] don't have wireless … but I did use all the technological [resources like the] DVDs, BQ tunes…" In the weekly logs, teachers commented they enjoyed the author videos, but would have liked more video resources.

> *"I enjoyed having all of the resources on the CD for the teachers."*
>
> **–A middle school teacher**
>
> *"They liked the…introduction videos"*
>
> **–A teacher**

The audio resources were also identified as a favorite by both middle school and high school teachers in multiple interviews and weekly logs. One teacher noted the audio resources, and the ability to use English subtitles on the videos, were particularly helpful for students whose first language was not English.

Website components

The website had mixed reviews. Many teachers and students appreciated having the full textbook available online and that it included the ability to search the pages. However, teachers expressed frustration from the beginning that the website logged them out extremely quickly and teachers could not assign homework using the website components because if a student did not have extremely high speed internet at their house they were unable to access the resources in a timely manner. Additionally, two teachers remarked that the website was not user-friendly and could benefit from reorganization.

> *"There's a lot of technology built into it, which was nice, it got the kids to buy into it a little bit more because they are such technologically advanced students now."*
>
> **–A high school teacher**

Nonetheless, when the technology was available, teachers gave high remarks on the website, especially the interactive vocabulary games. Many teachers simply said their students loved having the online components. Students gave their overall online experience ratings of 2.90, 2.86, and 2.62 (out of 4) for seventh, eighth, and tenth grades, respectively.

Product Satisfaction Summary

Overall, student and teacher users of the *Prentice Hall Literature* (2010) program were satisfied with the program. Treatment students rated nearly all aspects of their textbook significantly better than control students. Students particularly enjoyed the arts and graphics in the book as well as the online vocabulary games. Middle school students were much more likely to report liking both *BQ Tunes* and *Reality Central* books. Teachers reported liking the *Big Question* and academic vocabulary in particular. While teachers expressed frustration in some technology (either being logged out of the website too quickly or not being able to use technology components in their classrooms), they generally liked the technology components of the program. Teachers also liked the ability to choose one of the paired selections, however, their choice of reading selection was primarily based on students' interest and not reading level.

Section Seven: Discussion

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

Efficacy Study Key Findings

Program Components Were Well Liked and Used Often, but Not Always as Intended

It is clear that students and teachers liked using the *Prentice Hall Literature* (2010) program. We have seen numerous examples of this based on student survey responses, and both formal and informal discussions with teachers. While teachers clearly identified their favorite program components, such as having paired selections available for use, or guiding the unit with the *Big Question*, they did not always use the program as specified by the program creators. For example, most teachers reported that they appreciated having two selections to choose from in each section, especially when they could identify that one was more or less accessible for students. However, what often guided their selection choice was how engaging or interesting it might be to students as a primary concern, rather than basing the decision purely off of difficulty level. Teachers also stated that they chose a story that they were familiar with or liked instead of necessarily choosing something new.

For the *Big Question*, teachers reported liking this more than students did, and we saw references to the *Big Question* often during classroom observations. Teachers seemed to naturally use the *Big Question* in introducing a unit for the first time, or attempting to integrate it back into discussions throughout the unit, but were less interested to finish the unit *Applying the Big Question*, as the discussion throughout had seemed to have served its purpose. Another facet of the *Big Question* that was noted were differences in discussion depending on the nature of the *Big Question* and what was required of students in terms of the depth of their responses. For example, from grade eight, the *Big Question* in Unit 1 was "Is truth the same for everyone?" In a student's view, an appropriate response might simply be a "yes" or "no", which would not necessarily facilitate an active discussion among the class. This is contrasted with the *Big Question* in Unit 4 "What is the secret to reaching someone with words?" This question perhaps prompts a

much more meaningful discussion beyond a "yes" or "no" response. Therefore, it seems that the quality of discussions when integrating the *Big Question* varied with the nature of the question, something that was apparent during observations.

Resolving Differences in Students' Scores

It is necessary to resolve differences in students' reading and writing scores observed during the study. First, we saw that medium and high implementers of the *Prentice Hall Literature* (2010) program (treatment students) significantly increased their vocabulary test scores in comparison to control group students. This is contrasted with the fact that treatment and control group students increased their reading comprehension scores (and subsequently overall reading score) at the same rate from pretest to posttest. We believe a reasonable explanation for this is that the medium and high implementers used program components such as the academic vocabulary as well as the selection vocabulary and accompanying games often during the study. It is also highly likely that vocabulary was not emphasized as much in control classrooms. Therefore, it is reasonable to see that the treatment scores in vocabulary would exceed those of controls. Contrasted with this, control students were much more likely to engage in activities such as reading novels or writing research papers, which essentially had the same impact on their reading comprehension performance (and hence overall score) as the treatment group activities not related exclusively to vocabulary. In fact, one of the only items that treatment and control group students rated similarly on the posttest was "I learned how to read better from my English Language arts [*PH Literature*] textbook." It appears that students made similar attributions regarding improving their reading scores, and the objective test results supported this idea.

One finding that we cannot fully reconcile is that although treatment students reported that they did not like the Writer's Workshop component of the *Prentice Hall Literature* (2010) program, a subset of them (high implementation students) were still able to increase writing skills on the MAT8 beyond what the control students did (this was especially true of seventh grade students). It is possible that this and other program components actually helped them improve writing skills despite the fact that they did not like the major writing component in the program. This finding is supported again by student posttest ratings in which treatment students rated the following item significantly higher than control students "My

English Language arts [*Prentice Hall Literature*] textbook helped me learn how to write better than before I took this class." Therefore, despite not liking some writing components of the program, students did in fact make the positive attribution that their textbook helped them to learn how to write better, beyond what control students reported.

Teacher Influence and Enjoyment of Reading

A final noteworthy finding was related to student attitude scores and product satisfaction scores found on the student surveys. For both treatment and control students, their ratings of their teacher's influence dropped from pretest to posttest in roughly the same proportions. So, generally speaking students did not improve their attitudes about teachers or consider them more influential from fall to spring. Yet, treatment students reported that they *increased* their enjoyment of reading from pretest to posttest, while control students ratings for this item *remained the same* over this period. These findings are important in light of how much students reporting liking their language arts program. Treatment students overwhelming reported liking program components in the *Prentice Hall Literature* (2010) program than control students did, as observed in their product satisfaction ratings. Given these data, it is possible that students liked the *Prentice Hall Literature* (2010) program to the extent that it served as a direct and positive influence on their ratings of enjoyment of reading, and this was separate from anything that their teachers did to influence them.

This is also important given that teacher characteristics such as rapport did not have a significant impact on students' achievement scores, which is counter to what we have found in other studies. Therefore, the program in the current study does have the potential to impact students in a positive way given that students like the content and program components.

Study Limitations

Given that implementation level was not randomly assigned (e.g., teacher behavior and decisions about program used determined implementation level) we are less confident in our ability to state that high and medium implementation of the program caused students to have higher achievement in vocabulary, or the high implementation caused students to have higher writing scores. It is possible that better teachers were better implementers of the program, and consequently their students would have been recipients of a high-quality vocabulary or writing programs even without the use of the *Prentice Hall Literature* (2010) program. However, it is clear that those students in classrooms where more of the *Prentice Hall Literature* program was implemented did have higher achievement on the standardized vocabulary measure (GMRT subtest) and MAT8 writing test. Future efforts to examine this causal relationship could specifically manipulate this factor to see the extent to which it predicts higher student achievement beyond other teacher and student characteristics.

References

Center on Education Policy (2008). Instructional time in elementary schools: A closer look at changes for specific subjects. *Arts Education Policy Review, 109* (6), 23-27.

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

Ivey, G. & Broaddus, K. (2001). "Just plain reading": A survey of what makes students want to read in middle school classrooms. *Reading Research Quarterly, 36* (4), 350-377.

National Endowment for the Arts (2007). *To read or not to read: A question of national consequence.* Research Report #47. Office of Research and Analysis. Washington, D. C.

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

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

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

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) (2009). [Graph illustration from the 2009 Nation's Reading Report Card]. *1992, 1994, 1998, 2002, 2003, 2005, 2007, and 2009 Reading Assessments.* Retrieved from http://nationsreportcard.gov/reading_2009/nat_g8.asp?tab_id=tab1&subtab_id=Tab_1#tabsContainer

U.S. Department of Education, National Center for Education Statistics. (2010). *Digest of Education Statistics 2009* (NCES 2010-013), Chapter 2. Retrieved from http://nces.ed.gov/fastfacts/display.asp?id=147

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

Appendix A. Demographic Summary of Participants by Grade

High School (Tenth Grade)

	Response Options	Arizona Site 1 (*n* ~ 144)	California Site 2 (*n* ~ 390)	California Site 6 (*n* ~ 461)	Ohio Site 7 (*n* ~ 216)
Gender	Male	55.6%	51.5%	50.3%	58.3%
	Female	44.4%	48.5%	49.7%	41.7%
Ethnicity	Caucasian	55.1%	7.0%	12.0%	69.4%
	African-American	0.8%	17.5%	18.8%	23.4%
	Hispanic	39.8%	62.6%	63.8%	2.7%
	American Indian	3.4%	0.7%	0.5%	0.9%
	Asian	0.8%	2.3%	2.2%	0.9%
	Multiple Ethnicity/Other	–	9.9%	2.7%	2.7%
Primary Language	English	85.7%	87.3%	89.9%	99.1%
	Non-English	14.3%	12.6%	10.1%	0.1%

* Note. Site 1 = Wickenburg High School, Site 2 = Carter High School, Site 6 = Silverado High School, Site 7 = Pickerington High School

Middle School (Seventh and Eighth Grades)

State		California				Oregon	
		Site 3	Site 4	Site 4	Site 5	Site 8	Site 8
	Response Options	Eighth Grade (*n* ~ 318)	Seventh Grade (*n* ~ 293)	Eighth Grade (*n* ~ 265)	Seventh Grade (*n* ~ 334)	Seventh Grade (*n* ~ 155)	Eighth Grade (*n* ~ 153)
Gender	Male	50.6%	54.6%	55.5%	47.9%	49.0%	47.1%
	Female	49.4%	45.4%	44.8%	51.8%	51.0%	52.9%
Ethnicity	Caucasian	5.6%	7.2%	6.9%	6.8%	69.9%	61.1%
	African-American	14.3%	19.0%	20.1%	17.6%	–	3.8%
	Hispanic	74.5%	66.2%	70.9%	71.2%	11.8%	16.0%
	American Indian	–	3.1%	–	0.9%	16.9%	16.8%
	Asian	0.9%	2.1%	0.5%	–	0.7%	1.5%
	Multiple Ethnicity/Other	4.3%	2.0%	1.1%	3.7%	0.7%	0.8%
Primary Language	English	81.1%	78.0%	82.3%	75.2%	97.8%	96.9%
	Other	18.9%	22.0%	17.7%	24.8%	2.2%	3.1%

Appendix B. Teacher Training Description

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

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

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

Appendix C. Teacher Interview Protocols

Treatment Teacher Interview Protocol

1. *Navigation:* How easy or difficult was it to navigate the textbook? How about in comparison to other programs you have used in the past? Do you have any suggestions for improvement?

2. *Differentiated Instruction:* One definition of differentiated instruction is "A flexible approach to teaching in which the teacher responds to student differences in reading level, interests, and learning needs." Given this definition, how well do you feel you were able to differentiate instruction for your students using the *Prentice Hall Literature* (2010) materials? (e.g., prompts in the margin of the teacher edition text)

3. *Paired Reading Selections:* This question is related to the differentiated instruction, but we would like some specific feedback on the effectiveness of having *paired reading selections.* Did you feel that this was an effective feature of the new program?

   a. *Understanding by Design:* Please provide specific information about what you thought of the following program components. Please be as specific as possible. To what extent did each of the following engage students, motivate students, and reinforce vocabulary? a. The *Big Question*; b. *Understanding by Design* strategies

4. *Reality Central:* How did you use *Reality Central* with your students? Did you find it was an important addition to the program?

5. *Product use:* What components of the teacher's resources and ancillary materials do you prefer to use? What did you like about them? Which parts of the program did you avoid using and why?

6. *Digital Path:* How much of the *Digital Path* were you able to use in the classroom? Did students use it at home? What were the strong and weak points of the online services?

7. What was the strongest aspect of the *PH Literature* program? What was it missing?

Control Teacher Interview Protocol

1. Comment generally on your view of how your year has gone being part of the study and using your current Literature program. (Verify the control program)

2. *Navigation:* How easy or difficult was it to navigate the textbook?

3. *Differentiated Instruction:* (see question 2 above). Do you feel the program was able to meet the learning needs of all your students?

4. *Students' Understanding:* Please describe the elements of the curriculum that best contributed to students' understanding of literature. Are there particular things that you did as a teacher (assignments, activities, etc.) that you think helped contribute to their understanding?

5. *Control Textbook:* Please describe the best and worst features of the Literature program that you used this year.

6. Is there anything else you would like us to know about your classes this year?

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

Extraction Method: Principal Component Analysis

Rotation Method: Varimax Rotation with Kaiser Normalization

Three Factors Extracted

Factor 1: Student Interest and Enjoyment of Reading (Cronbach's alpha = .848)

1. I like to read about new things (.743)
2. I like hard, challenging books (.738)
3. I enjoy reading for pleasure (e.g., novels, poetry, short stories, etc.) (.726)
4. If a book is interesting, I don't care how hard it is to read (.688)
5. I like it when books make me think (.686)
6. I sometimes lose track of time when I am reading about an interesting topic (.654)
7. I usually learn about difficult things by reading (.605)
8. When the teacher discusses something interesting, I want to read more about it (.594)

Factor 2: Teacher's Influence of Learning (Cronbach's alpha = .792)

1. My English/language arts teacher explains things clearly (.809)
2. My English/language arts teacher helps me want to do my best in class (.801)
3. My teacher expects me to do well in my English/language arts class (.745)
4. During my English/language arts class it is clear what I am supposed to be doing at all times (.744)

Factor 3: Self Efficacy in Language Arts (Cronbach's alpha = .679)

1. I am just not good at writing even though I may be good at other things (reversed, .808)
2. I am just not good at reading even though I may be good at other things (reversed, .713)
3. I am a good writer (.681)
4. I am a good reader (.552)

Appendix E. Efficacy Study Implementation Guidelines (Grade Seven)

Before you begin using the book, please read the following:

- **"Leveled Reading Selections: A Key to Differentiation"** by Harvey Daniels on pages T64-66, Teachers Edition

- **"Big Questions: Teaching Literature by Design: Introducing the Big Questions"** by Grant Wiggins on pages T67-69, Teachers Edition

Pacing for a typical unit pacing plan covers six weeks; therefore, only some aspects of the unit will be required, but we encourage all teachers to complete as much of the unit as possible. The following implementation guidelines correspond to the *Unit Overview and Pacing Plan* on page 2a-2b of the Teacher's Edition.

Please note that all content in the textbook is available online at PHLitOnline.com. The online content is an optional format to deliver the contents (as opposed to reading from the textbook) although some online content is required. The following guidelines (and subsequent page numbers) correspond to *Unit 1: Fiction and Nonfiction,* but the same principles apply to all units in the book and therefore these guidelines should be applied to all units covered throughout the year, with the exception of the initial Diagnostic Test.

1. **Required: Administer the Diagnostic Test (e.g., page 5 of Unit 1 Resources)**

2. **Required: Introduce the Unit Big Question (pp. 2-3)**
   - **Required: Show *Big Question* video** (online or See It! Video Program)
   - **Required: *Big Question* Vocabulary review** (*Vocabulary Central* online or in book)
     - o Optional: Big Question Vocabulary worksheet (online or Unit Resources)
   - **Required: Introducing the *Big Question* (pp. 2-3)- Discuss it: Write What you Know;** *Discuss It: Explain What You Know; Write About It: Tell What You Think*
   - **Required: BQ Tunes for corresponding unit** (*Vocabulary Central* online)—lyrics are in Unit Resources (optional)

3. **Required: Introduce the Unit author and the Unit forms, fiction and nonfiction (pp. 4-7)**
   - Strongly Recommended: Meet the Author Video (online or See It! Video Program)

4. **Required: Teach the Model selections (pp. 8-21)**
   - **Complete section using textbook resources**

5. **Required: Teach one selection from Pairing 1- *Papa's Parrot* or *mk* (pp. 22-47)**
   - **Required: Review Lesson Pacing Guide (p. 22a-d) for *Papa's Parrot* or *mk*, then following training instructions on how to use lesson pacing guide**
   - **Required: Making Connections (p. 24/32)**
   - **Required: *Reality Central* selection**[3]
   - Optional: *Reality Central Writing Journal*
   - **Strongly Recommended: Get Connected Video for *Papa's Parrot* or *mk*** (online or See It! Video Program)
   - **Strongly Recommended: Background Video for *Papa's Parrot* or *mk*** (online or See It! Video Program)

6. **Required: Teach one selection from Pairing 2- *An American Childhood* or *The Luckiest Time of All* (pp. 48-69)**
   - **Required: Follow Lesson Pacing Guide (p. 48a-d) for *An American Childhood* or *The Luckiest Time of All***
   - **Required: Making Connections (p. 50/60)**
   - **Required: *Reality Central* selection;** Optional: *Reality Central* Writing Journal
   - **Strongly Recommended: Get Connected Video for *An American Childhood* or *The Luckiest Time of All*** (online or See It! Video Program)
   - **Strongly Recommended: Background Video for *An American Childhood* or *The Luckiest Time of All*** (online or See It! Video Program)

7. **Strongly Recommended: Complete the Test Practice: Reading (pp. 70-71)**

8. **Required: Teach Informational Texts**[4] **(pp. 72-77)**
   - **Teach entire section, be sure to address the *Big Question***

9. **Required: Teach Comparing Literary Works (pp. 78-91)**
   - **Teach entire section, be sure to address the *Big Question***

10. **Required: Have students complete the Writing Workshop (pp. 92-97)**
    - **Complete section using textbook resources**

11. **Strongly Recommended: Administer Benchmark Test 1 (Unit Resources, p. 127)**

12. **Required: Teach one selection from Pairing 3 (pp. 98-127)**

13. **Required: Teach one selection from Pairing 4 (pp. 128-151)**

14. Optional: Complete the Test Practice: Reading (pp. 152-153)

15. Optional: Teach Informational Texts (pp. 154-159)

16. Optional: Teach Comparing Literary Works (pp. 160-175)

17. Optional: Have students complete the Writing Workshop (pp. 176-183)

18. Required: Have students complete Applying the *Big Question* (pp. 184-185)
- **Complete section using textbook resources**

19. Required: Have students complete the Vocabulary Workshop (pp. 186-187)
- Complete section using textbook resources

20. Optional: Have students complete the Communications Workshop (p. 188)
- Complete section using textbook resources

21. Strongly Recommended: Complete the Test Practice: Unit 1 Review (pp. 190-195)

22. Strongly Recommended: Administer Benchmark Test 2 (Unit Resources, p. 234)

[1]NOTE: One (1) *Informational Texts* required per unit

Appendix F. Teacher Implementation for all Units

All Units	Teacher																	
	High School							Middle School										
	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15		T16	
Grade Level	10	10	10	10	10	10	10	8	8	8	8	7	7	7	7	8	7	8
Diagnostic Test*	100%	100%	–	100%	100%	–	–	100%	100%	100%	–	–	100%	100%	100%	100%	100%	100%
Introduce the Unit *Big Question*	75%	100%	100%	100%	100%	100%	66%	80%	100%	100%	66%	66%	100%	100%	100%	100%	100%	100%
Introduce Unit Author/Genre	75%	67%	100%	100%	100%	100%	33%	80%	100%	100%	33%	100%	100%	100%	100%	100%	100%	80%
Model Selections	100%	67%	100%	100%	100%	75%	66%	80%	75%	100%	33%	66%	100%	100%	100%	100%	60%	80%
Selection from Pairing 1/ Drama 1	50%	100%	75%	100%	100%	75%	66%	80%	100%	100%	100%	33%	66%	100%	100%	100%	100%	100%
Selection from Pairing 2/ Drama 2	75%	100%	50%	100%	80%	100%	100%	60%	75%	100%	33%	33%	66%	100%	100%	100%	80%	100%
Selection from Pairing 3**	75%	100%	75%	100%	100%	100%	100%	75%	100%	66%	50%	66%	66%	100%	75%	75%	75%	75%
Selection from Pairing 4**	50%	67%	75%	67%	75%	100%	100%	75%	100%	66%	50%	–	66%	50%	75%	100%	50%	75%
Benchmark Test	–	33%	50%	67%	20%	75%	–	40%	75%	–	–	33%	66%	50%	60%	80%	–	–
Test Practice: Reading	–	–	25%	67%	80%	25%	66%	–	–	33%	–	–	66%	100%	60%	80%	40%	60%
Test Practice: Unit Review	25%	–	25%	67%	20%	–	–	–	–	33%	–	–	66%	75%	40%	60%	20%	40%
Informational Texts	25%	67%	100%	67%	80%	50%	66%	40%	50%	33%	33%	–	66%	100%	40%	40%	80%	60%
Comparing Literary Works***	25%	100%	100%	67%	75%	33%	50%	–	–	66%	50%	–	33%	50%	75%	60%	75%	75%
Writing Workshop	50%	33%	25%	100%	80%	50%	66%	20%	50%	100%	66%	33%	100%	75%	20%	40%	60%	40%
Applying the Big Question	50%	33%	50%	100%	60%	75%	66%	80%	100%	33%	–	66%	33%	25%	80%	80%	80%	80%
Vocabulary Workshop	25%	–	–	100%	50%	100%	100%	60%	75%	33%	33%	–	66%	50%	60%	80%	20%	40%
Communications Workshop	–	–	25%	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–

Note: Percentages represent total percent of that section for all units covered by each teacher

* Diagnostic test only completed once; ** Selections from Pairing 3 / 4 not in Unit 5; *** Comparing Literary Works not in Unit 5

Teacher	Total # Units Covered	Unit 1	Unit 2	Unit 3	Unit 4	Unit 5	Unit 6
1	4	X	X	X	X		
2	3	X	X				X
3	4	X	X	X	X		
4	3	X	X	X			
5	5	X	X	X		X	X
6	4		X	X	X	X	
7	3		X	X		X	
8	5	X		X	X	X	X
9	4	X		X	X	X	
10	3	X	X	X			
11	3	X		X		X	
12	3	X	X		X		
13	3	X	X		X		
14	4	X	X	X	X		
15 (grade 7)	5	X	X	X	X	X	
15 (grade 8)	5	X	X	X	X	X	
16 (grade 7)	5	X	X	X	X	X	
16 (grade 8)	5	X	X	X	X	X	
Total		16	15	15	12	10	3

Appendix G. School Site Observation Summaries

Arizona High School- Site 1. There was one treatment and one control teacher at this site. The treatment teacher had four class sections and was observed for two. The treatment teacher used multiple sections of the program including the student edition textbook, and the Penguin Author video. Students worked on the Writing Workshop and Vocabulary Workshops during the fall observation and worked on *Introducing the Big Question* and Introducing the Unit Author and Unit Forms / Genre during the spring observation. The pacing of the lesson needed to be readjusted during the observation due to disruptive students.

California High School- Site 2. There were two control and two treatment teachers at this site. The first treatment teacher had four class sections while the second treatment teacher had two class sections. During observation, one treatment teachers' class utilized the *All-in-One Workbook* and reviewed worksheets as a class. There were no significant disruptions and though pacing seemed a little slow, the vast majority of students were on-task the entire class period. For the second treatment teacher's observation, students used the *Reader's Notebook* and the *Hear It! Audio Program* with a Paired Selection.

California Middle School- Site 3. This site had two treatment and two control teachers, each with multiple class sections. Each treatment teacher was observed for one class period in the fall and one class period in the spring. The first treatment teacher used the student edition textbook and the *All-in-One Workbook* during both observations. Students were generally engaged and there were no significant disruptions during either observation. The second treatment teacher used multiple sections of the program during both observations. In the fall, the teacher used the *Daily Bell Ringer Activity* and the *Penguin Author Video* in addition to discussing *Introducing the Unit Author*. In the spring, the teacher used the *Reality Central* text and a teacher-created worksheet relating to the *Reality Central* selection.

California Middle School- Site 4. There were three treatment and three control teachers at this site. Two of the three treatment teachers were observed in both fall and spring while the third was observed only in the fall. The first treatment teacher used the student edition text during both observations. In the fall, the class played a *Vocabulary Central* game and completed a small-group activity. During the spring observation, students watched

a *See It! Author video* and discussed the writing process in preparation for research papers. The students were mostly well-behaved and on-task. The second treatment teacher also used the student edition textbook during both fall and spring observations. In the fall, students worked on *Comparing Literary Works* and used the *All-in-One Workbook*. During the spring observation, students were reading the novel *The Outsiders* and completed a worksheet related to that book (non-*Prentice Hall* materials). Students completed a worksheet and reviewed the *Model Selection* in one class.

California Middle School- Site 5. There were two treatment teachers and one control teacher at this site. Both treatment teachers were observed in the fall and the spring. The first teacher used the student edition textbook and multiple other resources during both observations. In the fall, students used the *All-in-One Workbook* and worked on *Comparing Literary Works* and vocabulary words. In the spring, students watched a *See It! Author video* and discussed the difference between a research paper and other forms of writing. Students were mostly engaged throughout the lessons. The second teacher used the student edition text and the *Hear It! Audio Program* both in the fall and the spring. During the fall observation, students also used the *All-in-One Workbook* and worked on *Comparing Literary Works*. During the spring observation, students read a *Paired Selection* on poetry and completed *After You Read*.

California High School- Site 6. The two treatment teachers used many of the same teaching materials during fall and spring observations. Students were largely on-task and engaged in both classrooms during both observations. They used the student edition textbook and the *All-in-One Workbook* in addition to reading aloud and participating in class discussions.

Ohio High School- Site 7. Site 7 had two treatment teachers and one control teacher. During fall observations, the first treatment teacher used the student edition textbook but also composed five paragraph essays and referred to *Explore the Big Question* during the observation. In the spring, students utilized the textbook and *Unit Resources* along with *Paired Selection* and had a discussion about poetry. The second treatment teacher was only observed in the fall, but was observed for four separate class periods. Students used the textbook, specifically *Informational Texts* and completed an assignment that was adapted from the *Writer's Workshop*.

Oregon Middle School- Site 8. Site 8 consisted of two treatment and one control teachers, all of whom taught both seventh and eighth grades. Both treatment teachers were observed for one seventh grade class period and one eighth grade class period in the fall and the spring. During fall observations, students read aloud and used the *Reader's Notebook* in one class. The other class completed the *Writer's Workshop.* During spring observations, classes again read aloud and used *Comparing Literary Works.* For the second teacher in the fall, both classes worked on *Introducing the Big Question,* watched the *Big Question video,* and completed worksheets from the *Unit Resources.* During spring observations, students completed a grammar lesson in one class and read aloud with a class discussion in the other class.

Appendix H. Random-Intercept Models with Covariates

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

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

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

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

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

Outcome variables:

1. GMRT verbal scaled scores posttest

2. GMRT comprehension scale scores posttest

3. GMRT total scaled scores posttest

4. MAT8 total scores posttest

Student background characteristics variables:

1. Proxies for prior academic achievement
  - prior GMRT (verbal, comprehension, and total respectively) scaled scores pretest
  - MAT8 total scores pretest

2. Gender: female (0) and male (1)

3. Race indicators
  - African American (1)
  - Latino (1)
  - Other ethnicity (1)
  - White (reference group)

4. Primary language indicator
  - Not English (1)
  - English (reference group)

5. Mother's education: ordinal scale ranging from "less than high school" (1) to "doctoral/professional" (6).

6. Grade level indicators
  - 7th (1)
  - 8th (1)
  - 10th (reference group)

7. Pre-survey composite affective variables:
  a. Interest and Enjoyment of Reading
  b. Self-efficacy in language arts

Teacher/classroom/school characteristics variables:

1. Condition
  - Treatment Implementation-low (1)
  - Treatment Implementation-medium (1)
  - Treatment Implementation-high (1)
  - Control (reference group)

2. Teacher classroom management

3. Years of teaching experience (Number of years teaching)

Appendix I. Study Attrition

Sample attrition is defined as those students who completed pretests on any of the primary outcome measures (i.e., GMRT, MAT8, and a Student Survey), yet did not complete a posttest on any of these measures. There were 1,907 participating students that completed at least one pretest and posttest. Of the 1,907 students, 1,152 completed all pretests and posttest. The original pretest sample (students that completed at least one pretest measure) included 2,397 students (670 seventh grade; 651 eighth grade; 1,076 tenth grade).

The difference between the original sample of students that took any or all of the pretests and the final number of students is 490 (119 seventh grade students; 91 eighth grade students; 280 tenth grade students). Of the 490 students, 349 were not included on the rosters for the second semester and most likely moved during the first semester. The remaining 141 students were either absent during posttesting or had moved during the second semester of the 2009-10 school year.

An overall summary of the attrition data is provided in Table a. This table shows that there were no noteworthy differences when comparing the treatment and control groups in the percentage of students that did not complete a posttest after completing a pretest assessment. Overall, the treatment and control groups had a close percentage (21.6% vs. 19.3%, respectively) of students that did not complete at least one posttest after completing a pretest. After accounting for attrition, there were 934 participating students (78.4% of treatment students with a pretest)

in treatment classrooms and 973 participating students (80.7% of control students with a pretest) in control classrooms who had completed at least one matched pretest and posttest. Tables b through d show the same information as Table a for grades seven, eight, and ten. Table b shows that seventh grade was similar in the number of students that dropped out of the study. Table c showed a slightly higher percentage of control students that dropped from the study for eighth grade. Table d showed that a larger percentage of treatment students that left the study in tenth grade classrooms.

All further analyses focus only on the combination of all assessments (i.e., we have not reported individual results on the GMRT, MAT8, and the student survey). We found that these data are sufficiently represented when looking at all assessments together. To examine whether sample attrition created differences between the treatment and control groups, we compared the original sample (those students who had completed at least one pretest) to the final sample of students (those students who had completed both a pretest and posttest) across student demographic characteristics. Table e shows the key demographics of our original sample and those students from the original sample that completed a posttest. The table shows that there are no major differences between the original sample and the final sample of students: gender is split evenly between both samples, a majority of the students are Latino (46.0% and 46.9%), and most speak English (70.7% and 71.5%).

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

Assessment	Condition	Students with Complete Pretest and Posttest	Students Missing Posttest Only (Attrition)	Total for Each Assessment
GMRT	Treatment	849 (75.1%)	282 (24.9%)	2,245
GMRT	Control	865 (77.6%)	249 (22.4%)	2,245
MAT8	Treatment	823 (76.5%)	253 (23.5%)	2,038
MAT8	Control	713 (74.1%)	249 (25.9%)	2,038
Student Survey	Treatment	697 (73.2%)	255 (26.8%)	2,002
Student Survey	Control	761 (72.5%)	289 (27.5%)	2,002
At Least One Assessment Complete	Treatment	934 (78.4%)	257 (21.6%)	2,397
At Least One Assessment Complete	Control	973 (80.7%)	233 (19.3%)	2,397

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

Assessment	Condition	Students with Complete Pretest and Posttest	Students Missing Posttest Only (Attrition)	Total for Each Assessment
GMRT	Treatment	263 (81.4%)	60 (18.6%)	628
	Control	247 (81.0%)	58 (19.0%)	
MAT8	Treatment	247 (76.0%)	78 (24.0%)	633
	Control	239 (77.6%)	69 (22.4%)	
Student Survey	Treatment	224 (79.2%)	59 (20.8%)	556
	Control	213 (78.0%)	60 (22.0%)	
At Least One Assessment Complete	Treatment	284 (82.6%)	60 (17.4%)	670
	Control	267 (81.9%)	59 (18.1%)	

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

Assessment	Condition	Students with Complete Pretest and Posttest	Students Missing Posttest Only (Attrition)	Total for Each Assessment
GMRT	Treatment	289 (84.3%)	54 (15.7%)	617
	Control	204 (74.5%)	70 (25.5%)	
MAT8	Treatment	278 (89.1%)	34 (10.9%)	536
	Control	171 (76.3%)	53 (23.7%)	
Student Survey	Treatment	256 (81.0%)	60 (19.0%)	559
	Control	179 (73.7%)	64 (26.3%)	
At Least One Assessment Complete	Treatment	318 (87.6%)	45 (12.4%)	651
	Control	242 (84.0%)	46 (16.0%)	

Differential Attrition

Almost any experimental study has participant attrition, particularly in applied research settings (i.e., schools) where students may leave before the year is over due to circumstances outside of the control of the school, teacher, or researchers. What is important to determine, however, is whether there was differential attrition such that students in one group (treatment or control) were more likely to exit the study in comparison to the other group before completing posttest measures; two sets of analyses were conducted to test this. The first set of analyses used demographic characteristics to examine the extent to which students that completed both a pretest and posttest differ from students that completed only a pretest. The second set of analyses sought to determine of the students who dropped out of the study, the treatment and control students differ in their achievement scores on the GMRT and MAT8 assessments. This second set of analyses is discussed in Section Five under the Attrition section.

Based on our sample attrition analysis, there were 490 students (119 seventh grade, 91 eighth grade, and 280 tenth grade) who took at least one pretest assessment but did not

take a posttest. Tables f, g, and h explore the demographic characteristics of these students at each grade level to see whether there was any systematic differential attrition between the groups. Despite small differences between demographic characteristics of the students that left the study, the overall sample of students that remained in the study and were available for analysis of outcome data was relatively similar across the treatment and control groups.

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

Assessment	Condition	Students with Complete Pretest and Posttest	Students Missing Posttest Only (Attrition)	Total for Each Assessment
GMRT	Treatment	297 (63.9%)	168 (36.1%)	1,000
GMRT	Control	414 (77.4%)	121 (22.6%)	1,000
MAT8	Treatment	298 (67.9%)	141 (32.1%)	869
MAT8	Control	303 (70.5%)	127 (29.5%)	869
Student Survey	Treatment	217 (61.5%)	136 (38.5%)	887
Student Survey	Control	369 (69.1%)	165 (30.9%)	887
At Least One Assessment Complete	Treatment	332 (68.6%)	152 (31.4%)	1,076
At Least One Assessment Complete	Control	464 (78.4%)	128 (21.6%)	1,076

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

Demographic Characteristics (Percent %)		Original Pretest Sample n = 2,397	Sample with at least one Pretest and Posttest n = 1,907
Gender	Male	51.5	50.4
Gender	Female	48.4	49.4
Ethnicity	Caucasian	18.2	19.9
Ethnicity	Latino	46.0	46.9
Ethnicity	Multi-ethnic/Other	2.9	2.5
Ethnicity	African-American	12.6	11.7
Ethnicity	Asian	1.1	1.3
Primary Language	English	70.7	71.5
Primary Language	Other	11.2	11.9
Primary Language	Unknown	18.0	16.6

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

Table f. Seventh Grade Students with Complete Pretest and Posttest vs. Complete Pretest Only

Demographic Characteristics (Percent %)		Pretest and Posttest Complete n = 551		Complete Pretest Only n = 119	
		Treatment n = 284	Control n = 267	Treatment n = 60	Control n = 59
Gender	Male	52.1	46.1	43.3	61.0
	Female	47.9	53.9	56.7	39.0
Ethnicity	Caucasian	23.2	16.9	11.7	10.2
	Latino	41.9	49.4	31.7	55.9
	Multi-ethnic/Other	1.4	0.4	10.0	3.4
	African-American	10.9	12.4	10.0	10.2
	Asian	0.4	1.1	1.7	0
Primary Language	English	69.0	65.9	61.7	66.1
	Other	12.7	18.4	8.3	16.9
	Unknown	18.3	15.7	30.0	16.9

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

Table g. Eighth Grade Students with Complete Pretest and Posttest vs. Complete Pretest Only

Demographic Characteristics (Percent %)		Pretest and Posttest Complete n = 560		Complete Pretest Only n = 91	
		Treatment n = 318	Control n = 242	Treatment n = 45	Control n = 46
Gender	Male	47.5	53.7	51.1	69.6
	Female	52.2	46.3	48.9	30.4
Ethnicity	Caucasian	19.8	14.9	6.7	8.7
	Latino	50.9	50.8	35.6	56.5
	Multi-ethnic/Other	1.6	2.0	4.4	2.2
	African-American	10.1	13.2	17.8	8.7
	Asian	1.3	0.4	0	0
Primary Language	English	72.0	71.5	68.9	58.7
	Other	13.5	9.5	6.7	21.7
	Unknown	14.5	19.0	24.4	19.6

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

Table h. Tenth Grade Students with Complete Pretest and Posttest vs. Complete Pretest Only

Demographic Characteristics (Percent %)		Pretest and Posttest Complete *n* = 796		Complete Pretest Only *n* = 280	
		Treatment *n* = 332	Control *n* = 464	Treatment *n* = 152	Control *n* = 128
Gender	Male	49.1	53.0	59.9	50.8
	Female	50.6	46.8	40.1	49.2
Ethnicity	Caucasian	19.3	22.8	11.2	15.6
	Latino	41.3	47.8	30.3	53.1
	Multi-ethnic/Other	2.4	5.6	2.0	5.2
	African-American	12.0	12.1	19.7	18.0
	Asian	1.8	1.9	1.3	0
Primary Language	English	68.1	78.4	61.8	80.5
	Other	7.5	11.0	1.3	9.4
	Unknown	24.4	10.6	36.9	10.2

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

Appendix J. Student Ratings of Textbook Elements

Survey Item	Mean Rating (SD) (*n* ~ 1316)
The Big Question	2.35 (.806)
Meet the Author	2.57 (.754)
Model Selection	2.53 (.771)
Reading Selections (e.g. short stories, essays, poems)	2.96 (.788)
Writing Workshop	2.30 (.821)
Informational Texts	2.61 (.791)
Vocabulary Workshop	2.71 (.850)
Communication Workshop	2.52 (.834)
The pictures and artwork	3.30 (.738)
Reality Central book	2.82 (.882)

Note. Survey responses were provided on a scale of one to four

(One= Strongly Dislike, Four = Strongly Like)

Appendix K. Online and Electronic Component Satisfaction Ratings

Survey Item	Seventh Grade Mean Rating (SD)	Eighth Grade Mean Rating (SD)	Tenth Grade Mean Rating (SD)
The Big Question Video	2.81 (.823)	2.84 (.806)	2.68 (.778)
BQ Tunes	2.69 (.872)	2.80 (.873)	2.45 (.862)
Illustrated Vocabulary Words	2.64 (.854)	2.75 (.825)	2.90 (.739)
Interactive Vocabulary Games	3.09 (.790)	2.97 (.791)	2.94 (.776)
Online Worksheets	2.57 (1.01)	2.48 (.869)	2.40 (.925)
Author Videos	2.58 (.908)	2.66 (.878)	2.69 (.816)
Get Connected Videos	2.81 (.854)	2.74 (.866)	2.68 (.795)
Interactive Journals	2.25 (.910)	2.50 (.943)	2.39 (.831)
Vocabulary Flashcards	2.50 (1.07)	2.84 (.883)	2.74 (.933)
Background Video	2.96 (.883)	2.86 (.855)	2.80 (.781)
Your Overall Online Experience	2.90 (.956)	2.86 (.981)	2.62 (.900)

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

Notes