

# Final Evaluation Report AGS Globe's AMP Reading System Efficacy Study

Submitted To:

Marcy Baughman  
Pearson Education  
Director of Academic Research  
501 Boylston Street, Suite 900  
Boston, MA 02116

Submitted By:

Mid-continent Research for Education and Learning  
4601 DTC Blvd., Suite 500  
Denver, CO 80237



## EXECUTIVE SUMMARY

Proficiency in literacy is not only foundational for ensuring academic successes but also for ensuring that students are prepared for life. While such goals have long been recognized, over the past decade interest in ensuring a literate populace has heightened. This interest has certainly been punctuated by the emergence of Reading First and the No Child Left Behind Act (Biancarosa & Snow, 2004). Further, there is growing interest in addressing the literacy needs of adolescent and adolescent readers, a time during which adolescents reading becomes increasingly sophisticated and during which the groundwork for lifelong reading habits are laid. The importance of supporting (and improving) adolescent reading has resulted in the establishment of funding opportunities, such as the Striving Readers discretionary grant program of 2005, aimed at raising reading levels for middle and high school students who are reading below their grade level (see, for instance: [http://www.reading.org/downloads/positions/ps1052\\_supporting.pdf](http://www.reading.org/downloads/positions/ps1052_supporting.pdf) and <http://www.ed.gov/programs/strivingreaders/index.html>). Recognizing the significance of raising the reading skills of middle school children, Pearson AGS Globe recently developed materials aimed at improving adolescent readers' skills and abilities.

AGS Globe strongly believes that its products must demonstrate proven effectiveness in increasing student learning. Therefore, Pearson contracted with Mid-continent Research for Education and Learning (McREL), an external, independent, and nationally recognized leader in educational research and evaluation ([www.mcrel.org](http://www.mcrel.org)), to conduct an independent efficacy study of its reading curriculum materials aimed at improving the reading skills among struggling middle school students. The program being assessed, AMP Reading System (AMP Reading) was intended to help students attain critical reading strategies. Broadly speaking, this study of AMP Reading was intended to assess teachers' implementation the materials and measure their effects on student achievement.

The evaluation study was designed as an experimental, randomized-control trial to measure the impact of these materials on student learning in reading. Specifically, the study sought to address the following key evaluation questions:

- (1) Do teachers implement the AMP Reading curriculum appropriately?
- (2) Do students in treatment groups (using AMP Reading) demonstrate significant learning gains in reading during the study period?
- (3) How does the reading performance of students in treatment groups (using AMP Reading) compare to that of students in control groups (not using AMP Reading)?

Both qualitative and quantitative methods were used to address these questions, but specific methods of data collection were used for each question. For example, to understand how teachers implemented the curriculum, participating teachers completed an implementation log, and to assess student gains, students completed a standardized assessment (the *Gates-MacGinitie Reading Test-4, Level 6 [GMRT-4]*).

Sixth-, seventh-, and eighth-grade classrooms in 16 schools implemented AMP Reading during the 2005-2006 school year. The implementation logs revealed that participating teachers varied in the

extent to which they implemented the AMP Reading program, with some teachers considered “higher” implementers than others. Classroom observations were utilized to complement classroom implementation data collected from teacher logs. Observations revealed that the majority of students in classrooms were engaged in AMP Reading lessons and that the majority of AMP Reading teachers followed the program guidelines (although no teachers completed all lessons and many worked at a slower pace than suggested in Teacher’s Guides).

In terms of student achievement, the vocabulary and comprehension subtests and total scores of the *GMRT-4* were used to assess students’ reading achievement at the beginning and end of teachers’ implementation of the AMP Reading program. The following are key findings from the analyses:

- Comparing all students with available pretest and posttest total Extended Scale Scores (ESSs) on the *GMRT-4* revealed significant pretest to posttest gains for the total score and for both the vocabulary and comprehension subtest scores, with the larger gains occurring in the vocabulary subtest.
- Student performance from pretest to posttest on the *GMRT-4* for students in the AMP Reading classrooms as compared to students in the control condition revealed no significant differences on total or subtest score gains, indicating that there were no differences on test scores between those students who used the AMP materials and those who did not. However, the students in the control condition were initially significantly higher-achieving on the *GMRT-4* than were the AMP Reading treatment group, as indicated by their pretest scores. The students using AMP Reading improved at the same rate as the control students. Therefore, the gap between the initially lower-achieving student using AMP Reading and the initially higher-achieving control group did not grow larger.
- Students in the sixth grade made significantly greater progress during the study period than did students in other grades.
- In analyses intended to measure the effect of level of implementation on score gains in treatment classrooms (defined as the proportion of AMP Reading lessons completed by teachers<sup>1</sup>), implementation level was found to account for most of the change in total *GMRT-4* total scores from the pretest to the posttest. This was also true for the vocabulary subtest. On the comprehension subtest, when the influence of implementation was held constant the time variable still had a significant effect. However, the interaction of implementation level and time was also significant. Therefore, students still improve more in comprehension as teachers implement more lessons of AMP Reading, but gains in comprehension were less dependent on implementation level.
- Ancillary descriptive analyses of student outcome by AMP Reading units completed were also conducted. Based on estimated corresponding grade levels at PR50, researchers calculated the average grade level advanced over the course of the study; these ranged between -.5 and 2.

---

<sup>1</sup> To explore the effects of AMP Reading implementation on changes in student test scores, researchers developed an implementation variable with a range of 0-10 to describe what proportion of the lessons had been completed in each treatment classroom.

- Finally, because several of the implementation sites lent themselves to follow-up analyses, McREL researchers examined an additional question. Researchers determined that smaller class sizes make a difference (that implementing reading in classrooms with low student-to-teacher ratios may result in higher student gains).

Given the importance of reading both in terms of student academic success and lifelong success, what is perhaps most important for this inquiry into the effects of the AMP Reading materials will be determining if there are distal outcomes associated with using AMP Reading—for instance, are those students who were poor and reluctant readers at the beginning of the study more interested in and eager to engage in reading at the end? Comparisons of standardized test scores reveal student gains over the course of the intervention (the school year in this case); however, these should be accompanied with additional considerations around potential, additional, and perhaps long-range benefits of participating.

Anecdotally, teachers in AMP Reading reported that students in their classrooms who are typically “underperformers” benefited more from AMP Reading than their students who were typically higher achievers. Additional analyses were conducted to examine whether teachers’ perceptions of the differences for these students revealed that “underperformers” did, indeed, fare better in the program than their higher performing counterparts. Management issues in AMP Reading classrooms were sometimes a result of teachers being unable to keep children on the tasks in their AMP Reading Student Guides rather than reading ahead in their AMP Reading trade books—student interest in the AMP Reading materials was very high. Reports from site coordinators and teachers suggested that students enjoyed the materials and seemed to be more engaged, positive signs that these adolescent readers may have become more interested in reading.

## TABLE OF CONTENTS

Executive Summary .....	ii
Introduction.....	1
Evaluation Approach and Design .....	1
Method .....	2
Procedures.....	2
Training and Study Orientation .....	2
Incentives.....	2
Settings.....	2
Participants.....	3
Teacher Demographics .....	3
Student Characteristics.....	4
Results.....	5
Teacher Implementation .....	5
Implementation Logs .....	5
Observations .....	12
Unit Summaries .....	13
Student Performance Results .....	17
Performance in the Classrooms .....	17
Performance between AMP Reading and Control Classrooms .....	21
Performance Within and Between Grades.....	24
Performance within the Treatment Group with an Implementation Covariate.....	25
Performance By Units Completed .....	26
Additional Analysis .....	30
Conclusions.....	32
References.....	34
Appendix A: Average Ratings Unit 1 – 6 Summaries.....	35
Appendix B: Total Score Extended Scale Scores As Corresponding to Median Achievement Level .....	41
Appendix C: Effect Size Calculation Methods.....	42

## INDEX OF TABLES

Table 1. Number of Participating Teachers X Condition .....	3
Table 2. Teacher Education (Highest Degree).....	4
Table 3. Teacher Experience.....	4
Table 4. Number of Students by Condition and Attrition Rate .....	4
Table 5. Ethnicity of Participating Students (Frequencies) .....	4
Table 6. Gender of Students in Classrooms (Frequencies).....	5
Table 7. Perceptions of Amount of Material by Grade.....	6
Table 8. Perceptions of Pacing by Grade.....	7
Table 9: Reported Reading Instructional Time by Grade for AMP and Control Teachers .....	7
Table 10: AMP Teacher's Reported Preparation Time .....	8
Table 11. Proportion of Reading Time Spent Using AMP Reading by Grade.....	8
Table 12. Reasons For Not Implementing AMP Reading .....	9
Table 13. Teacher Responses to Unit Summaries .....	13
Table 14. Teacher Perceptions of Implementation Difficulty By Unit.....	14
Table 15. Average Item Ratings for Individual Unit Summaries, All Units.....	15
Table 16. AMP Reading Teachers' Perceived Barriers to Implementation.....	16
Table 17. Descriptive Statistics for ESSs .....	18
Table 18. Paired Sample <i>t</i> -Tests for AMP Students in Reading Classrooms .....	18
Table 19: Descriptive Statistics for ESSs by Performance Level.....	19
Table 20. Means and Standard Deviations for ESS Vocabulary Scores.....	21
Table 21. Means and Standard Deviations for ESS Comprehension Scores.....	23
Table 22. Means and Standard Deviations for ESS Total Scores.....	24
Table 23: Means for ESS Total Scores by Grade and by Units Completed, Treatment Schools Only (estimation of corresponding grade levels based on ESS at PR50 and estimation of grade levels advanced over one year shown).....	27
Table 24: Means for ESS Total Scores by Grade and by Units Completed, Treatment Schools Only (estimation of Percentile Ranks based on ESS-Total scores over one year) .....	28
Table 25: Means for ESS Total Scores by Grade and by Units Completed, Treatment Schools Only (estimation of Grade Equivalents based on ESS-Total scores over one year).....	29
Table 24. Means and Standard Deviations for ESS Total Scores by Student-Teacher Ratio Classrooms.....	30

## INDEX OF FIGURES

Figure 1: ESS Scores by Performance on Pretest, 6 <sup>th</sup> Grade .....	20
Figure 2: ESS Scores by Performance on Pretest, 7th Grade .....	20
Figure 3: ESS Scores by Performance on Pretest, 8th Grade .....	21
Figure 4. Vocabulary ESS Scores by Condition and Time.....	22
Figure 5. Comprehension ESS Scores by Condition and Time.....	23
Figure 6. Total ESS Scores by Condition and Time.....	24
Figure 7. Differences between Grades on ESS Total Scores.....	25
Figure 8. ESS Total Scores by Class Size.....	31

## INTRODUCTION

The Pearson AGS Globe strongly believes that its products must demonstrate proven effectiveness in increasing student learning. Therefore, Pearson contracted with Mid-continent Research for Education and Learning (McREL), an external, independent, and nationally recognized leader in educational research and evaluation ([www.mcrel.org](http://www.mcrel.org)), to conduct an independent pilot study of its reading curriculum materials aimed at improving the reading skills among struggling middle school students. The program being assessed, *AMP Reading System* (AMP Reading) is intended to help struggling middle school students attain critical reading skills. Broadly speaking, this study of AMP Reading was intended to assess teachers' implementation of AMP Reading and measure the effects of the materials on student achievement.

*AMP Reading* is a 32-week reading intervention program for middle school students. It specifically targets classes in which all students are reading two to three levels below their current grade. Therefore, it is intended as a whole-class intervention and includes options for summer school and after-school programs. According to the Teacher's Edition, AMP Reading is based on findings of the National Reading Panel report. The panel recommended that:

- Teachers should teach comprehension by focusing explicitly on a small number of skills and include many opportunities to practice them. Students should read text at a level they can understand.
- Vocabulary instruction should expose students to high-value words multiple times in a variety of contexts.
- Fluency should be taught directly with fluency cues, and practiced frequently.

*AMP Reading* is aimed at meeting these goals.

## EVALUATION APPROACH AND DESIGN

The evaluation study was designed as an experimental, randomized control trial<sup>2</sup>. The study sought to address the following evaluation questions:

- (1) Do teachers implement the AMP Reading curriculum appropriately?
- (2) Do students in treatment groups (using AMP Reading) demonstrate significant learning gains in reading during the study period?
- (3) How does the reading performance of students in treatment groups (using AMP Reading) compare to that of students in control groups (not using AMP Reading)?

---

<sup>2</sup> Randomized control trial designs provide a “means of comparing the yields of different treatments in a manner that rule[s] out most alternative explanations” [Cook, T. D. & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Boston: Houghton Mifflin.]. In other words, the process of randomization *theoretically* ensures that participants in treatment and control conditions are similar prior to implementation of any intervention; therefore, any effects of the intervention should be due to the intervention itself, rather than other factors (e.g., students' educational background, development, etc.).

Both qualitative and quantitative data were collected to address these questions. For instance, to understand how teachers implement the curriculum, participating teachers completed an implementation log and were observed twice during the school year. To assess student gains, students completed a standardized assessment (the *Gates-MacGinitie Reading Test-4, Level 6 [GMRT-4]*) both prior to the start of the study and at the close of the study.

## METHOD

This study was intended to examine the effect that participating in the AMP Reading program had on middle-school students' reading abilities. As previously noted, materials were implemented during the 2005-2006 school year.

### *Procedures*

#### **Training and Study Orientation**

Participating teachers and site coordinators were provided an on-site, hour-long overview of the research study. In addition, teachers assigned to the treatment condition were provided training on the use of AMP Reading materials. These trainings occurred between August and October, 2005. Following each training session, participants were asked to provide McREL with feedback on their perceptions of the AMP Reading training session<sup>3</sup>.

As is often true with new interventions, researchers and publishers anticipate that implementing teachers will have questions or raise concerns based on how the program appears when practiced. In early October, site coordinators were contacted to determine what additional, follow-up training would be needed for the participating sites. Based on site coordinator feedback and information gleaned from teacher Implementation Logs (described below), a follow-up training was organized and provided at all sites<sup>4</sup>.

#### **Incentives**

As an incentive for participation, treatment teachers received all of the AMP Reading materials and training free of charge (including Teacher's Editions for their assigned grade levels and all student materials). Control teachers received materials in the fall of 2007. In addition, treatment and control teachers as well as site coordinators each received a stipend totaling \$250 for their participation.

#### **Settings**

Sites were recruited by AGS Globe, PRES Associates, and McREL to participate in the study. Teachers from participating schools (n = 16) were randomly assigned to either utilize the AMP materials or to utilize their existing materials for the school year 2005 – 2006.

---

<sup>3</sup> This information was submitted to Pearson AGS Globe as part of the Efficacy Study Interim Report.

<sup>4</sup> Feedback gleaned from Implementation Logs revealed that AMP Reading teachers appreciated the follow-up training and were thankful that Pearson and McREL had attended to their concerns surrounding implementation of the intervention.

## Participants

Fifty-nine sixth, seventh and eighth grade teachers across eight districts took part in this randomized control trial. Sites were located in Midwestern, Appalachian, Southeastern and Northeastern United States.

Prior to the training session, all teachers from participating sites were randomly assigned to implement AMP Reading in their classrooms or to serve as a control classroom (utilizing their existing materials for teaching reading). Table 1 provides a breakdown of teachers by condition (treatment or control). For one site, AMP Reading implementation occurred over the course of the fall semester. Student data for analyses reported herein do not include this site, unless otherwise noted<sup>5</sup>. For another site, the student-teacher-ratio was quite low, with teachers assigned approximately five students to teach using the AMP Reading program. Although student analyses include data from this site, additional analyses were conducted to examine the effect of a low student-to-teacher ratio. In a final site, two very low performing 9<sup>th</sup> grade classrooms (one in the AMP Reading treatment condition and one in the control condition) were allowed to participate. Data from these students are included in the student analyses.

**Table 1. Number of Participating Teachers X Condition**

Condition	N	%	6 <sup>th</sup> Grade	7 <sup>th</sup> Grade	8 <sup>th</sup> Grade*
			n	n	N
Treatment	31	53%	10	11	10
Control	28	47%	8	9	11
Total	59		18	20	21

\*Includes students from two 9<sup>th</sup> grade classrooms allowed to participate in the study given very low test scores.

### *Teacher Demographics*

During their initial training on McREL's research (and, for teachers randomly assigned to the AMP Reading program, their introduction to AMP Reading), teachers were asked to provide information on their teaching background and experience. The following series of tables (Tables 2 – 3) present data on participating teachers. As shown, participating teachers from AMP Reading and control classrooms were very similar in educational background and teaching experience (none of the differences between AMP Reading teachers and control teachers were statistically significant). Collectively, teachers averaged 13.33 years of experience; the average number of years teaching among treatment teachers was 13.61 years whereas average number of years teaching among control teachers was 13.03. Although some teachers in both the treatment and control groups held reading certificates, in neither the treatment nor control group were *all* teachers charged with teaching reading for the year certified in reading.

---

<sup>5</sup> Later analyses refer to this classroom as an Abbreviated Implementation site.

**Table 2. Teacher Education (Highest Degree)**

Highest Degree	All Teachers (N = 60)		AMP Reading Teachers (n = 31)		Control Teachers (n = 29)	
	n	%	n	%	n	%
BA/BS	20	33.3	11	35.5	9	31.0
MA/MS/MEd	33	55.0	14	45.2	19	65.5
EdD/PhD	4	6.7	3	9.7	1	3.4
Other	3	5.0	3	9.7	--	--

**Table 3. Teacher Experience**

	All Teachers (N = 60)		AMP Reading Teachers (n = 31)		Control Teachers (n = 29)	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Years teaching	13.33	10.54	13.61	11.80	13.03	9.19
Years at current grade level	6.38	6.03	6.29	6.09	6.47	6.08
Years teaching at current school	6.53	6.73	6.48	6.50	6.57	7.09

***Student Characteristics***

Tables 4 – 6 provide demographic characteristics of students by treatment and control groups.

**Table 4. Number of Students by Condition and Attrition Rate**

Condition	Total # of students: Pretest	Total # of students: Posttest	Total # of students: Both tests	Total # Non-completers	Attrition Rate
Treatment	941	710	696	245	26%
Control	741	581	569	172	23%
Total	1682	1291	1265	417	

\* Numbers do not reflect students whose teachers were dropped from the study, nor students who did not have total scores available on the *GMRT-4* (e.g., students who took the whole pretest but only the vocabulary posttest)

**Table 5. Ethnicity of Participating Students (Frequencies)**

Classroom	Total # of students	African American	Hispanic / Latino	Caucasian	Asian / Pacific Islander	Native American	Multi-racial	Other	Missing
Treatment	931	355	196	150	31	5	43	36	115
Control	761	176	54	220	27	4	21	24	235

**Table 6. Gender of Students in Classrooms (Frequencies)**

<b>Classroom</b>	<b>Total # of students</b>	<b>Male</b>	<b>Female</b>	<b>Missing</b>
Treatment	931	467	449	15
Control	761	341	394	26

To ascertain whether a mortality bias due to differential attrition rates was evident in the sample, univariate analyses of variance were performed to determine whether pretest scores for students who completed the study and those who did not were significantly different<sup>6</sup>. Students who completed the study were not significantly different from those who did not complete the study in the total population,  $F(1, 1621) = 3.364, p = .067$ .

Finally, an analysis was conducted to ensure that students participating in AMP Reading were underperforming in reading. On average, the students enrolled in study classrooms (both in AMP Reading classrooms and in control classrooms) at the beginning of the study scored at least two grade levels below expected using total ESS scores compared to expected ESS scores for grade level.

## **RESULTS**

The evaluation questions for this efficacy study focused on (1) teacher implementation of the AMP Reading materials; (2) student performance in AMP Reading classrooms; and (3) student performance comparisons for participating and non-participating students.

### ***Teacher Implementation***

The first evaluation question (*Do teachers implement the AMP Reading curriculum appropriately?*) is addressed in this section of the report. Implementation data were collected and analyzed to provide feedback to AGS Globe on how the AMP Reading program was used in a classroom, the decisions participating teachers made regarding pacing, and the obstacles teachers faced in implementing AMP Reading in the classroom. These data were drawn from the *Teacher Implementation Logs* (administered to control and treatment teachers), classroom observations of AMP Reading teachers, and *Unit Summary* evaluations (administered to treatment teachers).

### **Implementation Logs**

Throughout the study, teachers completed an Implementation Log. Every Friday treatment and control teachers were sent an electronic link to an online implementation log, along with instructions to complete their online logs<sup>7</sup>. Teachers for whom electronic logs were not an option were provided paper copies of the logs which participants then completed and mailed or faxed back to McREL. Paper logs were then entered into a database populated by online Implementation Log entries.

---

<sup>6</sup> Only scores from students with both pre- and posttest *GMRT-4* total scores were utilized for this analysis.

<sup>7</sup> Logs were to be submitted by 5:00 p.m. the following Monday; when participants did not submit a weekly log, they were sent a follow-up reminder e-mail. In the event that the participant still failed to respond, site coordinators were contacted and asked to follow-up with non-respondents.

Each week Implementation Logs were reviewed by McREL staff to ensure that both treatment and control teachers completed the logs, that teacher questions communicated through the logs were addressed, and that teachers were “on pace.” For example, although Implementation Guidelines were provided to teachers and to site coordinators at the start of the study, these were sent a second time to teachers and site coordinators based on questions that were being raised in the Implementation Logs specifically regarding pacing of AMP Reading. Implementation Logs were also sent weekly to an AGS Globe Implementation Monitor (all identifying information was removed). In collaboration with the Implementation Monitor, McREL reviewed the logs and identified issues that needed to be addressed to ensure fidelity of the intervention. Midway through the study, AGS Globe elected to discontinue the services provided by the Implementation Monitor; determinations about teacher implementation were therefore made by McREL staff.

Data gleaned from the logs included information such as materials use, class time dedicated to AMP Reading lessons and activities, time spent in planning and preparation, teacher perceptions of pacing, and teacher perceptions of obstacles to implementation. The following paragraphs highlight data drawn from the Implementation Logs.

In their logs, AMP Reading teachers reflected on the amount of material to be covered in a given lesson (whether this amount was just right, too much, or not enough) and the pacing of the materials (based on the Teacher’s Guides, whether they perceived their AMP Reading instruction to be *Slowly paced and unbalanced*, *Moderately paced and unbalanced*, *Moderately paced and balanced*, *Fast paced and unbalanced*, or *Fast paced and balanced*). Table 7 and Table 8 provide teacher responses collected from all implementation logs by grade level. Of note is that the majority of teachers considered the AMP Reading materials to be *Just right* (indicating that they were able to implement all or most of their lessons during the week) and, as shown in Table 7, the majority of teachers across all grade levels considered their weekly AMP Reading lessons to be *Moderately paced and balanced*. Only a small percentage of responding teachers considered their weekly AMP Reading lessons *Slowly paced and unbalanced*. Moreover, even among those respondents who considered their AMP Reading lessons to be fast-paced, the majority found that the lessons were balanced (indicating that teachers felt they had *adequate time for all activities and could provide students with appropriate levels of emphasis where needed*).

**Table 7. Perceptions of Amount of Material by Grade**

Grade	Total N	Just Right – Teacher was able to implement all or most of the lesson this week		Too Much To Cover – Teacher didn’t have enough time to get to everything in the lesson during the week		Not Enough – Teacher wanted more material or more instructional options to use with students during the week	
		N	%	N	%	N	%
6 <sup>th</sup>	362	313	86.46%	30	8.29%	19	5.25%
7 <sup>th</sup>	291	248	85.22%	39	13.40%	4	1.37%
8 <sup>th</sup>	265	228	86.04%	25	9.43%	12	4.53%

**Table 8. Perceptions of Pacing by Grade**

Grade	Total N	Slowly paced and unbalanced		Moderately paced and unbalanced		Moderately paced and balanced		Fast paced and unbalanced		Fast paced and balanced	
		N	%	N	%	N	%	N	%	N	%
6 <sup>th</sup>	353	8	2.27%	15	4.25%	188	53.26%	11	3.12%	131	37.11%
7 <sup>th</sup>	291	5	1.72%	11	3.78%	207	71.13%	25	8.59%	43	14.78%
8 <sup>th</sup>	265	13	4.91%	6	2.26%	176	66.42%	9	3.40%	61	23.02%

AMP Reading teachers were also asked to report on the number of available minutes for teaching reading with AMP Reading during a typical day, the average amount of time students were exposed to the AMP Reading materials during the reporting week, and the amount of time teachers spent in planning and preparation for using the AMP Reading materials. Table 9 provides these data disaggregated by grade level. As shown, the average length of the reading periods ranged from approximately 42 minutes per day for 7<sup>th</sup> grade classrooms and approximately 53 minutes per day for 6<sup>th</sup> grade classrooms. The average number of minutes per day that students were exposed to AMP Reading ranged from approximately 44 minutes per day for 6<sup>th</sup> grade classrooms and approximately 36 minutes per day for 7<sup>th</sup> and 8<sup>th</sup> grade classrooms. Finally, although the standard deviations suggest rather high variability in the amount of time teachers reported spending in preparation and planning for utilizing the AMP Reading materials during the week, on average teachers spent more than 45 minutes preparing to teach with the AMP Reading materials. Medians are also provided in this table; these perhaps provide a better representation of the data reported for these items.

**Table 9: Reported Reading Instructional Time by Grade for AMP and Control Teachers**

		Grade 6			Grade 7			Grade 8		
		n	Mean	SD	n	Mean	SD	n	Mean	SD
Length of instructional period for reading (in minutes)	AMP	373	53.21	24.49	299	42.65	19.37	279	42.92	15.68
	Control	148	49.28	22.85	183	41.60	11.60	272	53.25	32.78
Amount of time (per day, in minutes) students exposed to reading instruction	AMP	373	44.22	19.01	301	36.49	18.50	276	36.38	15.51
	Control	148	48.37	22.22	183	61.27	64.36	272	52.40	52.20

**Table 10: AMP Teacher's Reported Preparation Time**

	Grade 6			Grade 7			Grade 8		
	n	Mean	SD	n	Mean	SD	n	Mean	SD
Amount of time spent in planning and preparation for the materials during the week (in minutes)	370	46.98	38.56	297	75.12	74.55	267	63.99	63.18

AMP Reading teachers also reported the proportion of their reading time they dedicated to utilizing the program. Table 11 provides the percentage of respondents selecting each time category by grade. Not surprisingly, regardless of grade level the majority of AMP Reading teachers indicated that they used the materials most of the time (between 76-100% of the reading time). Ideally, most responses to this question should fall into the 76% – 100% category; however, given that some of the school weeks included school closings or teacher inservice days, it would be expected that some of the weekly reporting would reflect less than ideal AMP Reading use. It is of interest to note that the overall percent of respondents indicating that they used AMP Reading between 76-100% of the time had increased from the mid-point examination of teacher progress, which is indicative of increased teacher utilization of the AMP Reading materials as the year progressed, and perhaps increased level of familiarity and comfort with the program.

**Table 11. Proportion of Reading Time Spent Using AMP Reading by Grade**

Percent Category	6 <sup>th</sup> Grade		7 <sup>th</sup> Grade		8 <sup>th</sup> Grade	
	n	%	n	%	n	%
0-25%	13	3.49%	10	3.39%	24	8.57%
26-50%	52	13.94%	26	8.81%	9	3.21%
51-75%	71	19.03%	24	8.14%	44	15.71%
76-100%	237	63.54%	235	79.66%	203	72.50%
Total	373		295		280	

Table 12 presents data gleaned from Teacher Implementation Logs regarding the reasons AMP Reading implementation was disrupted during the week. As shown, reasons for not implementing AMP Reading during the week varied, but included such factors as testing and assessment and, as noted above, school closings and teacher absences.

**Table 12. Reasons For Not Implementing AMP Reading**

Obstacle	N	%
Student testing and preparation	135	26.52%
Holiday	126	24.75%
Teacher absence	62	12.18%
Other	52	10.22%
Assembly	46	9.04%
Weather-related school closing	25	4.91%
Teacher professional development / Inservice	25	4.91%
Field trip	21	4.13%
Safety (fire drills, etc.)	8	1.57%
School fair	5	0.98%
Student absences	4	0.79%

Using data from Implementation Logs, we calculated each teacher’s progress in the AMP Reading program and used this as a covariate in student analyses (see student analyses).

On an ongoing basis, teacher comments recorded in the Implementation Logs were examined. Questions or concerns about the program were addressed when appropriate or related to Pearson AGS Globe for follow-up. Teacher comments about the program were also analyzed as a means of understanding teacher engagement in AMP Reading and teacher perceptions of student engagement in the program. This section provides examples of the Implementation Log comments aggregated by topic area.

Pacing was a concern for some of the teachers. Teachers made accommodations for students to ensure that AMP Reading met the needs of their students; as one teacher explained, “I paced my class to enable all to benefit from the program – it appears to take more time” (*8<sup>th</sup> grade teacher*). Other respondent’s comments express similar approaches to problems with pacing:

“I set the amount of material that I'm going to cover in my lesson plans. I know that my students are having a difficult time summarizing so I'm adjusting to meet their needs. I'm no longer following the given lesson plans” (*6<sup>th</sup> grade teacher*).

“This program is excellent, but in order for me to complete in a school year, I would have to change my teaching style set-up for my language arts class. I don't mind doing so because I believe the program works. I just didn't finish like I had expected.” (*7<sup>th</sup> grade teacher*)

Some teachers were pleased that AMP Reading seemed to complement the activities they were already doing with their students or their existing pedagogical strategies. In the words of a teacher who was able to incorporate AMP Reading into his existing approach, “AMP fits in with the summary warm-up that I have been doing since the beginning of the year. Many of the points I have addressed during warm-up have been reinforced and supported quite well with the AMP system” (*6<sup>th</sup> grade teacher*). Teachers also reported that they utilized additional strategies to maintain student interest. For example, an 8<sup>th</sup> grade teacher noted that she has places her students “...into a

review and enrichment mode with taught concepts that they enjoy. We took all the vocabulary words from this unit, and asked that they write interviews as if they were interviewing Olympic X-treme sports champions. They loved that! We look forward to finishing up with the preposition work this week, and moving into Unit Three. Finally!”

Several teachers considered the amount of student materials to be appropriate and catered to the needs of lower-level readers. For instance, an eighth grade and a sixth grade teacher lauded AMP Reading for its shorter readings and activities:

“Many of my students have a hard time when they are overwhelmed with tons of reading. I like how the program is broken into small sections of reading. It makes it easier for them to keep their focus.” (*8<sup>th</sup> grade teacher*)

“Each activity is short enough to accommodate the attention span of my 6th grade students. I have found that completion rates and interest levels have increased as a result.” (*6<sup>th</sup> grade teacher*)

In a related vein, several teachers recognized that AMP Reading is really meant to be utilized with low level readers as illustrated by the following comments:

“Some times it is repetitive and the kids get bored and tend to not try their hardest.” (*8<sup>th</sup> grade teacher*)

“I came to the realization that the repetition of this program making an impact on the low average reader. Most of them can provide the steps to summarization and are beginning to transfer this knowledge when the text is away from the AMP Reading materials.” (*6<sup>th</sup> grade teacher*)

Other teachers commented that the structure of AMP Reading – and the structure of the activities associated with it – appealed to struggling readers:

“AMP gave me an opportunity to work with students in small and large groups to revise, edit, and rewrite summary statements and discuss and model good writing techniques.” (*6<sup>th</sup> grade teacher*)

“The vocabulary review was a class favorite this week. Drawing the words from a bag and creating a sentence in pairs was really exciting for them...some even came up with themes to make the sentence relate. The reflection time for the peers to comment on fluency was challenging for them as well. They did provide feedback to one another and even though there was some debate their comments were sincere and appropriate. This also allowed everyone to get feedback in a timely manner.....love this strategy!!!” (*6<sup>th</sup> grade teacher*)

“My slower learners were able to keep up because of the uniformity of the lessons. Even though they were a little behind they were able to complete their work because they understood what to do.” (*6<sup>th</sup> grade teacher*)

“Students of many different ability levels were able to participate effectively in the lessons. It is cleverly written to allow for individual participation in workbook without unnecessarily "exposing" weaknesses to general at large.” (6<sup>th</sup> grade teacher)

A majority of teacher comments revolved around the nature of the AMP Reading materials and how materials were able to capture the interest of even the most reluctant readers:

“The students particularly enjoyed the reading material and wanted to read ahead. It was difficult trying to keep them from not doing so.”  
(7<sup>th</sup> grade teacher)

“The students are doing a good job of previewing and predicting what will be in the text. This helps them because it gets them interested in doing the reading. The reading itself is also very interesting. They really enjoyed reading about the alien invaders.” (8<sup>th</sup> grade teacher)

“This group was very excited about the reading material. They couldn't wait to begin the reading. That is really important, since the majority of the students in this class hate to read!” (8<sup>th</sup> grade teacher)

“I am having a challenge motivating my students to read outside of class. The reading material is high interest to them and it is such a relief that the in class is motivating their reading participation.” (6<sup>th</sup> grade teacher)

Another important benefit that surfaced in some teacher comments was that a transfer of learning had occurred – that students were utilizing the AMP Reading strategies in other classes. Thus, several teachers reported that AMP Reading students were transferring their learning to other content areas. For example, one teacher indicated that “The students enjoyed the Reflection, all aspects. They commented on how it had helped in other classes.” (7<sup>th</sup> grade teacher). Another teacher commented that, “The segment on note taking by utilizing headings was extremely useful in tying in the structure of the science and social studies books we currently use” (6<sup>th</sup> grade teacher).

Teachers also reported immediate and apparent student gains, particularly in vocabulary. The following is a sample of the many comments teachers made about increases in their students’ vocabulary:

“Although I have just started using AMP with this group of students, I am very pleased with the progress they are making. Most of them seem to easily be remembering the meanings of the vocabulary words, and seem to have little difficulty finding the topic of what they read.” (8<sup>th</sup> grade teacher)

“My class is all reading below grade level so they are all unique in different ways. They all benefited from the continuous vocabulary teaching.” (6<sup>th</sup> grade teacher)

On the other hand, one teacher noted that basic phonics rules were missing from the AMP Reading materials – and that some of her students would benefit from this: “Some students are still struggling with basic phonics rules. Although they can use context clues to infer meaning, they still are at a handicap in general comprehension.” (6<sup>th</sup> grade teacher)

In terms of skills, some teacher respondents specifically commented that AMP Reading, including its structure, supported strategies that were helpful for their struggling students:

“Many of our students have difficulty inferring. The activities we are currently completing help them practice this skill.” (*8<sup>th</sup> grade teacher*)

“Many of the students have a difficult time understanding prefixes. The way it was presented in the workbook was very easy to understand.” (*8<sup>th</sup> grade teacher*)

“Several of our students have gotten very good at making predictions. This was not their strong point before. They also love the reading.” (*8<sup>th</sup> grade teacher*)

“Several students have difficulty completing analogies. The way the material was presented was very effective. The students were able to understand and create their own.” (*8<sup>th</sup> grade teacher*)

However, several teacher respondents were worried that students were not given sufficient opportunity to write. Although writing is an important skill for middle school students to develop, this is not a focal point of AMP Reading. Some teacher respondents suggested that specific aspects of AMP Reading could be improved; overall teacher comments were positive and affirmed that students enjoyed AMP Reading.

## **Observations**

Researchers observed a sample of AMP Reading classrooms during fall 2005 and spring 2006. Forty classroom observations were conducted during the study year (22 observations were conducted in the fall and 18 observations were conducted in the spring). Observations were conducted at least once in all sites; teachers with multiple AMP Reading classes were observed more than once during a site observation when scheduling permitted. Two McREL evaluators conducted all site observations using an observation protocol.

Observations enabled researchers to examine teacher and student use of the AMP Reading materials as well as teacher-student and student-student interactions. Observations revealed the ways teachers and students used and did not use the AMP Reading materials. Site coordinators were asked to assist in the scheduling of AMP Reading teacher observations—to determine when teachers would be available for observation and to ensure that teachers were aware of the McREL observation schedule. When site coordinators were on site, McREL staff briefly met with them to ask about implementation and to collect program feedback. These conversations and follow-up activities were intended to facilitate teachers’ full, continued participation in the study of AMP Reading.

Although Implementation Logs revealed that teacher implementation pace – on the whole – was slow, the observations revealed that most teachers were following the AMP Reading program, with few modifications. During both fall and spring observations, the majority of students in the observed classes appeared on task. During the observations teachers primarily questioned students, presented information, gave directions, and interacted directly with students (typically walking around the classroom while students completed their AMP Student Guides or read AMP Reading

books in groups, in pairs, or alone). In some of the observed classrooms, teachers also provided students with demonstrations of the AMP Reading lessons, worked with small groups of students, or corrected student behavior. During the observations students were typically listening, asking questions, answering questions, or completing worksheets or responding to questions from their AMP Reading books. Other student activities included writing assignments and reading aloud in pairs.

## Unit Summaries

At the end of each unit, teachers in the treatment classrooms were asked to complete Unit Summaries; these assess how teachers used the materials, their perceptions of the quality and utility of the unit, and the effects of their instruction on student learning. These also provide a means of examining implementation, including teachers’ perceptions of their fidelity to implementation for a given unit and barriers to implementation of a given unit. The following paragraphs highlight findings from the unit evaluations.

A total of 66 Unit Summaries were completed by participating teachers. Table 13 provides the number of Unit Summaries completed by AMP Reading teachers per instructional unit.

**Table 13. Teacher Responses to Unit Summaries**

AMP Reading Unit	Number of Unit Summary Responses
Summarizing	26
Questioning	22
Predicting	11
Text Structure	4
Visualizing	2
Inference	1

Teachers reported how they used the materials in their classrooms—whether the AMP Reading materials were used as the core reading instruction for students (the intended use of the materials) or whether *all*, *most* or *some* of the AMP Reading materials were used to supplement the core reading instruction. Almost all teachers (n = 58; 95.08%) responding to the end-of-unit summary indicated that they used the AMP Reading materials as their core reading instruction for their students. Three of the teachers reported that they used *all* or *most* of the AMP Reading materials (for a given unit) to supplement core reading instruction.

Teachers were also asked to use a seven-point scale, ranging from a low [1] of “not at all difficult” to a high [7] of “very difficult,” to rate the ease with which they were able to adapt the AMP Reading materials to the teaching setting. Overall, respondents indicated that the materials were not at all difficult to implement (mean across all Unit Summaries = 1.67, standard deviation = 1.10). Descriptive statistics for each unit are reported in Table 14.

**Table 14. Teacher Perceptions of Implementation Difficulty By Unit**

	AMP Reading Unit					
	1 Summarizing	2 Questioning	3 Predicting	4 Text Structure	5 Visualizing	6 Inference
<b>Mean</b>	2.00	1.59	1.36	1.25	1	1
<b>SD</b>	1.41	0.91	0.67	0.5	0	n/a
<b>N</b>	26	22	11	4	2	1

Participating AMP Reading teachers were asked to use a five-point scale where [1] indicated “strongly disagree” and [5] indicated “strongly agree” to rate the quality and utility of the unit (a total of 16 questions were asked in this area), and rate their perceptions of the unit’s effect on instruction and student learning (a total of 12 questions were asked in this area). Table 15 provides aggregate descriptive data for items regarding unit quality, utility and effect on learning for Unit Summaries 1-6. Appendix A provides descriptive statistics for each unit.

As shown, responses to the Unit Summaries reflected above-average ratings for all items. Aggregating across all units, teachers gave their highest ratings to items about the ease of the lesson plans (that lesson plans were easy to follow; mean = 4.55), the instructional design of the materials (mean = 4.48), and the physical design of the materials (mean = 4.43). Respondents disagreed that the units were difficult to implement (mean = 2.70). Given these results and the high average ratings for lesson plans and design, AMP Reading teachers appeared comfortable with implementing the units of the program. Finally, responding AMP Reading teachers perceived that their students were engaged in the reading assignments and that the AMP Reading materials helped their students to understand the strategy being addressed in the AMP Reading Unit.

**Table 15. Average Item Ratings for Individual Unit Summaries, All Units**

Survey Items	Units 1 – 6	
Quality and Utility	Average Item Rating	Standard Deviation Item Rating
The layout, or physical design of the materials was well organized.	4.43	0.64
The pedagogy, or instructional design, of the materials was easy to follow.	4.48	0.62
The lesson plans were easy to follow.	4.55	0.53
Alignment with standards enhanced the materials' utility.	4.10	0.79
The standards to which the materials were aligned were relevant to my instruction planning.	4.07	0.79
The materials' alignment to standards was instrumental in helping students to meet those standards.	4.11	0.74
The assessment strategies enhanced the materials' utility.	3.92	0.89
The materials allowed me to engage in ongoing assessment of my teaching.	3.84	0.84
The materials allowed me to engage in ongoing assessment of student learning.	3.77	0.89
The materials included appropriate instructional practices for meeting the needs of diverse students.	3.98	0.88
The materials were culturally appropriate and included activities for minority students.	4.11	0.75
The materials were developmentally appropriate.	4.25	0.61
The materials were appropriate for special education students.	4.00	0.93
The materials were appropriate for ELL students.	3.71	0.87
Overall, the materials are of high quality.	4.36	0.57
Overall, the materials are very useful.	4.34	0.60
Effect on Instruction and Learning	Average Item Rating	Standard Deviation Item Rating
I felt comfortable using this unit.	4.16	0.62
The training prepared me to implement this unit.	3.81	0.73
The training provided me with the knowledge and skills I needed to implement this unit.	3.78	0.75
Implementing this unit was challenging.	2.70	1.05
Using these materials improved my understanding of how students learn to read.	3.69	0.87
Using these materials gave me a better understanding of how to facilitate learning of comprehension strategies.	3.95	0.81
Using the materials enhanced my capacity to teach reading.	3.89	0.74
I would recommend this unit to a colleague.	4.00	0.73
The materials enhanced students' interest in reading.	3.83	0.88
Students were more interested in reading as a result of using these materials.	3.73	0.93
Students were engaged in the reading assignments.	4.02	0.70
The materials improved students' abilities to use the [Unit] strategy...	4.11	0.65

Respondents were also asked to reflect on whether they thought their use of the materials had resulted in any changes to their instructional practices and to explain their responses. Fewer than half of the respondents indicated that the use of materials had changed their instructional practice, though it is important to note that a small number of teachers randomly assigned to utilize the AMP Reading materials did not consider themselves “reading teachers” and felt as if the materials had not changed their practices for teaching in other content areas. Respondents’ explanations for how the use of AMP Reading had changed their practices varied—some felt that they always followed the instructions for using materials, so there was no change in their instructional practices, or that their normal practices allowed for more flexibility for creativity and catering to individual student needs (instructional practices necessarily changed to accommodate the scripted nature of the program). Others considered the script useful for effectively teaching their students. These findings corroborate feedback reported in the Implementation Logs and observational data collected at the sites.

In terms of barriers to implementation, respondents indicated that they had limited class time for implementation or that other barriers than those included in McREL’s survey impeded their implementation. Table 16 provides descriptive data on the barriers that respondents indicated; details on the “Other” category are provided below the table. It is important to note that across all Unit Summaries nearly 30% did not think that they encountered any barriers to implementing AMP Reading.

**Table 16. AMP Reading Teachers’ Perceived Barriers to Implementation**

Barrier	Number of Responses (all Unit Summaries)	Percent (all Unit Summaries)
Limited class time available	17	23.94%
Poor alignment with the curriculum	9	12.68%
No access to necessary materials and resources	2	2.82%
Lack of training and support	2	2.82%
Not appropriate for grade level	1	1.41%
Lack of sufficient planning time	1	1.41%
Other	21	29.58%
None	18	25.35%

Other barriers to AMP Reading implementation that teachers identified included: the lack of comprehension questions following each reading from the AMP Reading materials; insufficient coverage of some state student performance indicators; lack of materials aligned with state writing standards and/or writing assignments aligned with AMP Reading units; insufficient overlap with content covered on state tests; additional assessment materials; too few activities and strategies for extremely low readers and English language learners (including decoding skills and basic word recognition strategies); the repetitive nature of the program, including the predictable layouts; lack of pronunciation keys in the student reading materials; and the problems associated with implementing a new program (the teacher learning curve).

The three sources of data employed to address the first evaluation question (*Do teachers implement the AMP Reading curriculum appropriately?*) provided sufficient evidence to suggest that, for the most part,

teachers did implement AMP Reading per the specifications. In order to ensure an appropriate amount of student time was available for active learning, all teachers adjusted their pacing. This may have been due to a number of factors, including shortened class periods or the need to meet different students' needs.

### ***Student Performance Results***

To answer the second and third evaluation questions (*Do students in treatment groups (using AMP Reading) demonstrate significant learning gains in reading during the study period? and How does the reading performance of students in treatment groups [using AMP Reading] compare to that of students in control groups (not using AMP Reading)?*), pre and posttest student data from the *GMRT-4* were analyzed. This section of the report focused on these student questions.

The *GMRT-4* subtests for vocabulary and comprehension were used to assess the students' reading achievement at the beginning and end of teachers' implementation of the AMP Reading program. This section provides student performance findings, comparing students of the teachers utilizing AMP Reading (treatment teachers) before and after the AMP Reading materials were introduced, and provides student performance findings comparing students of the teachers utilizing AMP Reading to those of the teachers not utilizing AMP Reading (the control teachers).

In order to examine the differences in student performance from pretest to posttest on the *GMRT-4* for students in the AMP Reading classrooms, researchers conducted univariate and multivariate analyses (paired sample *t*-tests and repeated measures ANOVAs). For these analyses, researchers used Extended Scale Scores (ESSs) for the vocabulary and comprehension subtests as well as the combined total scores. The ESS provides values for the *GMRT-4* across the total range of levels (Grade 1 through Grade 12) on a single, continuous scale, and is useful for comparing gains in reading over time. ESSs rank student achievement in such a way that it is related to the "entire range of achievement during the school years" (MacGinitie, MacGinitie, Maria, & Dreyer, 2000, p. 27). In addition, because ESSs are interval data they can be used to compute and compare means. Finally, the use of ESSs is advised for pre- and post-testing because they are not dependent on the time of year the test was administered (MacGinitie et. al., 2000). ESSs can also be compared against tabulated achievement levels.

#### **Performance in the Classrooms**

Table 17 provides descriptive statistics for the *GMRT-4* pre- and posttest for treatment and control groups. As shown in the table, the average total ESS scores on the pre-test for students in the treatment group was 495.49, whereas it was 522.16 for students in the control group. MacGinitie et. al. (2000) suggest that an ESS of 495 corresponds to a percentile rank of 53 expected among students in the spring of fourth grade, and 522 corresponds to a percentile rank of 51 expected among students in the spring of sixth grade (Appendix B provides the total extended scale scores corresponding to grade level expectations). Therefore, as noted in the interim report, the treatment and control groups represented different average reading levels at the beginning of the study. This table includes all available student data. Later analyses utilize only the data appropriate for the given analyses.

**Table 17. Descriptive Statistics for ESSs**

Measure		PRE			POST		
		N	Mean	SD	N	Mean	SD
GMRT-4 ESS Vocabulary Score	Treatment	912	491.46	29.07	698	505.48	31.36
	Control	755	519.84	40.37	584	536.48	39.34
GMRT-4 ESS Comprehension Score	Treatment	889	499.56	30.25	664	504.66	29.28
	Control	747	526.66	36.87	581	530.27	37.18
GMRT-4 ESS Total Score	Treatment	884	495.49	25.25	661	505.03	25.54
	Control	741	522.16	34.26	581	531.98	34.90

Comparing pre- to posttest gains in the treatment classrooms provides an indicator of student gains or losses in language arts that may have resulted from teachers’ use of materials during the course of the study<sup>8</sup>. As shown in Table 18, *t*-test analyses revealed significant gains pre- to posttest for both the vocabulary and comprehension subtests (and thus for the total score), with the larger gains occurring on the vocabulary subtest.

**Table 18. Paired Sample *t*-Tests for AMP Students in Reading Classrooms**

Measure	N	Pre-test Mean	Post-test Mean	SD	<i>t</i> value	<i>p</i> level	ES (Cohen’s <i>d</i> ) <sup>9</sup>
GMRT-4 ESS Vocabulary Score	687	492.39	505.26	23.50	14.35**	<.001	.78
GMRT-4 ESS Comprehension Score	656	501.73	504.83	24.93	3.19*	.001	.18
GMRT-4 ESS Total Score	649	496.85	504.96	16.98	12.17**	<.001	.68

\* significant at *p* < .01.  
 \*\* significant at *p* < .001

An additional analysis was conducted to examine mean differences between students in AMP Reading classrooms who were considered “higher” performers and those who were considered “lower” performers. Using tabled median values for Total GMRT-4 ESS scores by grade level, researchers categorized students in AMP Reading classrooms as either “lower” or “higher” performing. The GMRT-4 Total pretest ESS median score at two grade levels below expected was

<sup>8</sup> Many other factors affect student performance; however, as noted previously, the power of randomized control trials is that the design can help rule out pre-existing difference.

<sup>9</sup> See Appendix C.

used as a cut-point for this categorization; students with scores below this were considered “lower performing” and students with scores above this were considered “higher performing”.

In order to examine the differences in student performance from pretest to posttest on the *GMRT-4* vocabulary subtest for students in AMP Reading classrooms who were categorized “lower performing” and “higher performing”, researchers conducted repeated-measures ANOVA for grades 6, 7, and 8. Table 21 provides descriptive statistics for the *GMRT-4* pre- and posttest for treatment students by this performance level.

**Table 19: Descriptive Statistics for ESSs by Performance Level**

Grade		Performance Level	N	Mean	SD
6	<i>GMRT-4</i> ESS Total Score: Pre	“lower”	88	467.85	10.030
		“higher”	131	506.80	17.666
	<i>GMRT-4</i> ESS Total Score: Post	“lower”	88	483.49	19.125
		“higher”	131	515.67	22.796
7	<i>GMRT-4</i> ESS Total Score: Pre	“lower”	133	482.23	13.244
		“higher”	112	522.90	19.570
	<i>GMRT-4</i> ESS Total Score: Post	“lower”	133	490.98	17.939
		“higher”	112	522.84	22.632
8	<i>GMRT-4</i> ESS Total Score: Pre	“lower”	116	488.66	16.619
		“higher”	30	532.17	16.899
	<i>GMRT-4</i> ESS Total Score: Post	“lower”	116	499.84	19.440
		“higher”	30	533.17	21.346

The ANOVA for the interaction of time and performance level was significant in each of the grades. For grade 6, Wilks’s  $\Lambda = .642$ ,  $F(1, 217) = 9.215$ ,  $p < .001$ , multivariate  $\eta^2 = .669$ ; for grade 7, Wilks’s  $\Lambda = .927$ ,  $F(1, 243) = 19.173$ ,  $p < .001$ , multivariate  $\eta^2 = .967$ ; for grade 8, Wilks’s  $\Lambda = .946$ ,  $F(1, 144) = 8.192$ ,  $p < .001$ , multivariate  $\eta^2 = .600$ . Therefore, students who were “lower performing outperformed the “higher performing students in all grades. Moreover, this difference was found across the entire treatment group, regardless of grade level. Figures 1 – 3 depict grade differences.

Figure 1: ESS Scores by Performance on Pretest, 6<sup>th</sup> Grade

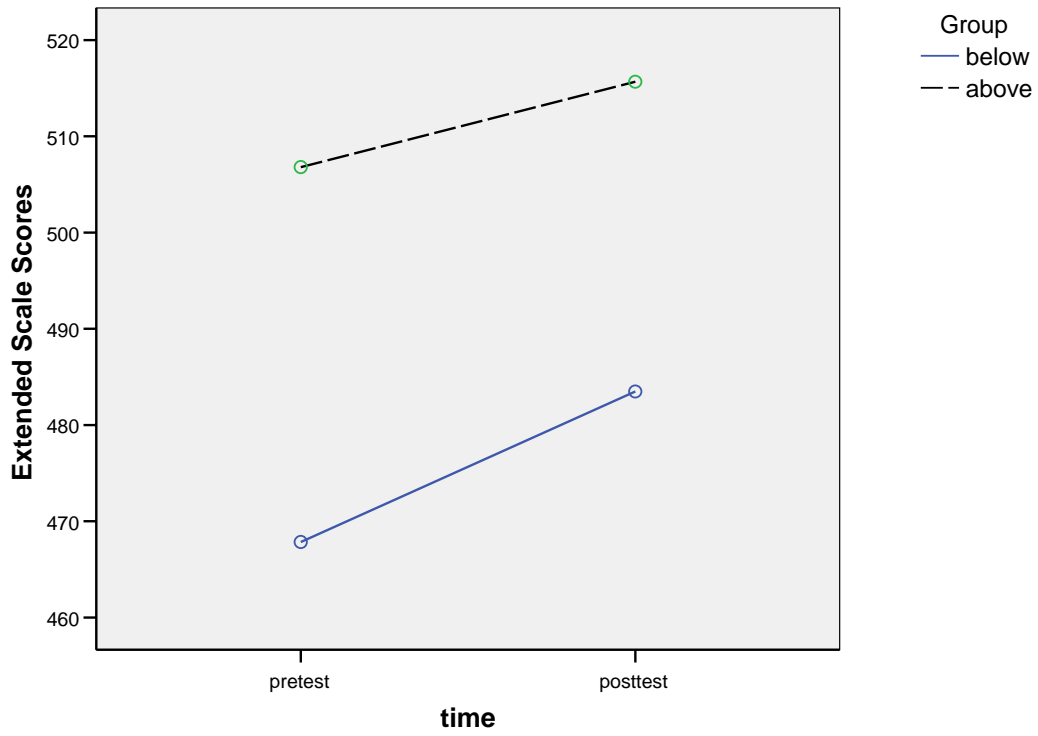


Figure 2: ESS Scores by Performance on Pretest, 7<sup>th</sup> Grade

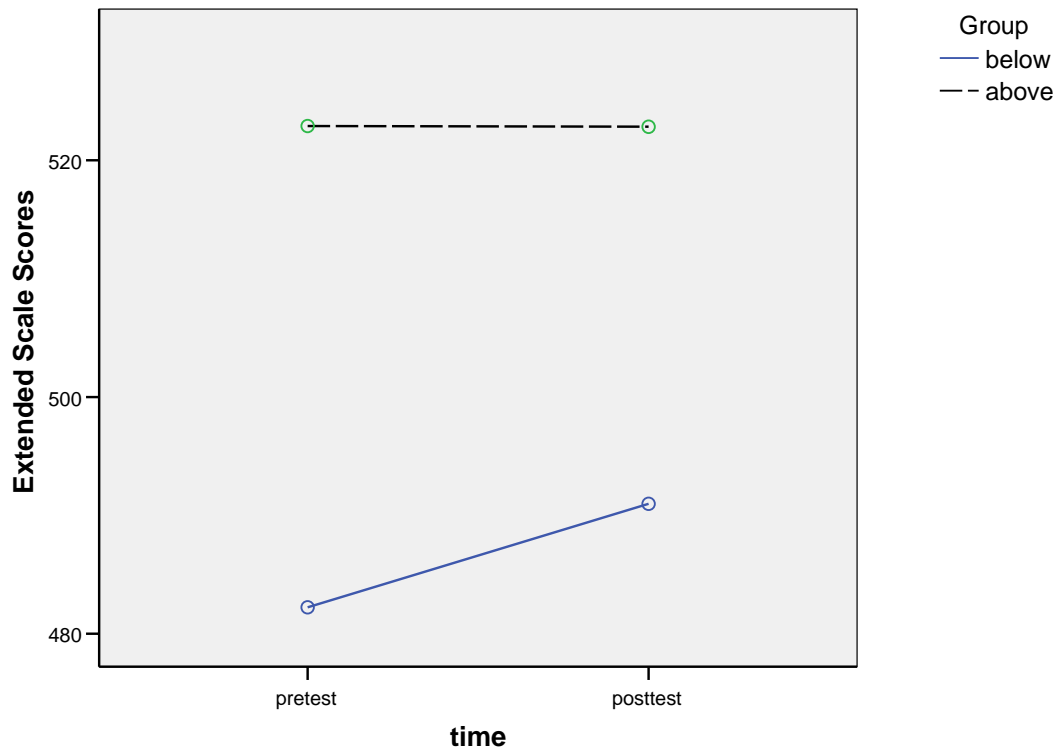
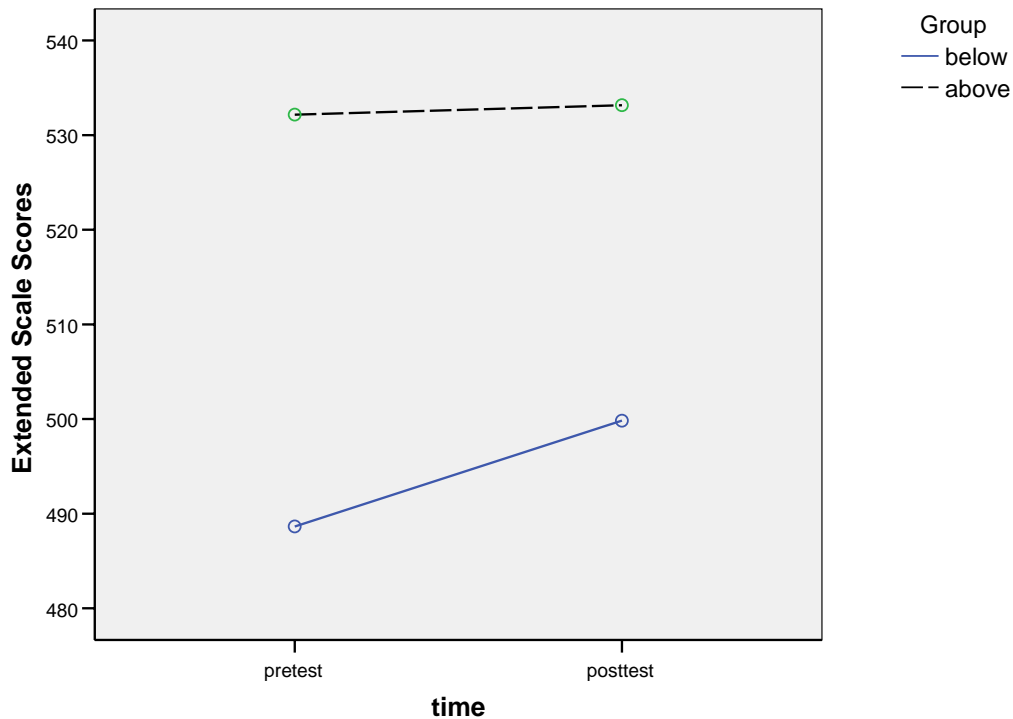


Figure 3: ESS Scores by Performance on Pretest, 8th Grade



**Performance between AMP Reading and Control Classrooms**

*Vocabulary Score.* In order to examine the differences in student performance from pretest to posttest on the *GMRT-4* vocabulary subtest for students in AMP Reading classrooms as compared to control classrooms, researchers conducted repeated-measures ANOVA. Table 20 displays the pretest and posttest performance of treatment and control students. There was a significant difference between the pretest and posttest for all students, Wilks’s  $\Lambda = .74$ ,  $F(1, 1265) = 446.23$ ,  $p < .001$ , multivariate  $\eta^2 = .26$ . However, the interaction between time and group was not significant. This indicates that statistically there were no differences on test scores between those students who used the AMP Reading

Table 20. Means and Standard Deviations for ESS Vocabulary Scores

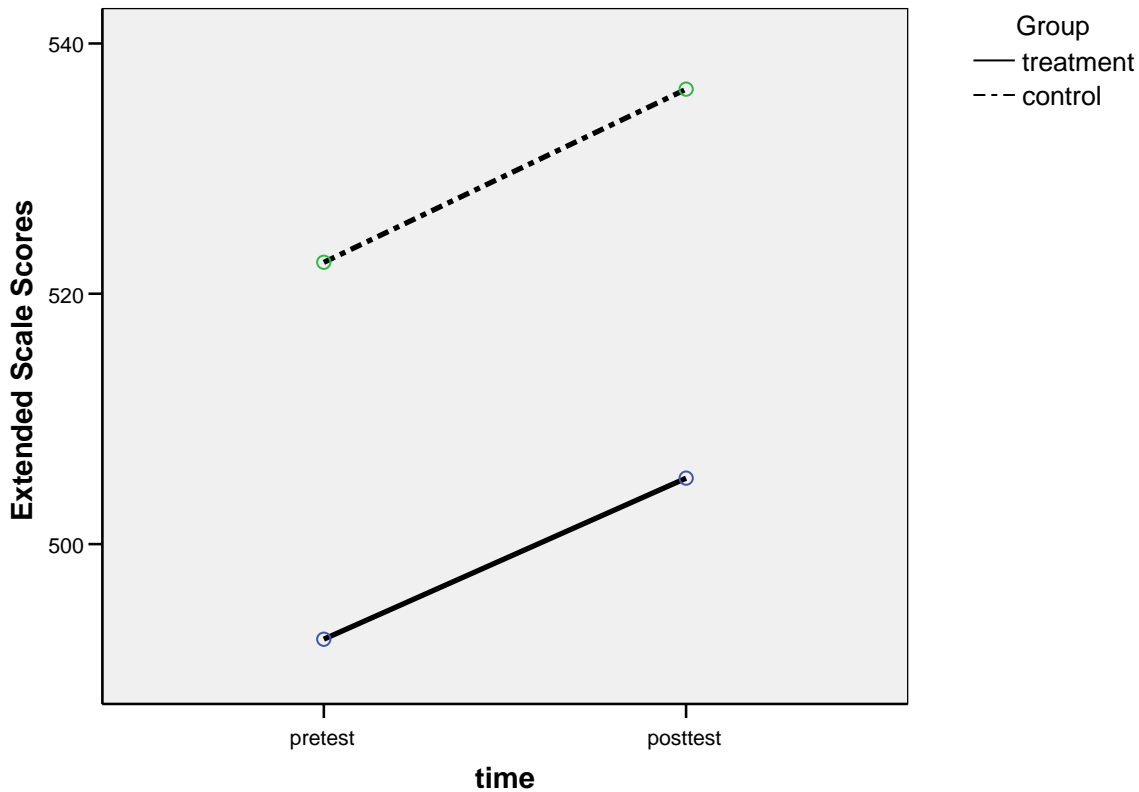
	Treatment Students n = 687		Control Students n = 580	
	Pretest	Posttest	Pretest	Posttest
M	492.39	505.26	522.51	536.34
SD	29.17	31.31	39.89	39.19

Note: Numbers represent students with matched pretest and posttest scores.

materials and those who did not (Figure 4 provides a visual depiction of the vocabulary *GMRT-4* ESS scores). However, since the control group began the program with a significantly higher

vocabulary pretest score than the treatment group<sup>10</sup>, it can be said that the AMP Reading program allowed the treatment students to make the same amount of progress as the initially higher-achieving control group.

Figure 4. Vocabulary ESS Scores by Condition and Time



*Comprehension Score.* In order to examine the differences in student performance from pretest to posttest on the *GMRT-4* comprehension subtest for students in AMP Reading classrooms as compared to control classrooms, researchers again conducted repeated-measures ANOVA. Table 21 displays the pretest and posttest performance of treatment and control students. There was a small but significant difference between the pretest and posttest for all students, Wilks's  $\Lambda = .99$ ,  $F(1, 1227) = 6.4$ ,  $p < .05$ , multivariate  $\eta^2 = .005$ .

<sup>10</sup> Although random assignment of teachers to treatment and control groups was employed, control group students, on average, were higher achievers than AMP Reading students at the beginning of the study. That said, it is imperative to note that any differences in pretest scores were counteracted during the statistical analysis.

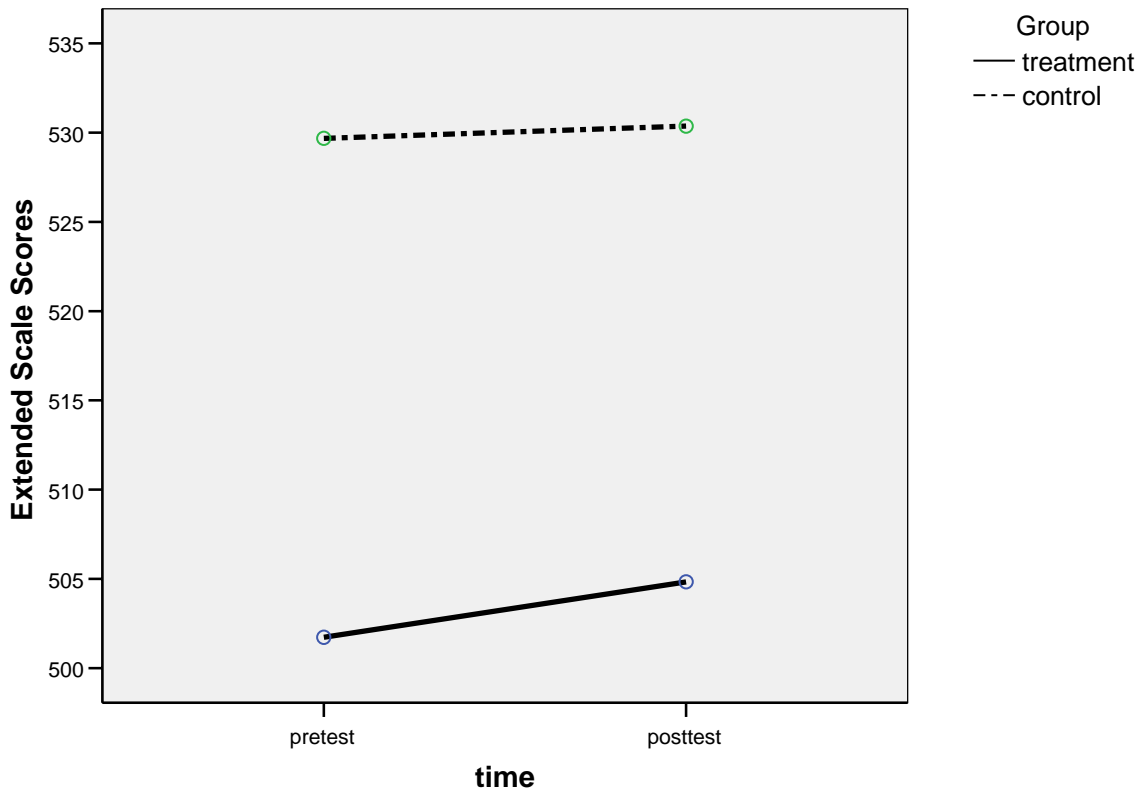
**Table 21. Means and Standard Deviations for ESS Comprehension Scores**

	Treatment Students n = 656		Control Students n = 573	
	Pretest	Posttest	Pretest	Posttest
M	501.73	504.83	529.68	530.37
SD	30.01	29.27	36.62	37.23

Note: Numbers represent students with matched pretest and posttest scores.

Although the comprehension score gains in the treatment group were larger than those in the control group, this interaction between time and group was again not statistically significant. Figure 5 provides a visual depiction of the comprehension *GMRT-4* ESS scores.

**Figure 5. Comprehension ESS Scores by Condition and Time**



*Total Score.* Finally, the pretest to posttest gains on the *GMRT-4* total score for students in both conditions were analyzed with repeated-measures ANOVA. Table 22 displays the pretest and posttest performance of treatment and control students on the total score. As with the two subtest scores, there was a significant difference between the pretest and posttest for all students, Wilks's  $\Lambda = .85$ ,  $F(1, 1216) = 217.71$ ,  $p < .001$ , multivariate

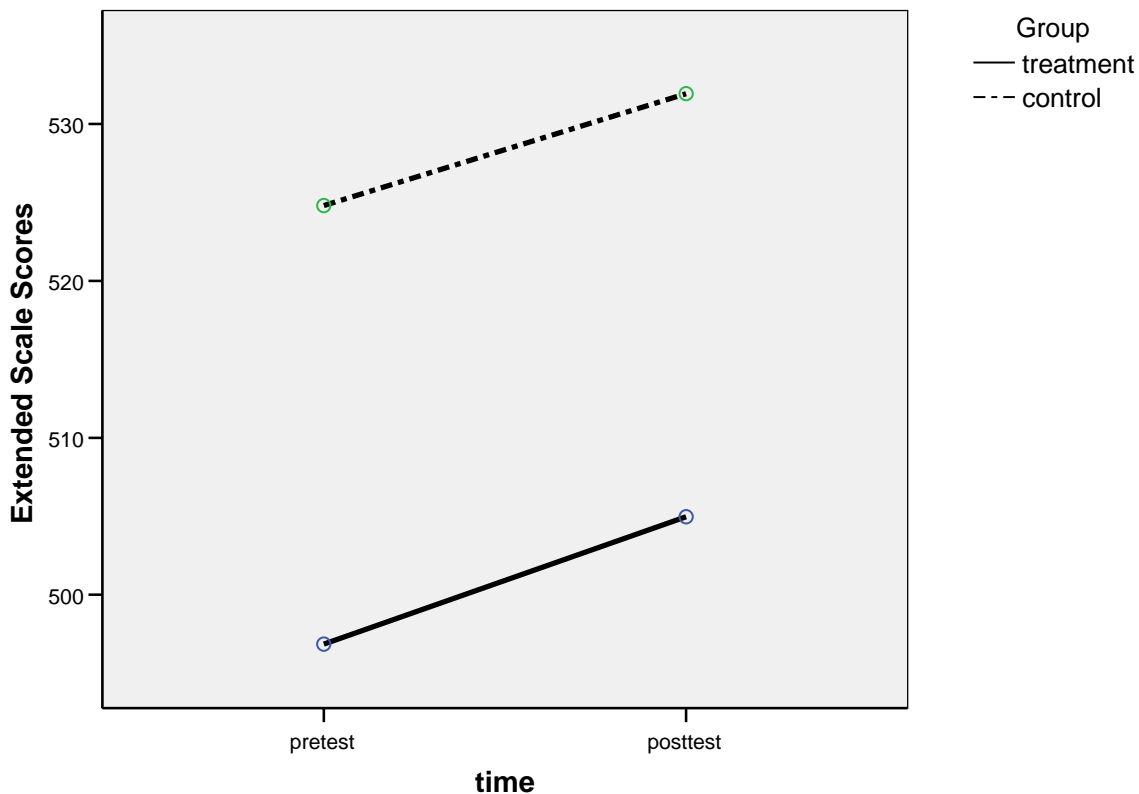
**Table 22. Means and Standard Deviations for ESS Total Scores**

	Treatment Students n = 656		Control Students n = 573	
	Pretest	Posttest	Pretest	Posttest
M	496.85	504.96	524.80	531.93
SD	25.37	25.56	34.06	34.80

Note: Numbers represent students with matched pretest and posttest scores.

$\eta^2 = .152$ . Again, however, the interaction between time and group was not statistically significant, although readers should note that the initially lower-achieving treatment group improved at the same rate as the initially higher-achieving control group. Figure 6 provides a visual depiction of the total *GMRT-4* ESS scores.

**Figure 6. Total ESS Scores by Condition and Time**



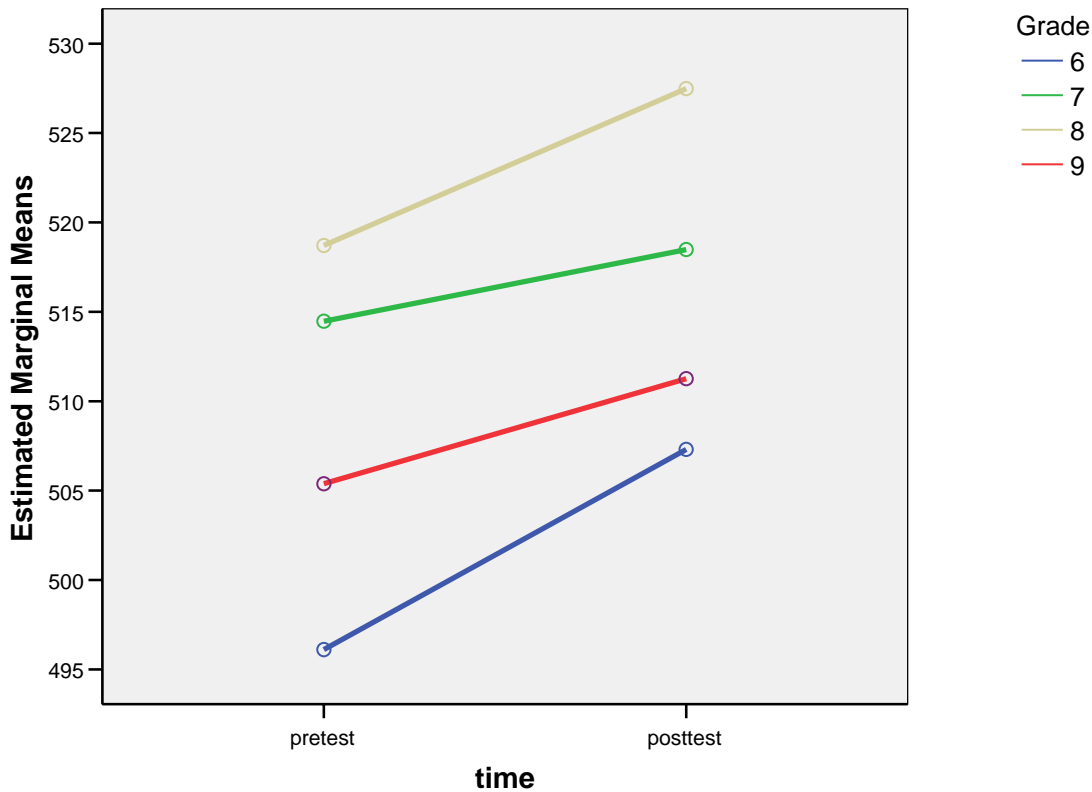
***Performance Within and Between Grades***

*Performance of groups within grades.* In order to determine if there were significant differences between treatment and control groups within particular grades, researchers again used repeated-measures ANOVA. Sixth-, seventh-, eighth-, and ninth-grade *GMRT-4* ESS vocabulary, comprehension, and total scores for each group were compared in separate analyses. For all analyses, the only significant differences between treatment and control were found with seventh-grade *GMRT-4* ESS comprehension scores. In this comparison, each group’s scores (see Table 23) decreased between

pretest and posttest. *GMRT-4* ESS scores declined significantly less than those of the treatment group.

*Performance between grades.* Because there were generally no differences between groups in the different grades, researchers then explored differences between grades for the entire sample. Repeated-measures ANOVAs were conducted to determine the effect of time (pretest to posttest), the interaction of time and grade, the interaction of time and group, and the interaction of time, grade, and group on *GMRT-4* ESS total scores. In this analysis, there were no significant differences found in the interaction of time and group and the interaction of time, grade, and group. The ANOVA for the interaction of time and grade was significant, however (Wilks's  $\Lambda = .97$ ,  $F(3, 1210) = 11.31$ ,  $p < .001$ , multivariate  $\eta^2 = .027$ ). Follow-up pairwise comparisons revealed that the differences were between the 6<sup>th</sup> and 7<sup>th</sup> grade scores ( $p < .001$ ) and between the 6<sup>th</sup> and 8<sup>th</sup> grade scores ( $p < .001$ ). Therefore, students in the 6<sup>th</sup> grade improved at a significantly higher rate than the 7<sup>th</sup> and 8<sup>th</sup> grade students. Figure 7 depicts these differences.

**Figure 7. Differences between Grades on ESS Total Scores**



***Performance within the Treatment Group with an Implementation Covariate***

To explore the effects of AMP Reading implementation on changes in *GMRT-4* ESS scores, researchers developed an implementation variable (possible range 0-10) that described what proportion of the lessons had been completed in each treatment classroom. For treatment teachers, the implementation scores ranged from 1.75 to 9.29, with a median of 3.73. This implementation

variable was used as a covariate in a repeated-measures ANCOVA analysis of the treatment group (the control group was not included because those teachers did not have an AMP Reading implementation score).

*Vocabulary subtest.* On the vocabulary subtest, when the influence of implementation was held constant as a covariate, the time variable no longer had a significant effect. This means that the level of implementation in treatment classrooms accounted for most of the change in vocabulary scores from the pretest to the posttest. In fact, the interaction of implementation level and time was significant (Wilks's  $\Lambda = .97$ ,  $F(1, 606) = 17.95$ ,  $p < .001$ , multivariate  $\eta^2 = .029$ ). Therefore, students improved more in vocabulary as teachers implemented more lessons of the AMP Reading program.

*Comprehension subtest.* On the comprehension subtest, when the influence of implementation was held constant as a covariate the time variable still had a significant effect (Wilks's  $\Lambda = .99$ ,  $F(1, 577) = 7.05$ ,  $p < .01$ , multivariate  $\eta^2 = .012$ ). This means that the comprehension score increased significantly independent of the teachers' level of AMP Reading implementation. However, the interaction of implementation level and time was also significant (Wilks's  $\Lambda = .97$ ,  $F(1, 577) = 15.68$ ,  $p < .001$ , multivariate  $\eta^2 = .026$ ). Therefore, students still improved more in comprehension as teachers implemented more lessons of the AMP Reading program, although gains in comprehension were less dependent on implementation level.

*Total score.* The results for the *GMRT-4* ESS total score were similar to those for the vocabulary subtest, which is expected because the pretest-posttest score difference in vocabulary was larger than it was for comprehension. When the influence of implementation on total score was held constant as a covariate, the time variable no longer had a significant effect. Therefore, the level of implementation in treatment classrooms accounted for most of the change in total *GMRT-4* ESS scores from the pretest to the posttest. As in the vocabulary subtest, the interaction of implementation level and time was significant (Wilks's  $\Lambda = .96$ ,  $F(1, 570) = 24.48$ ,  $p < .001$ , multivariate  $\eta^2 = .041$ ). Therefore, students' reading scores increased as teachers implemented more lessons of AMP Reading. In this study, the interaction of time and implementation was responsible for 4.1% of the increase in total *GMRT-4* ESS scores from pretest to posttest.

### ***Performance By Units Completed***

In addition, McREL examined student outcome data by units completed. These analyses supplement the analyses presented in the section above titled ***Performance within the Treatment Group with an Implementation Covariate***. To conduct these descriptive ancillary analyses, we rounded down to the nearest AMP-Reading Unit completed; hence, if a student was exposed to the 10<sup>th</sup> lesson in Unit 4, their Unit of record was 3.

Table 23 provides assessments of corresponding grade levels for the average ESS-Total scores, based on PR 50. The row labeled “Overall” for each grade level reflects combined data for all treatment students in a grade. The “Unit” rows illustrate data disaggregated by the number of Units completed. In both instances, a determination was made regarding where students fell in terms of the tabled grade levels provided in the *GMRT Manual for Scoring and Interpretation*. For those cases where the average ESS-Total score was the same as – or only fractions away from – tabled values, does the grade level reflect a single semester such as Fall or Spring. Grade levels reflect our interpretations of the *semester range* where average ESS-Total scores place the groups of students and

thus may not bear direct correspondence to data presented in subsequent tables. Based on estimated corresponding grade levels at PR50, researchers calculated the average grade level advanced. These ranged between -.5 and 2.

**Table 23: Means for ESS Total Scores by Grade and by Units Completed, Treatment Schools Only (estimation of corresponding grade levels based on ESS at PR50 and estimation of grade levels advanced over one year shown)**

Units Completed	Pretest n	Pretest Mean	Corresponding Grade at PR 50, pretest	Posttest n	Posttest Mean	Corresponding Grade at PR 50, posttest	Grade Levels Advanced (based on PR 50)
<b>6<sup>th</sup> grade</b>							
<i>Overall</i>	281	489.03	<i>Fall4-Spring4</i>	221	502.59	<i>Fall5-Spring5</i>	1
1 Unit	115	485.38	Fall4-Spring4	82	497.63	Spring4-Fall5	0.5
2 Units	24	482.46	Fall4	19	498.05	Spring4-Fall5	0.75
3 Units	41	493.46	Spring4-Fall5	30	503.63	Fall5-Spring5	0.75
4 Units	58	485.68	Fall4-Spring4	51	504.47	Fall5-Spring5	1
5 Units	22	501.38	Fall5-Spring5	20	504.90	Fall5-Spring5	0
6 Units	21	489.03	Fall4-Spring4	19	519.37	Fall6-Spring6	2
<b>7<sup>th</sup> Grade</b>							
<i>Overall</i>	259	497.02	<i>Spring4-Fall5</i>	215	503.49	<i>Fall5-Spring5</i>	1
1 Unit	28	511.86	Spring5-Fall6	28	512.96	Spring5-Fall6	0
2 Units	165	497.87	Spring4-Fall5	129	502.68	Fall5-Spring5	0.75
4 Units	66	488.59	Fall4-Spring4	58	500.71	Fall5	0.5
<b>8<sup>th</sup> Grade</b>							
<i>Overall</i>	223	498.52	<i>Spring4-Fall5</i>	148	506.44	<i>Fall5-Spring5</i>	1
1 Unit	35	488.74	Fall4-Spring4	6	478.67	Spring3-Fall4	-0.5
2 Units	5	509.40	Spring5-Fall6	5	509.80	Spring5-Fall6	0
3 Units	31	507.29	Spring5	20	518.55	Fall6-Spring6	0.75
4 Units	152	498.62	Spring4-Fall5	117	505.65	Fall5-Spring5	0.75

Researchers also examined students’ actual average percentile rank (based on both tabled norms for in- and out-of-level norms). The average PR 50 advanced for each grade for the “Overall” grade group and for each grade by Units completed was examined. These ranged from -4 to 27. See Table 24.

**Table 24: Means for ESS Total Scores by Grade and by Units Completed, Treatment Schools Only (estimation of Percentile Ranks based on ESS-Total scores over one year)**

Units Completed	Pretest n	Pretest Mean	Corresponding GMRT PR (vs own grade level)	Posttest n	Posttest Mean	Corresponding GMRT PR (vs own grade level)	PR Advanced
<b>6<sup>th</sup> grade</b>							
<i>Overall</i>	281	489.03	22	221	502.59	32	10
1 Unit	115	485.38	19	82	497.63	27	8
2 Units	24	482.46	16	19	498.05	27	11
3 Units	41	493.46	27	30	503.63	33	6
4 Units	58	485.68	20	51	504.47	33	13
5 Units	22	501.38	35	20	504.90	33	-2
6 Units	21	489.03	22	19	519.37	49	27
<b>7<sup>th</sup> Grade</b>							
<i>Overall</i>	259	497.02	17	215	503.49	20	3
1 Unit	28	511.86	33	28	512.96	32	-1
2 Units	165	497.87	18	129	502.68	20	2
4 Units	66	488.59	11	58	500.71	18	7
<b>8<sup>th</sup> Grade</b>							
<i>Overall</i>	223	498.52	13	148	506.44	16	3
1 Unit	35	488.74	8	6	478.67	4	-4
2 Units	5	509.40	21	5	509.80	19	-2
3 Units	31	507.29	20	20	518.55	28	8
4 Units	152	498.62	13	117	505.65	16	3

Finally, researchers converted ESS-Total scores to Grade Equivalents for the “Overall” grade group and for each grade by Units completed<sup>11</sup>. Table 25 provides these data.

**Table 25: Means for ESS Total Scores by Grade and by Units Completed, Treatment Schools Only (estimation of Grade Equivalents based on ESS-Total scores over one year)**

Units Completed	Pretest n	Pretest Mean	Corresponding GMRT Grade Equivalent	Posttest n	Posttest Mean	Corresponding GMRT Grade Equivalent	Grade Equivalent Levels Advanced*
<b>6<sup>th</sup> grade</b>							
<i>Overall</i>	281	489.03	4.5	221	502.59	5.3	0.8
1 Unit	115	485.38	4.3	82	497.63	5.1	0.8
2 Units	24	482.46	4.1	19	498.05	5.1	1
3 Units	41	493.46	4.8	30	503.63	5.5	0.7
4 Units	58	485.68	4.4	51	504.47	5.5	1.1
5 Units	22	501.38	5.3	20	504.90	5.5	0.2
6 Units	21	489.03	4.5	19	519.37	6.5	2
<b>7<sup>th</sup> Grade</b>							
<i>Overall</i>	259	497.02	4.9	215	503.49	5.3	0.4
1 Unit	28	511.86	6.1	28	512.96	6.1	0
2 Units	165	497.87	5.1	129	502.68	5.4	0.3
4 Units	66	488.59	4.5	58	500.71	5.3	0.8
<b>8<sup>th</sup> Grade</b>							
<i>Overall</i>	223	498.52	5.1	148	506.44	5.5	0.4
1 Unit	35	488.74	4.5	6	478.67	4	-0.5
2 Units	5	509.40	5.8	5	509.80	5.8	0
3 Units	31	507.29	5.7	20	518.55	6.6	0.9
4 Units	152	498.62	5.2	117	505.65	5.6	0.4

\*Mean ESS scores were transformed into Grade Equivalents (GEs). However, it is important to note that Ges do not represent equal units. GEs reflect yearly growth achievement of average students. GEs are not meaningful when they are not within the intended grade ranges of the test. GEs are not useful for comparing scores.

<sup>11</sup> Although Grade Equivalents (and the actual Percentile Ranks) are more easily quantified, these results are not as readily interpretable<sup>11</sup>.

### ***Additional Analysis***

As previously noted, one site implemented AMP with smaller groups of students. In order to examine the differences in student performance from pretest to posttest on the *GMRT-4* for students in AMP Reading classrooms with low student-to-teacher ratios, students in AMP Reading classrooms with typical student-to-teacher ratios, students in control classrooms with low student-to-teacher ratios, and students in control classrooms with typical student-to-teacher ratios, researchers conducted repeated-measures ANOVA. Table 24 displays the pretest and posttest performance of treatment and control students for both the low student-to-teacher classrooms and for the typical student-to-teacher ratio classrooms. There was a significant interaction between time and group, Wilks's  $\Lambda = .99$ ,  $F(3, 1214) = 3.964$ ,  $p < .01$ , multivariate  $\eta^2 = .01$ . This indicates that statistically there were differences on test scores between students who used the AMP Reading materials in low student-to-teacher settings and those who did not use the materials in such settings. Follow-up pairwise comparisons revealed that the differences were between (a) AMP Reading and control students (previously reported), (b) typical student-to-teacher ratio control classrooms and low-student-to teacher ratio control classrooms (showing an advantage for the smaller class size), and (c) typical student-to-teacher ratio control classrooms and low-student-to teacher ratio AMP Reading classrooms (showing a slight advantage for the smaller class size). These findings need to be interpreted with caution; however, as differences in cell sizes (and a loss of statistical power) could lead to spurious effects. Additional research on the effects of smaller class sizes is warranted.

**Table 26. Means and Standard Deviations for ESS Total Scores by Student-Teacher Ratio Classrooms**

	Treatment Students – Typical Student-to-Teacher Ratio n = 613		Treatment Students – Low Student-to-Teacher Ratio n = 39		Control Students – Typical Student-to-Teacher Ratio N = 543		Control Students – Low Student-to-Teacher Ratio n = 26	
	<u>Pretest</u>	<u>Posttest</u>	<u>Pretest</u>	<u>Posttest</u>	<u>Pretest</u>	<u>Posttest</u>	<u>Pretest</u>	<u>Posttest</u>
M	497.38	505.25	488.56	500.54	526.55	531.18	488.23	505.69
SD	25.46	25.76	22.74	22.54	33.66	34.80	18.04	22.72

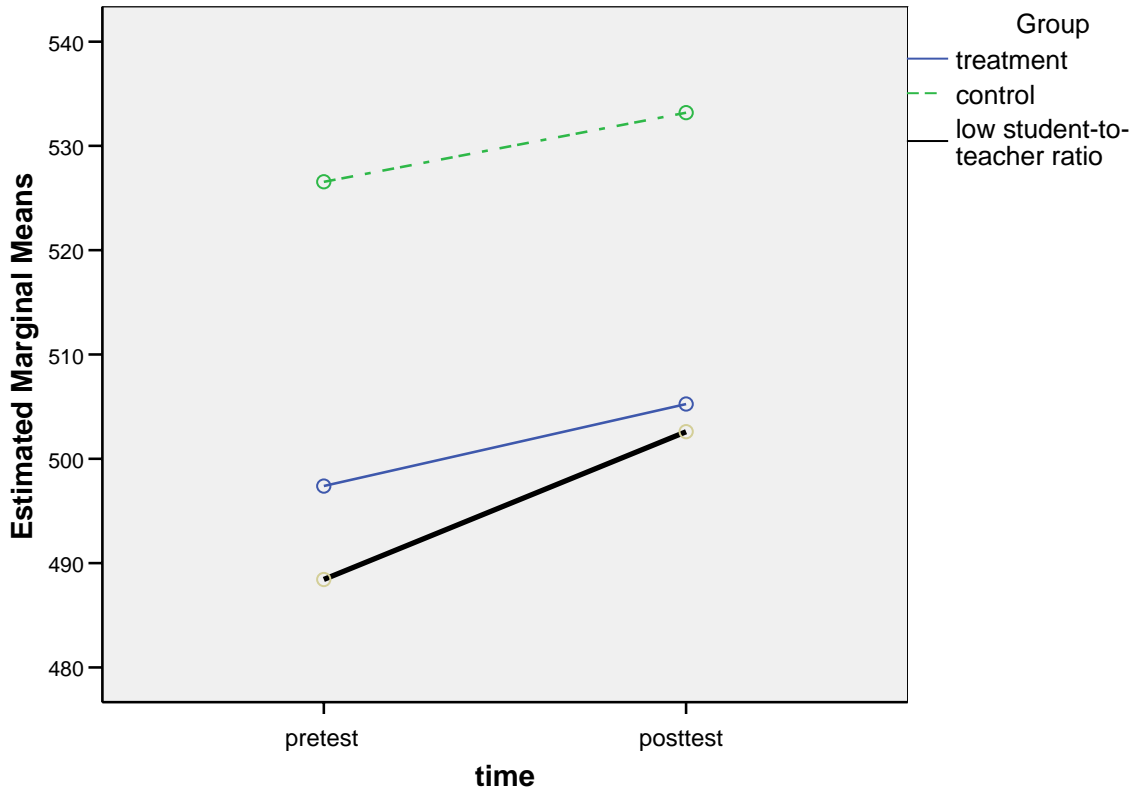
Note: Numbers represent students with matched pretest and posttest scores.

An additional analysis was conducted to compare the effects of a low student-to-teacher ratio independent of the intervention implemented to the effects for students in AMP Reading classrooms with typical student-to-teacher ratios and students in control classrooms with typical student-to-teacher ratios<sup>12</sup>. Repeated-measures ANOVA were again conducted and a significant interaction between time and group was revealed, Wilks's  $\Lambda = .99$ ,  $F(2, 1215) = 5.213$ ,  $p < .01$ , multivariate  $\eta^2 = .009$ . This indicates that statistically there were differences on test scores between students in low student-to-teacher settings and those in classrooms with typical student-to-teacher ratios. Follow-up pairwise comparisons revealed that the differences were between AMP Reading and control students (previously reported) and typical student-to-teacher ratio control classrooms and low-student-to teacher ratio classrooms (again showing an advantage for smaller class size – this

<sup>12</sup> Collapsing the AMP Reading and control classrooms in the site with a low student-to-teacher ratio increased the power of the analysis; however, this analysis is useful only in terms of understanding the effects of smaller class sizes, *regardless of AMP Reading*.

time, independent of intervention). Still, these findings need to be interpreted with caution as differences in cell sizes could lead to the detection of a spurious effect. Figure 9 provides a visual depiction of these differences.

**Figure 8. ESS Total Scores by Class Size.**



## CONCLUSIONS

The primary purpose of this evaluation was to examine the implementation and effectiveness of AMP Reading with respect to student achievement. Sixth-, seventh-, and eighth-grade classrooms in 16 schools were randomly assigned to either implement AMP Reading during the 2005-2006 school year or to teach reading using typical materials.

Weekly implementation logs revealed that participating teachers varied in the extent to which they implemented the AMP program, with some teachers considered “higher” implementers than others because they advanced much farther into the program with their classes. Classroom observations were utilized to complement teacher implementation log data and revealed that the majority of students in classrooms were engaged in AMP Reading lessons and that the majority of AMP Reading teachers followed the program guidelines (although no teachers completed all lessons for a given grade level and many teachers worked at a slower pace than suggested in Teacher’s Guides).

On average, AMP Reading teachers reported spending more than 60 minutes a week in planning and preparation for utilizing the study materials. However, implementing a new program is invariably more difficult compared to using a familiar program. Teachers are learning along with the students, a fact that can sometimes result in lower-than-anticipated test scores. Control teachers and their students, on the other hand, were allowed to utilize familiar materials and to supplement their materials in any way that they saw fit, including drawing from additional resources or materials used in the past.

In terms of student achievement, the vocabulary and comprehension subtests and total scores of the *GMRT-4* were used to assess students’ reading achievement at the beginning and end of teachers’ implementation of the AMP Reading program. On average, teachers completed three of the AMP units. Researchers found that:

- Comparing all students with available pretest and posttest total Extended Scale Scores (ESSs) on the *GMRT-4* revealed significant pretest to posttest gains for the total score and for both the vocabulary and comprehension subtest scores, with the larger gains occurring in the vocabulary subtest. Vocabulary appears to be an area that AMP Reading impacts.
- Student performance from pretest to posttest on the *GMRT-4* for students in the AMP Reading classrooms as compared to students in the control condition revealed no significant differences on total or subtest score gains, indicating that there were no differences on test scores between those students who used the AMP materials and those who did not. However, the students in the control condition were initially significantly higher-achieving on the *GMRT-4* than were the AMP Reading treatment group, as indicated by their pretest scores. The students using AMP Reading improved at the same rate as the control students. This is an important finding—the gap between the initially lower-achieving student using AMP Reading and the initially higher-achieving control group student did not grow larger.
- Students in the sixth grade made significantly greater progress during the study period than did students in all other grades.

- In analyses intended to measure the effect of level of implementation on score gains in treatment classrooms (defined as the proportion of AMP Reading lessons completed by teachers), implementation level was found to account for most of the change in total *GMRT-4* total scores from the pretest to the posttest. This was also true for the vocabulary subtest. On the comprehension subtest, when the influence of implementation was held constant the time variable still had a significant effect. However, the interaction of implementation level and time was also significant. Therefore, students still improve more in comprehension as teachers implement more lessons of AMP Reading, but gains in comprehension were less dependent on implementation level.
- Finally, because several of the implementation sites lent themselves to follow-up analyses, McREL researchers examined an additional question and determined that smaller class sizes make a difference (that implementing reading in classrooms with low student-to-teacher ratios may result in higher student gains).

Given the importance of reading both in terms of student academic success and lifelong success, what is perhaps most important for this inquiry into the effects of the AMP Reading materials will be determining if there are affective outcomes associated with using AMP Reading—for instance, are those students who when starting AMP Reading were poor and reluctant readers more interested in reading and eager to engage in reading afterwards? Comparisons of standardized test scores revealed student gains over the course of the intervention (the school year in this case); however, these should be accompanied with additional considerations around potential, additional, and perhaps long-range benefits of participating.

Anecdotally, teachers in AMP Reading reported that students in their classrooms who are typically “underperformers” benefited more from AMP Reading than their students who were typically higher achievers. Additional analyses were conducted to examine whether teachers’ perceptions of the differences for these students revealed that “underperformers” did, indeed, fare better in the program than their higher performing counterparts. Management issues in AMP Reading classrooms were sometimes a result of teachers being unable to keep children on the tasks in their AMP Reading Student Guides rather than reading ahead in their AMP Reading trade books—student interest in the AMP Reading materials was very high. Reports from site coordinators and teachers suggested that students enjoyed the materials and seemed to be more engaged, positive signs that these adolescent readers may have become more interested in reading.

## REFERENCES

Biancarosa, G., & Snow, C. E. (2004). *Reading next—A vision for action and research in middle and high school literacy: A report from Carnegie Corporation of New York*. Washington, DC: Alliance for Excellent Education. Retrieved October 3, 2005, from <http://www.all4ed.org/publications/ReadingNext/index.html>

Lipsey, M. W. (1990). *Design sensitivity: Statistical power for experimental research*. Thousand Oaks, CA: Sage.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., and Dreyer, L. G. (2000). *Gates-MacGinitie Reading Tests: Manual for Scoring and Interpretation* (4th ed.). Itasca, IL: Riverside Publishing.

## APPENDIX A: AVERAGE RATINGS UNIT 1 – 6 SUMMARIES

### Average Item Ratings for Individual Unit Summaries, *Unit 1: Summarizing*

Survey Items	Unit 1: <i>Summarizing</i>	
Quality and Utility	Average Item Rating	Standard Deviation Item Rating
The layout, or physical design of the materials was well organized.	4.46	0.58
The pedagogy, or instructional design, of the materials was easy to follow.	4.46	0.58
The lesson plans were easy to follow.	4.46	0.58
Alignment with standards enhanced the materials' utility.	4.20	0.65
The standards to which the materials were aligned were relevant to my instruction planning.	4.12	0.67
The materials' alignment to standards was instrumental in helping students to meet those standards.	4.24	0.66
The assessment strategies enhanced the materials' utility.	3.96	0.92
The materials allowed me to engage in ongoing assessment of my teaching.	3.77	0.91
The materials allowed me to engage in ongoing assessment of student learning.	3.77	0.91
The materials included appropriate instructional practices for meeting the needs of diverse students.	3.92	1.08
The materials were culturally appropriate and included activities for minority students.	4.08	0.91
The materials were developmentally appropriate.	4.31	0.68
The materials were appropriate for special education students.	4.10	1.02
The materials were appropriate for ELL students.	3.73	1.08
Overall, the materials are of high quality.	4.44	0.51
Overall, the materials are very useful.	4.42	0.58
Effect on Instruction and Learning	Item rating	
I felt comfortable using this unit.	4.32	0.56
The training prepared me to implement this unit.	3.96	0.84
The training provided me with the knowledge and skills I needed to implement this unit.	3.92	0.88
Implementing this unit was challenging.	2.88	1.17
Using these materials improved my understanding of how students learn to read.	3.72	0.98
Using these materials gave me a better understanding of how to facilitate learning of comprehension strategies.	3.96	0.84
Using the materials enhanced my capacity to teach reading.	4.00	0.71
I would recommend this unit to a colleague.	4.12	0.73
The materials enhanced students' interest in reading.	3.96	0.84
Students were more interested in reading as a result of using these materials.	3.92	0.91
Students were engaged in the reading assignments.	4.20	0.65
The materials improved students' abilities to use the comprehension strategy ( <i>Summarizing</i> ).	4.25	0.68

**Average Item Ratings for Individual Unit Summaries, *Unit 2 Questioning***

Survey Items	Unit 2: Questioning	
	Average Item Rating	Standard Deviation Item Rating
The layout, or physical design of the materials was well organized.	4.36	0.73
The pedagogy, or instructional design, of the materials was easy to follow.	4.41	0.73
The lesson plans were easy to follow.	4.55	0.51
Alignment with standards enhanced the materials' utility.	4.10	0.79
The standards to which the materials were aligned were relevant to my instruction planning.	4.14	0.79
The materials' alignment to standards was instrumental in helping students to meet those standards.	4.10	0.70
The assessment strategies enhanced the materials' utility.	3.77	0.97
The materials allowed me to engage in ongoing assessment of my teaching.	3.80	0.83
The materials allowed me to engage in ongoing assessment of student learning.	3.57	0.98
The materials included appropriate instructional practices for meeting the needs of diverse students.	3.95	0.79
The materials were culturally appropriate and included activities for minority students.	4.05	0.74
The materials were developmentally appropriate.	4.14	0.64
The materials were appropriate for special education students.	3.89	1.05
The materials were appropriate for ELL students.	3.61	0.85
Overall, the materials are of high quality.	4.27	0.70
Overall, the materials are very useful.	4.27	0.70
<b>Effect on Instruction and Learning</b>	<b>Item rating</b>	
I felt comfortable using this unit.	3.86	0.71
The training prepared me to implement this unit.	3.59	0.73
The training provided me with the knowledge and skills I needed to implement this unit.	3.59	0.73
Implementing this unit was challenging.	2.73	1.03
Using these materials improved my understanding of how students learn to read.	3.64	0.79
Using these materials gave me a better understanding of how to facilitate learning of comprehension strategies.	3.82	0.80
Using the materials enhanced my capacity to teach reading.	3.77	0.87
I would recommend this unit to a colleague.	3.86	0.71
The materials enhanced students' interest in reading.	3.55	0.91
Students were more interested in reading as a result of using these materials.	3.55	0.91
Students were engaged in the reading assignments.	3.86	0.77
The materials improved students' abilities to use the comprehension strategy ( <i>Questioning</i> ).	4.00	0.62

**Average Item Ratings for Individual Unit Summaries, *Unit 3: Predicting***

Survey Items	Unit 3: Predicting	
Quality and Utility	Average Item Rating	Standard Deviation Item Rating
The layout, or physical design of the materials was well organized.	4.50	0.53
The pedagogy, or instructional design, of the materials was easy to follow.	4.60	0.52
The lesson plans were easy to follow.	4.70	0.48
Alignment with standards enhanced the materials' utility.	4.22	0.67
The standards to which the materials were aligned were relevant to my instruction planning.	4.11	0.78
The materials' alignment to standards was instrumental in helping students to meet those standards.	4.10	0.74
The assessment strategies enhanced the materials' utility.	3.90	0.88
The materials allowed me to engage in ongoing assessment of my teaching.	4.11	0.60
The materials allowed me to engage in ongoing assessment of student learning.	4.10	0.57
The materials included appropriate instructional practices for meeting the needs of diverse students.	4.00	0.82
The materials were culturally appropriate and included activities for minority students.	4.22	0.44
The materials were developmentally appropriate.	4.20	0.42
The materials were appropriate for special education students.	3.89	0.78
The materials were appropriate for ELL students.	3.80	0.63
Overall, the materials are of high quality.	4.40	0.52
Overall, the materials are very useful.	4.44	0.53
Effect on Instruction and Learning	Item rating	
I felt comfortable using this unit.	4.40	0.52
The training prepared me to implement this unit.	3.70	0.48
The training provided me with the knowledge and skills I needed to implement this unit.	3.60	0.52
Implementing this unit was challenging.	2.50	1.08
Using these materials improved my understanding of how students learn to read.	3.80	0.63
Using these materials gave me a better understanding of how to facilitate learning of comprehension strategies.	4.10	0.74
Using the materials enhanced my capacity to teach reading.	3.70	0.48
I would recommend this unit to a colleague.	3.90	0.88
The materials enhanced students' interest in reading.	4.00	1.05
Students were more interested in reading as a result of using these materials.	3.80	1.14
Students were engaged in the reading assignments.	3.80	0.79
The materials improved students' abilities to use the comprehension strategy ( <i>Predicting</i> ).	4.10	0.74

**Average Item Ratings for Individual Unit Summaries, Unit 4: Text Structure**

Survey Items	Unit 4: Text Structure	
	Average Item Rating	Standard Deviation Item Rating
The layout, or physical design of the materials was well organized.	4.25	0.96
The pedagogy, or instructional design, of the materials was easy to follow.	4.50	0.58
The lesson plans were easy to follow.	4.75	0.50
Alignment with standards enhanced the materials' utility.	3.50	1.73
The standards to which the materials were aligned were relevant to my instruction planning.	3.00	1.73
The materials' alignment to standards was instrumental in helping students to meet those standards.	3.75	1.50
The assessment strategies enhanced the materials' utility.	4.50	0.58
The materials allowed me to engage in ongoing assessment of my teaching.	3.67	1.53
The materials allowed me to engage in ongoing assessment of student learning.	3.75	1.26
The materials included appropriate instructional practices for meeting the needs of diverse students.	4.50	0.58
The materials were culturally appropriate and included activities for minority students.	4.50	0.58
The materials were developmentally appropriate.	4.75	0.50
The materials were appropriate for special education students.	4.33	0.58
The materials were appropriate for ELL students.	3.67	0.58
Overall, the materials are of high quality.	4.50	0.58
Overall, the materials are very useful.	4.25	0.50
<b>Effect on Instruction and Learning</b>	<b>Item rating</b>	
I felt comfortable using this unit.	4.25	0.50
The training prepared me to implement this unit.	4.25	0.50
The training provided me with the knowledge and skills I needed to implement this unit.	4.25	0.50
Implementing this unit was challenging.	2.50	0.58
Using these materials improved my understanding of how students learn to read.	3.50	1.29
Using these materials gave me a better understanding of how to facilitate learning of comprehension strategies.	4.25	0.96
Using the materials enhanced my capacity to teach reading.	4.00	0.82
I would recommend this unit to a colleague.	4.00	0.82
The materials enhanced students' interest in reading.	3.75	0.50
Students were more interested in reading as a result of using these materials.	3.25	0.50
Students were engaged in the reading assignments.	4.25	0.50
The materials improved students' abilities to use the comprehension strategy ( <i>Text Structure</i> ).	4.00	0.82

**Average Item Ratings for Individual Unit Summaries, *Unit 5: Visualizing***

Survey Items	Unit 5: Visualizing	
Quality and Utility	Average Item Rating	Standard Deviation Item Rating
The layout, or physical design of the materials was well organized.	4.50	0.71
The pedagogy, or instructional design, of the materials was easy to follow.	4.50	0.71
The lesson plans were easy to follow.	4.50	0.71
Alignment with standards enhanced the materials' utility.	4.00	0.00
The standards to which the materials were aligned were relevant to my instruction planning.	4.00	0.00
The materials' alignment to standards was instrumental in helping students to meet those standards.	4.00	0.00
The assessment strategies enhanced the materials' utility.	4.00	0.00
The materials allowed me to engage in ongoing assessment of my teaching.	4.00	0.00
The materials allowed me to engage in ongoing assessment of student learning.	4.00	0.00
The materials included appropriate instructional practices for meeting the needs of diverse students.	4.00	0.00
The materials were culturally appropriate and included activities for minority students.	4.00	0.00
The materials were developmentally appropriate.	4.00	0.00
The materials were appropriate for special education students.	4.00	0.00
The materials were appropriate for ELL students.	4.00	0.00
Overall, the materials are of high quality.	4.00	0.00
Overall, the materials are very useful.	4.00	0.00
<b>Effect on Instruction and Learning</b>	<b>Item rating</b>	
I felt comfortable using this unit.	4.00	0.00
The training prepared me to implement this unit.	4.00	0.00
The training provided me with the knowledge and skills I needed to implement this unit.	4.00	0.00
Implementing this unit was challenging.	2.00	0.00
Using these materials improved my understanding of how students learn to read.	4.00	1.41
Using these materials gave me a better understanding of how to facilitate learning of comprehension strategies.	4.00	1.41
Using the materials enhanced my capacity to teach reading.	4.50	0.71
I would recommend this unit to a colleague.	4.50	0.71
The materials enhanced students' interest in reading.	4.50	0.71
Students were more interested in reading as a result of using these materials.	4.00	1.41
Students were engaged in the reading assignments.	4.00	0.00
The materials improved students' abilities to use the comprehension strategy ( <i>Visualizing</i> ).	4.00	0.00

**Average Item Ratings for Individual Unit Summaries, Unit 6: Inference**

Survey Items	Unit 6: Inference	
Quality and Utility	Average Item Rating	Standard Deviation Item Rating
The layout, or physical design of the materials was well organized.	5.00	n/a
The pedagogy, or instructional design, of the materials was easy to follow.	5.00	n/a
The lesson plans were easy to follow.	5.00	n/a
Alignment with standards enhanced the materials' utility.	3.00	n/a
The standards to which the materials were aligned were relevant to my instruction planning.	4.00	n/a
The materials' alignment to standards was instrumental in helping students to meet those standards.	3.00	n/a
The assessment strategies enhanced the materials' utility.	4.00	n/a
The materials allowed me to engage in ongoing assessment of my teaching.	4.00	n/a
The materials allowed me to engage in ongoing assessment of student learning.	4.00	n/a
The materials included appropriate instructional practices for meeting the needs of diverse students.	4.00	n/a
The materials were culturally appropriate and included activities for minority students.	4.00	n/a
The materials were developmentally appropriate.	4.00	n/a
The materials were appropriate for special education students.	4.00	n/a
The materials were appropriate for ELL students.	4.00	n/a
Overall, the materials are of high quality.	4.00	n/a
Overall, the materials are very useful.	4.00	n/a
Effect on Instruction and Learning	Item rating	
I felt comfortable using this unit.	4.00	n/a
The training prepared me to implement this unit.	4.00	n/a
The training provided me with the knowledge and skills I needed to implement this unit.	4.00	n/a
Implementing this unit was challenging.	2.00	n/a
Using these materials improved my understanding of how students learn to read.	3.00	n/a
Using these materials gave me a better understanding of how to facilitate learning of comprehension strategies.	4.00	n/a
Using the materials enhanced my capacity to teach reading.	4.00	n/a
I would recommend this unit to a colleague.	4.00	n/a
The materials enhanced students' interest in reading.	4.00	n/a
Students were more interested in reading as a result of using these materials.	4.00	n/a
Students were engaged in the reading assignments.	4.00	n/a
The materials improved students' abilities to use the comprehension strategy ( <i>Inference</i> ).	4.00	n/a

**APPENDIX B: TOTAL SCORE EXTENDED SCALE SCORES AS CORRESPONDING TO  
MEDIAN ACHIEVEMENT LEVEL**

<b>Grade</b>	<b>Semester</b>	<b>ESS at Percentile Rank 50</b>
3	Fall	459
3	Spring	472
4	Fall	482
4	Spring	492
5	Fall	500
5	Spring	507
6	Fall	514
6	Spring	520
7	Fall	526
7	Spring	531
8	Fall	536
8	Spring	540
9	Fall	544
9	Spring	548

## APPENDIX C: EFFECT SIZE CALCULATION METHODS

Because paired *t*-test analyses examined pre- to posttest differences, the following modification of the standard effect size measure to take into account the correlation of the paired values:

$$ES = \frac{\mu_t - \mu_c}{\sigma\sqrt{1 - r_{tc}}}$$

(where  $\mu_t$  is the posttest,  $\mu_c$  is the pretest, and  $r_{tc}$  represents the correlation of the paired values; Lipsey, 1990).

Effect size for repeated measures ANOVA was calculated using partial  $\eta^2$ . The following formula was used for this calculation:

$$ES = \sqrt{\frac{\eta^2}{1 - \eta^2}}$$