PEARSON READING STREET EFFICACY STUDY

2009-10 FINAL REPORT

Principal Investigator

Guido G. Gatti
Gatti Evaluation Inc.
162 Fairfax Rd.
Pittsburgh, PA 15221
(412) 371-9832
gggatti@gattieval.com

Co-Principal Investigator

Katya Petrochenkov
Gatti Evaluation Inc.

Primary Stakeholder

Funded By **Pearson**

For Information Please Contact:

Marcy Baughman
Director of Academic Research
(724) 863-1621
marcy.baughman@pearsoned.com

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Pearson partnered with Gatti Evaluation to conduct a longitudinal scientific research study during the 2009-10 and 2010-11 school years to support the assertion that the Reading Street © 2011 curriculum effectively increases student English language arts achievement scores and attitudes. The first year of this longitudinal effort is testing the Reading Street program during the first year of implementation, the most challenging year for any new program to impact student achievement. The Reading Street program was tested against comparison classrooms randomly selected within each school that did not implement the Reading Street program. Rather, the comparison classrooms were not required to change reading curricula and teachers continued to utilize their current and preferred reading programs and instructional practices. The secondary goal for this project was to collect information on teacher and student attitudes towards features and aspects of the Reading Street curriculum.

The sample is comprised of students and teachers from 82 diverse elementary grade classrooms in nine schools distributed across six different states (i.e., AZ, CO, MA, MT, OH, and WA). This report describes the methods and results for the 2009-10 school year. All study schools are located in public school districts. Five schools are located in large suburban areas, two are located in a rural or rural fringe area, and one site is located in a mid-sized city. The study schools show considerable variation in ethnicity, students eligible for free/reduced priced lunch, as well as a wide range of reading ability. The evaluation team sought out diversity in the study sample to ensure the Reading Street curriculum would be used by learners of all abilities and backgrounds, thus reflecting the reality that is today's elementary classrooms.

The final study sample consisted of 472 kindergarten (i.e., Reading Street = 269, comparison = 203), 491 1st grade (i.e., Reading Street = 261, comparison = 230) and 517 4th grade (i.e., Reading Street = 265, comparison = 252) students. An assessment battery comprised of the Group Reading Assessment and Diagnostic Evaluation (GRADE) and the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) was used to measure gains in student achievement over the course of the school year. The GRADE subtests included Word Study, Reading Comprehension (i.e., 1st and 4th only), and Listening Comprehension (i.e., kindergarten and 1st only). The DIBELS scales included Phoneme Segmentation and Nonsense Word Fluency at kindergarten and 1st grade, as well as Oral Reading Fluency at 1st and 4th grade. In addition, an Academic Attitude survey and the Elementary Reading Attitude (i.e., ERAS or Garfield) Survey were used to measure gains in student attitudes.

Four different published reading/ELA programs were utilized by those teachers randomly assigned to the comparison condition at kindergarten, six programs at 1st grade, and seven programs at 4th grade. A similar portion of Kindergarten and 1st grade comparison students had teachers that reported using mainly a *teacher-created* program while relatively fewer 4th grade comparison students received a teacher-created reading program. Comparison teachers at each level reported an average of four years using their current reading program. The vast majority of study teachers described their teaching philosophy as a balanced literacy approach. A majority of students attended classrooms where there was a person regularly assisting the classroom teacher (i.e., teacher's aide, paraprofessional, student teacher, reading coach, etc.). Students received an average of 95 minutes of daily reading instruction (Reading Street = 93 minutes, comparison = 96 minutes).

RQ: How well was the Reading Street program implemented?

Pearson provided for Reading Street curriculum specialists to conduct an initial full-day program training session as well as three additional follow-up observation and training sessions. Some Reading Street ancillary materials were not available at the beginning of the school year and the teachers were asked to implement Reading Street with the core components and ancillaries available at the time. As ancillary materials became available, they were distributed to the study schools. The majority of ancillary materials were available by November 2009, with the exception of the Reading Street practice stations. These were not available and thus not utilized. Teachers created their own center stations to go with the program.

Several schools experienced challenges implementing the writing portions of the program, and required extra training and assurance of how the writing portions matched their state standards. It was determined these schools did not fully implement the writing portion of the program with fidelity during the 2009-10 school year.

Reading Street teachers unanimously reported that instruction felt natural to them by winter break. A few months of use and discovery were necessary for the teachers to become fully comfortable with pacing and flow of the program. The same was true for their students falling into the structured routine offered by the Reading Street program. Average ratings from the final training session indicated that Reading Street teachers were implementing the required components well with good pacing. On a 1-10 scale with a rating of five 5 indicating average implementers, kindergarten = 7.4, 1st grade = 7.4, and 4th grade = 7.3. Seventy-seven percent of teachers received a rating of seven or higher.

The Reading Street program is comprised of six units. Each unit is comprised of two volumes, each with three weeks worth of content, totaling six weeks of content per unit. Pacing varied across the Reading Street teachers. Kindergarten teachers tended to complete more of the program than the other grades, with 27% reaching unit six by end-of-year testing, and the remaining teachers reaching unit five. Thirty-eight percent of 1st grade teachers reached unit five, while 50% reached unit four and 12% reached unit three. Fourth grade teachers were the most varied, with 15% reaching unit six, 39% reaching unit five, 15% reaching four, and 31% reaching unit three.

RQ: How did teachers and students react to Reading Street?

Focus group sessions were conducted at each school between April and mid-May. These sessions provided the evaluators with several insights into teacher and student experiences with the program. Teachers and students alike had positive experiences with the Reading Street program. When interviewed, the teacher response to the program was overall positive with 64% of the 799 recorded comments being positive in nature. Teachers appreciated the program's components that support differentiated and small group instruction (ex., leveled and decodable readers), the ongoing progress monitoring and assessment, vocabulary, and multi-day reading selections. Teachers also very much liked that lessons are organized around central themes and that the program adds structure to the weekly reading/ELA instruction.

Teachers also experienced some drawbacks. They felt that the pacing took some time to master and that the concept talk can run long or "*be wordy.*" Many teachers also felt the consumable notebook could be improved in its worksheet content and usability, and that the layout of the teacher's edition could be improved.

Teachers were overwhelmingly positive about their students' interactions with the program. Of the 219 recorded comments, 80% were positive in nature. Many teachers firmly believe that the Reading Street program increases student motivation, participation, and energy in their classrooms. Students responded particularly well to small group time and the structure offered by the program.

RQ: Do students using the Reading Street curriculum demonstrate significant gains in reading/language arts achievement during the first year of curriculum implementation?

Students in kindergarten, 1st, and 4th grades demonstrated educationally meaningful growth with the Reading Street program even when classroom teachers are implementing the program for the first time. The gains in GRADE scores for all three grade levels were very large and statistically significant, 1.74, 2.09, and 0.74 standard deviations for kindergarten, 1st, and 4th grade respectively. The gains for GRADE subtests were also large.

Kindergarten Scale	GRADE Effect Size[1]
GRADE Total	1.74
Word Study	1.72
Listening Comprehension	0.83

1. effect size = baseline to EOY Reading Street group gain / comparison sample standard deviation

1st Grade Scale	GRADE Effect Size[1]
GRADE Total	2.09
Word Study	1.54
Reading Comprehension	1.53
Listening Comprehension	0.65

1. effect size = baseline to EOY Reading Street group gain / comparison sample standard deviation

4th Grade Scale	GRADE Effect Size[1]
GRADE Total	0.74
Word Study	0.70
Reading Comprehension	0.50

1. effect size = baseline to EOY Reading Street group gain / comparison sample standard deviation

RQ: How does reading achievement differ for students using the Reading Street curriculum as compared to their peers using other elementary reading curricula?

For the following results, comparisons on assessment outcomes were made between study groups using model adjusted end-of-year raw score group mean differences. Adjusted group mean differences are calculated holding the effects of confounding factors constant for both groups (i.e., baseline scores, student demographic information, and classroom environment indicators are set to the overall sample mean), effectively removing their influence from the results. The statistical models were able to find small to moderate effect sizes statistically significant (i.e., effect size = estimated adjusted group difference / comparison sample standard deviation) at the Type I error rate of 5%.

Kindergarten Reading Street students statistically significantly outperformed their peers using other reading programs on the GRADE by 0.38 standard deviations or 15% in percentile rank. They also statistically outperformed their peers on the GRADE Listening Comprehension subtest, DIBELS Phoneme Segmentation Fluency, and the DIBELS Nonsense Word Fluency. Kindergarten Reading Street students performed statistically similarly to the comparison group on the GRADE Sound Matching, Rhyming, and Phoneme-Grapheme Correspondence subtests combined (i.e., Word Study).

Kindergarten Scale	Effect Size[1,2]
GRADE Total	0.38
Word Study	***
Listening Comprehension	0.22
Phoneme Segmentation Fluency	0.45
Nonsense Word Fluency	0.15

*** Indicates group means are not statistically significantly different

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

Although not statistically significant, 1st grade Reading Street users had greater gains on the GRADE than their comparison group peers by 0.13 standard deviations or 5% in percentile rank. Reading Street 1st grade students statistically significantly outperformed their peers using other reading programs on the GRADE Listening Comprehension subtest and DIBELS Oral Reading Fluency test. First grade comparison group students statistically outscored the Reading Street students on the DIBELS Phoneme Segmentation Fluency and the DIBELS Nonsense Word Fluency. Reading Street students and their comparison peers performed similarly on the GRADE Word Study (i.e., Word Reading and Word Meaning subtests combined) and Reading Comprehension (i.e., Sentence Comprehension and Passage Comprehension subtests combined) subtests.

1st Grade Scale	Effect Size[1,2]
GRADE Total	***
Word Study	***
Reading Comprehension	***
Listening Comprehension	0.16
Phoneme Segmentation Fluency	-0.62
Nonsense Word Fluency	-0.46
Oral Reading Fluency	0.29

*** Indicates group means are not statistically significantly different

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

Fourth grade Reading Street students performed statistically similar to their peers on the GRADE total score (i.e., 0.10 standard deviations or 4% in percentile rank). They also performed statistically similar to the comparison group on the GRADE Word Study subtest (i.e., vocabulary). The 4th grade comparison group statistically significantly outperformed the Reading Street group on the GRADE Reading Comprehension subtest (i.e., Sentence Comprehension and Passage Comprehension) and the DIBELS Oral Reading Fluency scale.

4th Grade Scale	Effect Size[1,2]
GRADE Total	***
Word Study	***
Reading Comprehension	-0.35
Oral Reading Fluency	-0.12

*** Indicates group means are not statistically significantly different

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

Results were also broken out and analyzed for a group of lower achieving students at each grade level as well as separate levels of four key demographic variables (i.e., English proficiency, ethnicity, gender, meal status). Results for subpopulations of students were consistent with those seen for the whole sample. The results by subpopulation for the GRADE subtests and DIBELS are included in the full report.

RQ: Do Reading Street students demonstrate more positive attitudes toward reading and reading instruction than their peers using other elementary reading programs?

The reading academic attitude survey was administered to students to examine general reading attitude, confidence, motivation, self-perceived aptitude, vocabulary and comprehension, as well as, recreational and academic reading. Kindergarten and 1st grade Reading Street students had

higher reading academic attitudes than their comparison group peers, although only 1st grade was statistically significant (i.e., 0.34 standard deviations or 13% in percentile rank). Fourth grade Reading Street students had slightly lower reading academic attitudes than their comparison peers, but it was not statistically significant.

The ERAS (i.e., Garfield) survey was also administered to measure student recreational and academic reading attitudes. The ERAS survey was administered because of its wide usage and recognition in the education field. The results from this survey indicate Reading Street students had more positive attitudes than their comparison group peers at all grade levels, although only the 4th grade difference was statistically significant (i.e., 0.32 standard deviations or 13% in percentile rank).

I. INTRODUCTION

As schools strive to meet the adequate yearly progress goals in reading achievement, many are attempting to maximize their efforts by implementing innovative basal reading programs. Scott Foresman Reading Street © 2011[1] is one such program. Gatti Evaluation is partnering with Pearson to evaluate the Reading Street program in a longitudinal study during the 2009-10 and 2010-2011 school years. Information gathered during the study will provide evidence of curriculum efficacy and help to inform future revisions of the curriculum.

Pearson is partnering with Gatti Evaluation to study the efficacy of the Reading Street program in a longitudinal study during the 2009-10 and 2010-2011 school years.

This report provides methods and results from the first year of the efficacy research conducted during the 2009-10 school year on the Reading Street program, including; the study methodology, nuanced program usage information, teacher and administrator attitudes, as well as student attitudinal and achievement gains. The sample is comprised of students and teachers from 82 diverse elementary grade classrooms in nine schools distributed across six different states (i.e., AZ, CO, MA, MT, OH, and WA).

Reading Street is a pre-Kindergarten through 6th grade basal reading program based on the priority skills model. The priority skills model incorporates phonemic awareness, phonics, fluency, vocabulary, and comprehension in appropriate amounts as each beginning reader progresses through subsequent grades.[2] Because children approach text in various ways in accordance with their own abilities and purposes, reading instruction must be differentiated.[3] Differentiated instruction within the Reading Street Program ensures success for students of varying ability levels and experiences. Award winning reading selections seek to motivate students to learn, with a focus on developing a Big Idea in each unit along with science and social studies concepts. Reading Street also helps teachers achieve adequate yearly progress through integrated progress monitoring and assessment plans.

Reading Street is a pre-Kindergarten through 6th grade basal reading program based on the priority skills model, which incorporates phonemic awareness, phonics, fluency, vocabulary, and comprehension in appropriate amounts as each beginning reader progresses through subsequent grades.

Theoretically, research-based reading curricula can increase student reading achievement. Although a reading curriculum may be skillfully applied to create an educational environment that significantly increases achievement, poorly designed and implemented programs will provide little or no benefit, and may even be detrimental. Poorly designed and implemented curricula can confuse and frustrate students and teachers, proving to be a waste of valuable

[1] http://www.pearsonschool.com/
[2] http://www.nationalreadingpanel.org
[3] Lyon, G. R. (July 10, 1997). Report on Learning Disabilities Research. Testimony before the Committee on Education and the Workforce, U.S. House of Representatives.

resources and learning time. For these reasons, the No Child Left Behind Act[4] requires publishers to conduct rigorous efficacy research to support their educational materials.

Both the federal government and state adoption committees require publishers to conduct rigorous research to support the efficacy of their educational materials.

Study Goals And Research Questions

The primary goal of the current project was to conduct rigorous research to support the assertion that the Scott Foresman Reading Street ©2011 curriculum significantly increases students' reading achievement and attitudes in the first school year of implementation. The first year of implementation, when the program is new to both teachers and students, is the most challenging year for any new program to impact student achievement. The Reading Street program was tested against comparison classrooms randomly selected within each school that did not implement the Reading Street program. Rather, the comparison classrooms were not required to change reading curricula and teachers continued to utilize their current reading programs and favorite instructional practices. The secondary goal for this project was to collect information on teacher and student attitudes towards features and aspects of the Reading Street curriculum.

The research questions for this study are:

RQ1: Do students using the Reading Street curriculum demonstrate significant gains in reading/language arts achievement during the first year of curriculum implementation?

RQ2: How does reading achievement differ for students using the Reading Street curriculum as compared to their peers using other elementary reading curricula?

RQ3: Do Reading Street students demonstrate more positive attitudes toward reading and reading instruction than their peers using other elementary reading programs?

RQ4: How well was the Reading Street program implemented?

RQ5: How did teachers and students react to Reading Street?

[4] http://www.ed.gov/nclb

II. METHODOLOGY

The Reading Street © 2011 elementary basal reading program was evaluated in 82 diverse elementary grade classrooms in eight schools distributed across six different states (i.e., AZ, CO, MA, MT, OH, and WA) during the 2009-10 school year. The program was evaluated via a two-group, classroom level randomized, baseline to post observation assessment research design. Teachers or sections within each school were randomly assigned to one of the two study groups (i.e., comparison or. Reading Street). Students in classrooms randomly assigned to use Reading Street received the program for basal reading/language arts instruction while students in the comparison classrooms received reading/language arts instruction from district adopted programs and those materials and methods preferred by their classroom teachers.

The Reading Street efficacy study employed a two-group, classroom level, randomized design. Reading Street students received the program for basal reading/language arts instruction while students in the comparison classrooms received instruction from those materials and methods preferred by their classroom teachers.

Gatti Evaluation provided participating schools all data collection materials, maintained constant communication with study participants, and followed clear data collection procedures throughout the study to ensure that both study and program implementation ran smoothly and effectively.

The following sections provide information on study procedures, including; student and teacher level data collection, site recruitment and selection, the nature of reading instruction at the study sites, program training and implementation, detail on educational settings at each study site, demographic information for study participants, and the statistical methodologies used to analyze outcomes.

Student Outcome Measures

A challenging assessment battery was group administered to students to measure achievement and academic attitude growth during the school year.

An assessment battery comprised of the *Group Reading Assessment and Diagnostic Evaluation* (GRADE), an academic attitude survey, the Elementary Reading Attitude (i.e., ERAS or "Garfield") Survey, and the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) was used to measure gains in student achievement over the course of the school year. The assessment battery was intended to challenge the students; attempting to adequately assess baseline reading knowledge, while also providing room for growth as knowledge is acquired during the school year.

The GRADE, academic attitude survey, and ERAS survey were group-administered by the classroom teachers at the beginning (i.e., corresponding to initial training) and end of school year

(i.e., within four weeks of end of school). The DIBELS assessment was individually administered at the beginning, middle, and end of school year.

Group Reading Assessment and Diagnostic Evaluation (GRADE)

The GRADE is a standardized, nationally norm-referenced reading achievement test published by Pearson Assessment. The GRADE was constructed with all fifty states' standards in mind, covering a wide range of content topics and skills. The GRADE includes 11 levels that span grades preK-12, each with two parallel forms (i.e., level K for kindergarten, level 1 for 1st grade, level 4 for 4th grade). Form A was administered at baseline and form B was administered at the end of the school year. The GRADE is not a timed test, but generally takes between 70 and 100 minutes to administer. The level K test has 84 questions, the level 1 test has 107 questions, and the level 4 test is made up of 99 questions. Schools returned completed student tests to the site coordinators, who then shipped the tests to Gatti Evaluation for scoring.

Both GRADE overall and subtest scores were reported. The subtest scores allowed the research team to evaluate the effectiveness of the Reading Street curricula on important dimensions of reading/language arts acquisition. For level K, the subtests reported are; *Word Study* (42 questions) and *Listening Comprehension* (18 questions). The *Early Literacy Skills* subtests (i.e., *Print Awareness* 4 questions, *Letter Recognition* 11 questions, *Same and Different Words* 5 and 4 questions) do not assess either word study or comprehension skills but rather very basic visual and literacy skills. The *Early Literacy Skills* subtests are not reported separately but are part of the total kindergarten GRADE score. *Word Study* is comprised of three subsections: *Sound Matching* (12 questions), *Rhyming* (14 questions), and *Phoneme-Grapheme Correspondence* (16 questions). Early Literacy Skills is also comprised of three subsections: *Print Awareness* (4 questions), *Letter Recognition* (11 questions), and *Same & Different Words* (9 questions).

The Level 1 GRADE test is comprised of three subtests: *Word Study*, *Reading Comprehension* (43 questions), and *Listening Comprehension* (17 questions). For level 1, the *Word Study* (47 questions) subtest is further broken down into *Word Reading* (20 questions) and *Word Meaning* (27 questions), while the *Reading Comprehension* section is comprised of *Sentence Comprehension* (19 questions) and *Passage Comprehension* (24 questions). For level 4, *Word Study* is comprised of thirty-five vocabulary questions, and the *Reading Comprehension* section is also broken down into *Sentence Comprehension* (19 questions) and *Passage Comprehension* (28 questions). *Listening Comprehension* is not included in the 1st grade total GRADE scores, as it is intended to be a separate optional subtest.

Reading Academic Attitude Survey

The reading academic attitude survey was developed by the Gatti Evaluation Principal Investigator. Students responded to self-report questions regarding general reading attitude, confidence, motivation, and self-perceived aptitude. Questions also pertained to vocabulary and comprehension, as well as, recreational and academic reading. Student responses were coded as 1 for a positive response, 0 for a neutral response, and -1 for a negative response. This scoring method anchors a completely neutral student at an overall score of zero with positive total scores indicating an overall positive attitude. The kindergarten and 1st grade versions of the survey contained 16 items, and the 4th grade version of the survey contained 20 items.

Elementary Reading Attitude Survey (ERAS)

The ERAS survey, commonly knows as the "Garfield survey", was developed by McKenna and Kear[5] as a measure of reading attitude levels for students in grades 1 through 6. The instrument is comprised of 20 questions falling into two subcategories: recreational reading attitudes (10 questions) and academic reading attitudes (10 questions). The survey uses a 4-point Likert scale with anchored depictions of Garfield at each point. At the first point on the Likert scale, Garfield appears to be happy and excited. By the 4th point on the Likert scale, Garfield is hunched over and appears annoyed.

Dynamic Indicators of Basic Early Literacy Skills (DIBELS)

DIBELS is a tool used to measure key early literacy skills in students in kindergarten through 6th grade. These early literacy skills fall into five main categories: phonemic awareness, alphabetic principals, phonics, accurate and fluent reading, vocabulary, and comprehension. The measures can be used to monitor students' progress and identify students who need additional help in reading. Research studies have found evidence of a positive relationship between student DIBELS scores and reading success.

For the purposes of this study, the following measures were used: *Phoneme Segmentation Fluency* (PSF) to measure phonemic awareness, *Nonsense Word Fluency* (NWF) to examine alphabetic awareness, and *Oral Reading Fluency* (ORF), to examine accuracy, fluency, and phonics. The PSF and NSF measures can be used beginning at mid-year for kindergarten, through the end of 1st grade. Oral Reading Fluency can be used beginning mid-year with 1st grade students, through the 6th grade.

For kindergarten, the mid-year and end-of-year measures for PSF and NWF were analyzed. In 1st grade, the baseline and end-of-year scores were analyzed for PSF and NWF. The end-of-year ORF scores for 1st grade were also analyzed using baseline PSF and NSF scores as proxy baseline measures. In 4th grade, the baseline and end of year ORF measures were analyzed. Scale ranges and benchmark goals (ex., ORF score between 71-92 for 4th grade at months 1-3 equates to a "some risk" status) for DIBELS can be found online at dibels.uoregon.edu.

Teacher Measures

In addition to the assessment battery, qualitative data collection methods were also employed. The research team collected qualitative data through self-report teacher logs and classroom observations, as well as teacher interviews and focus groups. The data was compiled and content analyzed to examine teacher attitudes, pedagogy and performance, as well as to illuminate the various ways teachers and students interact with the Reading Street program. The teacher and classroom data also increased the validity of the research findings by verifying results through

[5] McKenna, M. C, & Kear, D. J. (1990). Measuring attitude toward reading: A new tool for teachers. The Reading Teacher, 43, 626-639.

multiple data collection methods, by adding context to the achievement results through reporting the perspectives of various study participants, and by collecting data throughout the project period. Continuous monitoring of the study sites was of immense importance, and teachers were routinely asked for their opinions and concerns throughout the school year.

The research team collected achievement, attitudinal, as well as, observational and self-report data making the study both quantitative and qualitative in nature.

Weekly Teacher Logs

All study teachers were required to complete weekly self-report online logs in which they described their reading lessons. Information from the weekly logs was important for two reasons; to guarantee Reading Street teachers fully and regularly utilized all key components of the program in an attempt to positively influence student achievement, and to document the instructional model utilized by each study teacher, including classroom environment, teaching style, pacing and reading content and methods.

Teachers were asked not to spend more than 15 minutes per week completing the logs. It is clear several teachers spent more time, however, as many of the logs were returned with detailed comments. Teachers often shared candid weekly experiences with the project manager and were typically happy to provide documentation describing weekly instruction and learning experiences related to the program. Comparison group teachers summarized daily classroom reading instruction time, topics and methods. Reading Street teachers were also asked to provide usage information for each major component of the program.

Teacher Observations

Site visits took place between mid-October and early December, and again between April and mid-May. Classroom observations were conducted by the research team. All study classrooms from each school were observed at least once during routine reading/ELA lessons. Portions of the observation forms included; a description of the classroom environment, summary of the lesson taught, teacher interviews, student comments, observed teaching strengths and weaknesses, pacing, and supplemental instruction information. The observations also allowed researchers to observe general classroom environment and teaching styles, and to verify the ability and willingness of Reading Street teachers to properly implement the program for reading instruction.

It should be noted that two observations show just a snapshot of the classroom environment and instructional competence. Some teachers were required to change their normal class time due to scheduling conflicts, which occasionally resulted in the observer having less than optimal time to spend in the classroom. The observations are, however, worthwhile because they are the only opportunity the research team has to directly observe the study teachers in action and verify teacher reported information. It should also be noted that the Reading Street consultants had opportunities to observe most teachers using the Reading Street program during the follow-up training visits. Teachers that were not returning their weekly implementation logs, missing training sessions, or generally perceived as struggling with their implementation were prioritized for observation by the Reading Street consultants so they could provide support and assistance.

Teacher Surveys

All participating teachers were administered two surveys about their teaching background; a baseline paper-based survey administered at the study orientation, and an end-of-year online survey. The purpose of the baseline teacher survey was to collect information on teaching experience, reading curricula, and prior research study experience. Teachers were asked to indicate their highest level of education and the number of years teaching total, as well as years they had spent at their district, school, and grade level.

The end-of-year teacher survey focused on gathering details about school context, teaching philosophy, and reading curriculum implementation. Teachers were asked about their curriculum materials, technology usage, and teaching strategies. Teachers were also asked to describe ways in which their school and students are unique. All of this information allowed researchers to gain additional insight into the overall experience at each research site.

Reading Street Teacher Focus Group

Focus groups were executed by the research team to ascertain teacher attitudes toward the Reading Street program. The face-to-face nature of a focus group, though more labor intensive, can be superior to simple questionnaires in collecting detailed attitudinal information from participants. When properly conducted, the focus group discussion gravitates to those topics most important to the participants, and can provide more nuanced information. Collecting attitudinal data in person allows for a better understanding of participant tone and importance of responses, and provides opportunity to delve deeper into topics.

Focus group sessions were conducted at each school during site visits between April and mid-May. Representatives from the research team facilitated each session. The sessions lasted approximately 60 minutes. Forty-two of the 43 Reading Street teachers participated in the focus group sessions. One teacher who could not participate in the focus group session sent in responses to the focus group questions electronically.

These sessions provided a forum for teachers and administrators to respond to specific questions about the Reading Street curriculum, as well as express their professional and personal opinions about the curriculum. Each session held the teachers' comfort level as a high priority. The teachers were encouraged to speak without hesitation or inhibition, and to be as honest and candid as possible. Though the facilitator followed a structured interview format, the teachers were allowed to direct the discussion and provide their reactions to, and comment on, any and all aspects of the program. The focus group sessions provided extensive insight into teacher and student experiences with, and attitudes about, the Reading Street program. This information was supplemented with opinions informally shared by students during the observations.

Extensive notes were taken at each focus group session, allowing the research team to compile a large master file of participant responses. Following an exhaustive review of the teacher responses, a two-dimensional coding system was developed to organize the responses. Responses were categorized by *Topic Area* and *Attitude*. Topic area codes have a three digit numeric format, with the first digit on the left indicating general topic category and the remaining digits indicating a specific topic within each general category. The topic codes are further categorized by grade level, study site, and paired with either an 'N' to indicate neutral, a '+' to indicate positive, or a '–' to indicate a negative attitude toward an aspect of the program or the tone of the comment.

The focus group results describe what teachers and students liked about the Reading Street program, how the program could be improved, and how teachers are using specific features of the program.

Site Recruitment and Selection

Potential research schools were identified by Pearson sales representatives and via email blasts sent to districts with specific demographics. Schools that indicated interest were sent a study description that included responsibilities and incentives. Possible research schools were further vetted through local sales representatives. If the school indicated interest after reviewing the study description and being approved by the sales representative, they were asked to complete a detailed questionnaire. The intent of the questionnaire was to ensure participants understood all the requirements and benefits associated with participation. It was required that schools did not currently use the Reading Street program, that all participating teachers abide by the random assignment, and that all randomly assigned Reading Street classroom teachers fully implement the program with their students.

Table 1 Gatti Evaluation Reading Street RCT Site State Assessment Information

					School Results	State Wide Results
School Year	Grade	State	School	Met AYP	Meets Reading Standards	Meets Reading Standards
2008-09	4	AZ	1	Yes	59% (-15%)	74%
2008-09	4	CO	2	Yes	62% (-3%)	65%
2008-09	4	CO	3	No	68% (+3%)	65%
2008-09	4	MA	4	Yes	77% (+24)	53%
2008-09	4	MA	5	Yes	60% (+7%)	53%
2008-09	4	MT	6	***	***	81%
2008-09	4	MT	7	***	***	81%
2008-09	4	OH	8	Yes	86% (+14)	72%
2008-09	4	WA	9	Yes	83% (+9%)	74%

Parentheses indicate comparison to state percent meeting reading standards

*** Information not available.

After the questionnaire was reviewed and approved by the Principal Investigator, the school was invited to be a study participant. Finally, both a district level administrator (ex., curriculum director, superintendent) and a school level administrator (ex., principal) signed a memorandum of understanding outlining the responsibilities of each stakeholder. No available students of any socio-economic level, English proficiency level, or ethnic background, who opted to participate

in the study, were excluded from the study. The research team adhered to the informed consent requirements of each participating school and/or district.

The final study sample was comprised of schools from public school districts located in mid- to large cities, suburban, or rural fringe areas. One school from each of Arizona, Ohio, and Washington states participated in the study. Three schools from two districts in Massachusetts, two schools from two different districts in Colorado, and two schools from the same district in Montana also participated in the study.

Ethnic and socio-economic diversity among the student population were two criteria the evaluation team considered when recruiting study sites. A third criterion was that students exhibit a wide range of ability with respect to reading achievement. Table 1 shows, according to recent state achievement testing data, the percent of each school's students meeting state reading standards range between 15% below to 24% above statewide results. The evaluation team sought out diversity in the study sample to ensure the program would be used by learners of all abilities and backgrounds, thus reflecting the reality that is today's elementary classrooms.

Reading Instruction

Teachers assigned to the comparison condition were expected to implement the reading curricula currently being used in their school and/or district. Four published reading/ELA programs were used by the research schools at kindergarten, six programs at 1st grade, and seven programs at 4th grade. A similar proportion of kindergarten and 1st grade comparison students had teachers that reported primarily implementing a teacher-created reading program (kindergarten comparison = 50%, 1st grade comparison = 48%), while relatively fewer students received a teacher-created reading program at 4th grade (4th comparison = 27%). Instead, a widely published program that was not from the publisher of Reading Street was used with 28% of the 4th grade students. Comparison teachers at each grade level had an average of four years using their current reading program.

A majority of students (i.e., Reading Street = 72%, comparison = 74%) attended classrooms where there was a person regularly assisting the classroom teacher (i.e., teacher's aide, paraprofessional, student teacher, reading coach, etc.). Further, a majority of students had teachers that explained their teaching philosophy as balanced literacy as opposed to phonics or whole language (i.e., Reading Street = 97%, comparison = 84%), and received an average of 95 minutes of daily reading instruction (Reading Street = 93 minutes, comparison = 96 minutes). More kindergarten and 4th grade comparison students attended classrooms with extra assistance; the opposite was true for 1st grade students. Also, 1st grade students received the most reading instruction, and 1st grade comparison group students received an average of six more minutes of reading instruction than their Reading Street peers.

There was approximately one year or less difference between the study conditions in teaching experience at 1st and 4th grade (i.e., 4th grade comparison teachers had a year more experience at that grade level). The difference in teaching experience at the kindergarten level was more pronounced. Kindergarten Reading Street students had teachers with more teaching experience, 3.14 more years teaching overall and 6.46 more years teaching kindergarten.

The 1st and 4th grade comparison sample had a higher portion of students taught by a teacher with a Master's degree (i.e., 1st comparison 76% or 14% higher than Reading Street; 4th comparison 75% or 21% higher than Reading Street). The kindergarten sample had overall fewer students taught by a teacher with a Master's degree. Comparison teachers at each level had an average of four years using their current reading program. It should be noted that the statistical analyses adjusted for these, as well as other, factors known to impact the outcomes of interest.

Kindergarten	Reading Street	comparison
years teaching	17.73	14.59
years at current grade	12.43	5.97
master's degree	37%	39%
years using current program	6.77	3.99
minutes reading instruction	90.11	91.63
regular classroom assistance	68%	81%

Difference in years teaching was statistically significant.
Difference in years teaching at current grade level was statistically significant.
Difference in years using current reading/ELA program was statistically significant.
Difference in percent of students in classrooms receiving assistance was statistically significant.

1st Grade	Reading Street	comparison
years teaching	15.12	15.60
years at current grade	8.23	7.83
master's degree	62%	76%
years using current program	7.70	4.19
minutes reading instruction	104.44	110.39
regular classroom assistance	79%	64%

Difference in percent of teachers with master's degree was statistically significant.
Difference in years using current reading/ELA program was statistically significant.
Difference in minutes of reading instruction was statistically significant.
Difference in percent of students in classrooms receiving assistance was statistically significant.

4th Grade	Reading Street	comparison
years teaching	13.07	13.06
years at current grade	5.21	6.25
master's degree	54%	75%
years using current program	6.76	4.35
minutes reading instruction	85.13	86.90

regular classroom assistance	69%	76%

Difference in years teaching current grade level was statistically significant.
Difference in percent of teachers with master's degree was statistically significant.
Difference in years using current reading/ELA program was statistically significant.

Reading Street Implementation

Teachers received multiple training sessions by Pearson curriculum specialists. The trainings allowed teachers to fully implement the program and fostered positive teacher and student attitudes.

Reading Street Teacher Training

To initiate the study, Gatti Evaluation representatives conducted study orientations for all teachers at the start of the school year. The study orientation formally introduced the teachers to the research team, explained in detail the requirements and benefits of participation in the study, as well as, addressed any immediate questions or concerns about the research. All teachers were required to read and sign informed consent forms.

Pearson provided free product training and funding to cover the cost of substitute teachers during training. All Reading Street teachers, including the 2nd and 5th grade teachers who would participate in the second year (2010-11) of the research study, were required to attend training sessions facilitated by a curriculum specialist. The first curriculum training took place on-site over the course of two full school days. This training introduced administrators and teachers to the key components and instructional features of the Reading Street curriculum, including whole and small group instruction, and progress monitoring. Three follow-up training sessions were further provided to each school to support consistent usage and implementation fidelity of the Reading Street curriculum, and to acquaint teachers with new additions to the curriculum since some components were not available at their first training session. For training dates by site, please see Table 2.

Initial curriculum training sessions lasted a full two school days and typically began with a group presentation. Then teachers were separated into two groups by grade, primary (kindergarten-2nd grade) and secondary (4th and 5th grade), to get more personalized grade-level training. Subsequent training sessions coupled in-classroom observations with one-on-one meetings between teachers and consultants, and typically lasted one school day. After each visit the trainers provided feedback on each teacher's performance including strengths and weaknesses on program components, adherence to the study's implementation guidelines, and flagging those teachers in need of extra attention.

The trainings were well-received. The research team strongly believes that ongoing professional development can significantly affect the potential for a program such as Reading Street to foster positive teacher and student attitudes, meet students' needs, and ultimately increase student achievement.

Table 2			Gatti Evaluation Reading Street RCT Training Dates			
State	District	School	School Start Date	Initial Training Date	Follow-up Training Date	Additional Trainings
AZ	1	1	08/03/09	08/04/09	09/29/09	11/17/09 & 02/26/10
CO	1	1	08/13/09	08/11/09	10/0509	11/20/09
CO	2	1	08/24/09	08/17/09	10/26/09	01/11/10 & 03/16/10
MA	1	1	09/02/09	08/25/09	10/19/09	12/08/09 & 02/23/10
MA	1	1	09/01/09	08/20/09	10/21/09	12/09/09 & 02/24/10
MT	1	1	08/26/09	08/19/09	10/05/09	12/01/09 & 03/03/10
MT	1	2	08/26/09	08/19/09	10/05/09	12/01/09 & 03/03/10
OH	1	1	08/25/09	08/18/09	10/13/09	12/11/09 & 02/17/09
WA	1	1	09/01/09	09/02/09	11/12/09	12/15/09 & 02/19/10

Reading Street Curriculum Usage

This sub-section provides evidence in response to research question four, or

How well was the Reading Street program implemented?

Pearson ensured that research schools had full access to all Reading Street components that were available. It is worth noting the Reading Street ©2011 program was not released for public sale until January 2010. Therefore, some ancillary components of the Reading Street program were still under development when the study was launched in August 2009. As ancillary components became available, they were shipped to the research schools. Teachers assigned to use the Reading Street curriculum were asked to implement the core Reading Street components (i.e., Teacher Edition, Student Edition) to the best of their ability while awaiting ancillary materials. For example, while some of the posters and big books were not available, teachers simply made their own. It should be noted that the Reading Street practice stations were not completed by the end of the school year, and thus not used by the year one study teachers. Teachers created their own practice stations to go with the program.

Several sites were very concerned with strictly implementing the writing portions of the program and required extra training and assurance of how the writing portions could be applied to adhere with their state standards. Implementation of the writing portions and practice stations will be strictly adhered to in the second year of the study.

Reading Street teachers unanimously reported that instruction felt natural to them by winter break. A few months of use and discovery were necessary for the teachers to become fully comfortable with pacing and flow of the program. The same was true for their students. It took them a couple months to adapt to the structured routine offered by the Reading Street program.

At the fourth and last training session (i.e., February and March) the trainers rated each of their teachers' overall performance implementing the program with regard to the study guidelines on a 1-10 scale. A rating of one was reserved for those teachers that were putting forth no effort; no

teachers received a rating of one. A ten was reserved for those teachers using all the required components with distinction and proper pacing. These teachers represent the best implementing teachers the trainers have seen, in essence experts. No teachers received a rating of ten.

A rating of five 5 was applied to those teachers trying their best to implement the program but were still having some problems implementing the required components and/or pacing. These teachers could be described as average implementers with seven to nine months using the Reading Street program. Twelve percent of teachers performed at or below a rating of five.

The ratings three and seven acted as additional anchor points. A rating of three was applied to those teachers trying their best to implement the program but were still having substantial problems implementing the required components and/or pacing. One teacher received a rating of three and this was the lowest rating given out. Conversely a rating of seven was given to those teachers that were implementing the required components well with good pacing. Seventy-seven percent of teachers received a rating of seven or higher.

The average ratings for overall implementation were above average, indicating the teachers were ultimately implementing the program well with respect to the study guidelines (i.e., kindergarten = 7.4, 1st grade = 7.4, 4th grade = 7.3). The average ratings were consistent across grade levels with the standard deviations for the ratings ranging from 1.1 to 1.5. The variance components for the raters and grade levels were negligible or essentially zero. Ninety-five percent of the variation in the ratings is due to the sites and teachers. This means that the average implementation for each site and the implementation for each teacher accounts for the bulk of what makes the ratings differ, not a teacher's grade level or the trainer that rated the teacher.

The trainers also asked each of their teachers to rate themselves from 1-10 on how well they felt they were implementing the program and following the study guidelines. The trainers' rating and teachers' self ratings correlated moderate to high with a Pearson correlation coefficient of 0.55. The sample means for the two ratings do not differ statistically and practically do not differ at all (i.e., average trainer rating = 7.22, average teacher self rating = 7.31)

Average Reading Street teacher implementation ratings were consistent across grades and above average, 7.4 out of a possible 10.

The Reading Street program is comprised of six units. Each unit is comprised of two volumes, each with three weeks worth of curriculum content, totaling six weeks of content per unit. Pacing varied across the Reading Street teachers. Kindergarten teachers tended to complete more of the program than the other grades, with 27% reaching unit six by end-of-year testing, and the remaining teachers reaching unit five. Thirty-eight percent of 1st grade teachers reached unit five, while 50% reached unit four and 12% reached unit three. Fourth grade teachers were the most varied, with 15% reaching unit six, 39% reaching unit five, 15% reaching four, and 31% reaching unit three. Reading Street teachers used the program (i.e., Get Read to Read, Read and Comprehend, Language Arts, Wrap up your Week, and small group instruction) for an average of 7.65 hours per week (i.e., 94 minutes per day).

As for the major components of the program, teachers instructed from the Get Ready to Read and Read and Comprehend sections the most, three to four days a week. After these, teachers instructed from the Language Arts section and lastly the Wrap up your Week section. Reading

Street teachers were asked to record in their weekly logs their impression of the students' engagement as a group while using the program. On average, teachers reported high classroom engagement for 28% of the lessons, average engagement 54% of the time, and low engagement only 19% of the time. These results are further validated by the fact that Reading Street teachers were overwhelmingly positive about their students' interactions with the program. When interviewed, 80% of teachers' comments were positive in nature.

Settings

This section summarizes the educational model and environment for each study site, as well as a demographic breakdown. This information is crucial for determining how applicable results from this study may be to the consumers of this report.

Arizona District

One school from this district participated in the study. This participating Arizona school is located in a rural fringe area of a large city. In 2008-09, the school district served a community of 10,000. The median household income is approximately $50,000, indicating a middle class community. The school is mid- to large-sized, serving approximately 700 students in grades kindergarten through six. There is a strong sense of camaraderie and teamwork amongst the teachers at this school, and they strive to give their students the most supportive learning environment possible. The site is a Title 1 school with a great deal of learning diversity and high Hispanic (i.e., 70%) and ELL (i.e., 25%) populations. Many of the students come from low-income families and the school follows a strict dress code.

This school falls into the high range for participation in the nation's free or reduced-price lunch program, with 85% of students eligible to receive free/reduced price lunch. The school did meet AYP in the 2008-09 school year. The percentage of 4th grade students testing at standard in reading in the 2008-09 school year was 59%, lower than the statewide results by 15%. The student/teacher ratio is approximately 23 to 1.

Twelve teachers participated in the study, three kindergarten teachers and two teachers from each of 1st, 2nd, 4th, and 5th grades. In kindergarten, two teachers were randomly assigned to use Reading Street. In 1st, 2nd, 4th and 5th grade, one of the two teachers from each grade was randomly assigned to the Reading Street. Three of the comparison teachers primarily adhere to the district-adopted basal program with some supplementation, while two heavily supplement the district-adopted program. Teachers have used this basal reading program for an average of 3.4 years. Three of the comparison teachers reported having some training on the curriculum, one at a previous district. The other two did not receive any formal training on the program. One teacher was new to the school for the 2009-2010 school year, while two were new to their grade level.

Teachers' daily reading blocks range from 90 to 180 minutes. They conduct their reading lessons using whole group an average of 58% of the time. All comparison teachers incorporate leveled instruction, cooperative learning, independent reading, guided reading, and seat work into their literacy instruction. All but one of these teachers also incorporate centers and test prep for AIMS (Arizona's Instrument to Measure Standards) into their reading instruction.

The comparison teachers monitor their students for progress one to three times per week. Most of the teachers prefer using a balanced literacy approach to their reading instruction. However, one teacher prefers a whole language approach and two prefer a phonics approach. In terms of instructional materials, all teachers prefer using a combination of trade books and basal readers to using one over the other. Some teachers also incorporate technology into their reading instruction. Five of the 12 teachers reported using a variety of websites and educational computer programs, while two reported using interactive whiteboards. Two teachers had teaching aides in the classroom, though one only had the aide for about 40 minutes per week. Another teacher had some isolated help from an ELL teacher, totaling to about eight sessions during the school year.

Colorado District One

One school from this district participated in the study. The first thing you will notice walking into this Colorado elementary school is a two-story art installation made by teachers, students, and volunteers. Past installations have included representations of classic children's literature *The BFG* and *Where the Wild Things Are*. This exemplifies the effort that the teachers put into making learning a positive and exciting experience for their students. As this school is located in a largely military community, many students have family members that are deployed abroad. The student population is very diverse, and there is high mobility.

The school resides in a large suburb. In 2008-2009, the school district served a community of 20,000. The median household income is approximately $50,000, indicating a middle class community. The school is mid-sized, serving approximately 600 students in grades pre-kindergarten through five. The majority of the students are Caucasian (i.e., 50%), followed by African-American and Hispanic (i.e., 20% each). This school falls into the medium-high range for participation in the nation's free/reduced price lunch program with 52% of students eligible. The elementary school did meet AYP in the 2008-09 school year. The percentage of 4th grade students testing at standard in reading in 2008-09 was 62%, 3% lower than the statewide results. The student/teacher ratio is approximately 23 to 1.

Twenty teachers participated in the study, four kindergarten teachers, five 1st grade teachers, four 2nd grade teachers, four 4th grade teachers, and three 5th grade teachers. Originally two kindergarten teachers were randomly assigned to use the Reading Street program, and one kindergarten teacher was assigned to the comparison group. Shortly thereafter, the school hired a new kindergarten teacher who began using Reading Street, and thus was added to the treatment group. In November, one of the Reading Street kindergarten teachers was transferred to 3rd grade, and a new teacher took her place as a Reading Street kindergarten teacher. This teacher was trained immediately on the program. Two Reading Street teachers and one comparison teacher taught half-day kindergarten, and the other Reading Street teacher taught full-day kindergarten.

Three 1st grade teachers were randomly assigned to the Reading Street group, while the other two 1st grade teachers were assigned to the comparison group. Two 2nd grade and two 4th grade teachers were randomly assigned to the Reading Street group. Finally, in 5th grade, two teachers were randomly assigned to the Reading Street group, and one was assigned to the comparison group.

The district adopted a widely-published, elementary basal reading curriculum with a copyright date later than 2005. One comparison group teacher reported adhering to the district curriculum with some supplementation, while two comparison group teachers reported using the reading curriculum with heavy supplementation and another two reported primarily using supplemental materials. Two of the comparison group teachers use their own teacher-made programs. This information was not available for one of the comparison group teachers. Teachers using the district-adopted curriculum had an average usage of about three years, though most report receiving no formal training.

Participating kindergarten teachers in this district have an average of 15.5 years teaching. First grade teachers had an average of 7.8 years, 2nd grade teachers an average of 7.25, 4th grade teachers an average of 3.7 years, and 5th grade teachers, an average of 4.5 years. Four of the 20 teachers were new to this school, and four teachers were new to their grade level during the 2009-2010 school year.

Teachers have daily reading blocks of 60 to 120 minutes. The comparison group teachers frequently use a number of teaching strategies during reading instruction, including; leveled instruction, peer tutoring, centers, independent reading, seatwork, and test prep. Four teachers monitor their students for progress weekly, while three monitor for progress two to three times per week or more. Reading is taught using whole group instruction about 54% of the time.

Out of 20 teachers, 19 completed their end-of-year teacher survey, which inquired about teaching philosophy and preferred literacy materials. Out of the 19 of teachers providing information, 17 prefer using a balanced literacy approach to teaching reading, and two prefer a phonics approach. All teachers prefer to use a combination of basal readers and trade books, rather than just one or the other. Teachers also like using websites and educational computer games in the classroom, though only a few teachers use interactive white boards. Less than half of the teachers receive additional support in the classroom in the form of parent helpers, paraprofessionals, Title 1 and SPED staff, and older students.

Colorado District Two

Teachers at this Colorado school describe their students as coming from a range of backgrounds with little to moderate parental involvement. Teachers have seen an increase in emotional and behavioral problems with their students over the last year. Despite this, teachers are proud of the hurdles their students overcome to make strides in their learning.

This school resides in a large suburb. In 2008-2009, the school district served a community of 40,000. The median household income is approximately $80,000, indicating an upper-middle class community. The elementary school is mid- to small-sized, serving approximately 350 students in grades kindergarten through six. The school has one primary ethnic group, Caucasian, accounting for 83% of the school population. This school falls into the medium range for participation in the nation's free/reduced price lunch program, with 26% of students eligible and has a small ELL population. They did not meet AYP in the 2008-09 school year. The percentage of 4th grade students testing at standard in reading in the 2008-09 school year was 68%, higher than the statewide results by 3%. The student/teacher ratio is approximately 16 to 1.

A total of 10 teachers participated in the Reading Street study, two in each grade level (i.e., k, 1st, 2nd, 4th and 5th), of which one was randomly assigned to use Reading Street. Midway through the year, the research team was alerted the 2nd grade Reading Street teacher was going to retire at the end of the year. In order to ensure that there was a 2nd grade teacher trained on the program who would be prepared to teach Reading Street the following year, the comparison teacher was moved to the Reading Street group as well. A new hire in the second year of the study will become the 2nd grade comparison group teacher. One 4th grade teacher was new to the grade level, having taught 5th grade in years past.

Participating teachers were surveyed on their use of curricular materials, teaching strategies, and teaching philosophies. A total of eight out of ten teachers (e.g. three comparison teachers and five treatment teachers) responded to this survey. Comparison teachers continued to use their previously chosen materials for reading instruction during the study. There was a large range for how long teachers have been following their current programs, from two to 25 years. Comparison teachers use whole group instruction about 58% of the time. They frequently use centers, peer tutoring, seatwork, leveled instruction, independent reading, and guided reading in their literacy instruction. Teachers also reported sometimes using test prep strategies as well. Out of three current year comparison teachers, one teacher monitors students for progress four or five days per week, while the other two monitor their students' progress once per week.

This school district does not mandate the use of one particular reading curriculum. Rather, the district provides essential standards that all teachers must include; this helps guide teachers in their choices of materials. Daily reading blocks range from 105 minutes to 150 minutes. Six of the teachers prefer using a balanced-literacy approach to teaching reading, while one prefers a phonics approach, and one prefers a whole language approach. Seven teachers prefer to use a combination of trade books and basal readers for reading instruction, and one prefers to just use trade books. A few teachers use digital white boards and educational computer games frequently to enhance reading instruction. Websites are used by some teachers, but not very often. Four of the teachers have paraprofessionals to support them in the classroom, while another has a parent helper and a coach. Another teacher receives support from two graduate students in addition to the paraprofessional.

Massachusetts District One

Two schools participated from this district, an elementary and a middle school. The schools are mostly Caucasian (i.e., 95% and 97%) and the majority of students come from middle to high-middle class backgrounds. Parents and teachers are very supportive and have high expectations for their students' education. Pre-school is a high priority in this community, and many kindergarten students are not new to the school environment. The biggest challenge teachers face is continuing to motivate all of their students when teaching to multiple ability levels in the classroom.

Both schools in this Massachusetts district reside in a large suburb. In 2008-2009, the school district served a community of 10,000. The median household income is approximately $100,000 indicating an upper-class community. The elementary school is a medium-large sized school serving approximately 600 students in grades kindergarten through four. Few students, only 3%, are eligible for participation in the nation's free or reduced-price lunch program. The middle school is a large school serving approximately 900 students in grades five through eight.

This school also falls into the low range for participation in the nation's free or reduced-price lunch program, with 5% of students eligible.

The elementary school did meet AYP in the 2008-09 school year. The percentage of 4th grade students testing at standard in reading in the 2008-09 school year was 77%, higher than the statewide results by 24%. The student/teacher ratio is approximately 16 to 1. Students in 4th grade will move to the middle school in the second year of the study.

A total of 22 teachers participated in the study, four kindergarten teachers, four 1st grade teachers, two 1st/2nd combination class teachers, four 4th grade teachers, and four 5th grade teachers. In kindergarten, two teachers were randomly assigned to use the Reading Street program. One of these teachers had two sections of two-and-a-half day kindergarten, and so used the program with both classes. In 1st, 2nd, 4th, and 5th grade, two teachers per grade were randomly assigned to use the Reading Street program, as was one of the 1st/2nd combination classrooms.

Participating kindergarten teachers had an average of 19.75 years teaching, first grade had an average of 16.5 years, 2nd grade had an average of 19.25 years, 4th grade teachers had an average of 11 years, and 5th grade teachers had an average of 21 years. One teacher at the elementary school was new to the district. Daily elementary reading blocks were 90 minutes, while 5th grade ELA blocks were 60 minutes.

At the time of the study, intermediate classrooms (i.e., 4th and 5th grade) in the comparison group were using a newly published basal reading curriculum for the first time. Other comparison teachers continued to use materials they had used in the past. Prior to this school year, the elementary school teachers created their own programs from a collective of literacy materials. At the middle school, the comparison group teachers adhered strongly to the recently adopted program, with some supplementation.

Elementary school comparison teachers use whole group instruction an average of 36% of the time, and middle school comparison teachers use whole group instruction an average of 50% of the time. Most teachers have some form of additional support in the classroom, with only one elementary teacher and one middle school teacher not receiving additional support. Elementary comparison teachers frequently use leveled instruction, independent reading, guided reading, and seat work strategies. They also use test-prep strategies, though not very frequently, and many also occasionally use center activities and cooperative learning.

Comparison teachers at the middle school incorporate all of the above mentioned strategies into their reading instruction as well. Three of the elementary school comparison teachers prefer to monitor their students for progress daily, while one teacher monitors students two or three times a week, and two teachers once per week. The final two comparison group teachers at the elementary school report monitoring students for progress less than once per week. At the middle school, one teacher monitors students for progress once per week, and the other two to three times per week.

The teachers at both schools prefer a balanced literacy approach to reading instruction. One teacher at the elementary school prefers basal readers and eight teachers prefer using trade books. The rest of the teachers at both schools prefer using a combination of basal readers and

trade books. Most elementary teachers use websites and educational computer programs as a part of their reading instruction. At the middle school, educational websites are the most popular form of supplemental technology, with two of the four teachers reporting frequent usage. Computer games are also used by most, though infrequently.

Massachusetts District Two

This district had two participating schools, and is located in a large suburb. The primary school serves pre-kindergarten through 2nd grade while the intermediate school serves 3rd through 5th grade. Though the schools have separate administrations, they are connected by a common hallway and collaborate closely with one another. In 2008-2009, the schools served a community of 14,000. The median household income is approximately $100,000, indicating an upper class community.

The schools provided various learning opportunities for their students. In addition to traditional classrooms, there was a French Immersion program beginning in kindergarten, a Montessori program, and Spanish language available in the first grade. Classroom composition varied, including students at different learning levels such as gifted, on-level, ELL, and IEP. These were the only elementary schools for the district, therefore were quite large. They enjoy a good deal of support from parents and the community.

The elementary schools in Massachusetts are large and serve approximately 1,300 students in grades pre-kindergarten through five. The schools have one primary ethnic group, Caucasian, representing 94% of the school population. Only 3% of the students are eligible for free or reduced-price lunch. The primary school did meet AYP in the 2008-09 school year. Data was unavailable for the intermediate school. The percentage of 4th grade students testing at standard in English Language Arts for the 2008-09 school year was 60%, 7% higher than the statewide results. The student/teacher ratio at the primary school is approximately 13 to 1. It is 15 to 1 at the intermediate school.

A total of 26 teachers participated in the study from these schools, four kindergarten, six teachers each in 1st, 2nd, and 4th grades, and four 5th grade teachers. Two kindergarten teachers were randomly assigned to use Reading Street. In 1st, 2nd, and 4th grade, three of the six teachers at each grade level were randomly assigned to use Reading Street. Finally, two 5th grade teachers were randomly assigned to the Reading Street group. The kindergarten teachers have been teaching for an average of 16.75 years, 1st grade teachers for an average of 16 years, 2nd grade teachers for an average of 12.6 years, 4th grade teachers for an average of 18.2 years, and 5th grade teachers for an average of 14.7 years.

Daily reading blocks range from 60 to 120 minutes. Most teachers use and have received training on guided reading, and the Developmental Reading Assessment (DRA). The majority of the teachers, 15 of 26, have a paraprofessional in their classroom for support. The rest have a student teacher, teacher's aide, parent helper, or SPED teacher. Teachers at both schools prefer a balanced literacy approach to reading instruction, and prefer using a combination of basal readers and trade books rather than just one or the other. Most use websites and educational computer programs at least once per week as part of their reading instruction.

The district adopted a widely published elementary basal reading curriculum for the intermediate school. At the primary school, teachers mostly create their own programs that coincide with

their balanced literacy approach. The comparison group teachers have been following this approach for an average of six years. The comparison teachers use whole group instruction an average of 42% of the time. Comparison teachers frequently use a number of teaching strategies including; leveled instruction, independent reading, guided reading, centers, and seatwork. Test-prep is the least commonly used strategy, and three of the teachers do not use it at all. All comparison teachers monitor their students' progress at least weekly, while four monitor their students two to three times per week, and two teachers monitor their students four to five times per week.

Montana District

Two separate elementary schools participated from the Montana district. The first elementary school was built in the 1970's with no walls in between classrooms in what is called an "open school" structure. Students come from a variety of backgrounds, and parents are active and involved. Teachers work hard in support of their students and think of the school as a family. The second elementary school also reports having very involved and supportive parents. Teachers and staff practice a teamwork approach, and there is a large emphasis on respect and good manners. These schools have one primary ethnic group, Caucasian, representing a total of 86% of the school population. Both schools met AYP for the 2008-2009 school year.

The first participating school resides in a midsize city. In 2008-2009, the school served a community of 25,000. The median household income is approximately $60,000 indicating a middle class community. The school is small, serving approximately 350 students in grades pre-kindergarten through six. Approximately 20% of students are eligible to receive free or reduced-price lunch. The student/teacher ratio is approximately 15 to 1.

The second participating school also resides in a midsize city. In 2008-2009, the school served a community of 50,000. The median household income is approximately $55,000, also indicating a middle class community. This school is a small to medium sized, serving approximately 400 students in grades pre-kindergarten through six. Approximately 30% of students are eligible to receive free or reduced-price lunch. The student/teacher ratio is approximately 12 to 1.

Eleven teachers from the first school participated in the study; two kindergarten 1st, 4th, and 5th grade teachers, and three 2nd grade teachers. One teacher was randomly assigned to use Reading Street at each grade level. One teacher was new to the school, and three were new to their grade level. Thirteen teachers from the second school participated in the study; three teachers from kindergarten, 1st, and 2nd grade, and two teachers from 4th and 5th grade. In kindergarten, 1st, and 2nd grade, two of the participating teachers at each grade were randomly assigned to use Reading Street. At 4th and 5th grade, one teacher per grade level was randomly assigned to the Reading Street group. One study teacher was new to the school, and two were new to their grade level.

The primary grades at both schools use a widely published elementary basal reading program with a copyright date prior to 2005. In 4th and 5th grade, students are required by the district to read at least two novels, and so the reading program is novel-based. The comparison group teachers at the first elementary school have been following the current programs for an average of 6 years, and for an average of 7.8 years at the second elementary school. Teachers at both schools have daily language arts blocks of 60 to 120 minutes. The comparison teachers at the first elementary school use whole group instruction an average of 36% of the time, while comparison teachers at the second school teach whole group about 63% of the time. All but two

teachers at the first school receive some form of help during their literacy instruction in the form of a parent, a paraprofessional, literacy coach, or teacher's aide. This additional support is less common at the second school, with only four teachers having a parent, literacy coach, or paraprofessional in the classroom during reading instruction

Comparison group teachers in the first elementary school all frequently use a number of teaching strategies including; leveled instruction, cooperative learning, independent reading, guided reading, and seatwork. About half of the comparison teachers also use test prep, and all but one frequently use center activities. All comparison teachers at the first school monitor their students' progress two to three times per week.

At the second school, all teachers frequently use center activities, guided reading, and seat work. The majority also frequently uses leveled instruction, and guided reading. Cooperative learning strategies are also used by all teachers, though not as frequently. All but one comparison teacher conducts test-prep. All comparison teachers at the second school monitor their students for progress one to three times per week.

All but one teacher at both schools prefer a balanced literacy approach to reading instruction and all but one prefer using a combination of basal readers and trade books rather than just one or the other. All teachers at the first school supplement their literacy instruction with some form of technology. Most use educational computer programs at least once a week with their students, and more than half frequently use websites. All but two teachers at the second school incorporate some form of technology into their reading instruction, the most common being educational computer programs. About five teachers also use educational websites.

Ohio District

The Ohio elementary school opened recently. It is housed in a new building with state-of-the-art facilities, such as a digital white board in every classroom. Students have varied learning abilities, with any given class including both gifted and IEP students. The school resides in a rural area and served a community of 3,000 in 2008-2009. The median household income is approximately $50,000, indicating a middle class community. Teachers say that being a part of a smaller community allows them to know more about their students. They make an effort to meet the personalized needs of their students. They also credit the success of the school to having strong and supportive leadership.

The elementary school in Ohio is a small school serving approximately 320 students in grades kindergarten through six. The school has one primary ethnic group, Caucasian, representing 97% of the school population. Approximately 38% of students are eligible for free or reduced-price lunch. The elementary school did meet AYP in the 2008-09 school year. The percentage of 4th grade students testing at standard in reading in the 2008-09 school year was 86%, higher than the statewide results by 14%. The student/teacher ratio is approximately 22 to 1.

Nine teachers participated in the study, one kindergarten, and two from each of 1st, 2nd, 4th and 5th grade. Since there was only one kindergarten teacher, this teacher was assigned to use the Reading Street program with her two half-day classes. In 1st, 2nd, 4th, and 5th grades, one teacher from each grade level was randomly assigned to use the Reading Street program. None of these teachers were new to the school or district, and one was new to the grade level.

The district adopted a widely published elementary basal reading curriculum with a copyright date prior to 2005. Two out of three comparison teachers created their own reading programs, and one comparison teacher primarily uses the district-adopted program with some supplementation. The comparison group teachers have been following their current programs for an average of 5.7 years, ranging from one to 10 years. Daily literacy blocks range in duration from 50 minutes to 120 minutes. The comparison teachers use whole group instruction an average of 42% of the time. All teachers have additional classroom support during their literacy block from a paraprofessional, a Title 1 teacher, or a parent helper.

Comparison group teachers use each of the following strategies in their literacy instruction at least once per week: leveled instruction, cooperative learning, independent reading, guided reading, test-prep, and seatwork. Center activities are used often by one of the comparison teachers, and infrequently by the other two. All comparison teachers monitor their students for progress two to three times per week.

One teacher prefers a phonics approach to teaching literacy, while the rest of the teachers prefer a balanced literacy approach. Three of the teachers prefer using trade books, and five teachers like using a combination of trade books and basal readers. As mentioned above, all teachers have electronic white boards in their classroom, which are used frequently during literacy instruction. All teachers also make use of websites and all but one use educational computer games.

Washington District

The Washington elementary school resides in a large suburb. In 2008-09, the school district served a community of 18,000. The median household income is approximately $45,000, indicating a middle class community. This school serves approximately 300 students in grades kindergarten through six. The school has one primary ethnic group, Caucasian, representing 84% of the school population. Approximately 50% of students are eligible to receive free or reduced-price lunch.

The Washington state school has a number of different programs: a neighborhood program, a parent involvement program, a gifted program for grades 3-5, a special needs classroom, and a behavior disorder classroom. The many programs and opportunities reflect a school interested in meeting the needs of an ever-growing, diverse student population. The school has recently received awards for literacy and their administration. The elementary school did meet AYP in the 2008-09 school year and the percentage of 4th grade students testing at standard in reading was 83%, 9% more than the statewide results. The student/teacher ratio is approximately 16 to 1.

Nine teachers participated in the study. Two teachers each from kindergarten, 1st grade, and 2nd grade; one fourth grade teacher, one fourth and fifth grade combination class teacher, and one 5th grade teacher. One kindergarten, 1st, and 2nd grade teacher was randomly assigned to use Reading Street. In 4th and 5th grade, because there was only one full class at each grade level, these teachers were assigned to use Reading Street, and the 4th/5th grade combination teacher was assigned to the comparison group.

Teachers have daily language arts blocks of 60 - 90 minutes. Three of the four comparison teachers were piloting a widely published reading curriculum with a recent copyright. The

fourth teacher was still using the older district-adopted reading program. Two of these teachers primarily used their district-adopted program with some supplemental materials, another teacher heavily supplemented the district-adopted program, and the fourth teacher primarily used her own materials. The comparison teachers use whole group instruction an average of 40% of the time. All of the teachers have a paraprofessional or teacher's aid in their classroom for support, and some additional help from parent helpers, or in the case of one teacher, a local college student.

All comparison group teachers frequently use leveled instruction and seat work to complement their reading instruction. Three out of four frequently use center activities, and three teachers also use cooperative learning strategies at least once per week. In terms of progress monitoring, three of the comparison teachers monitor students for progress once a week, and the fourth monitored students less than once per week.

All teachers at the Washington school take a balanced literacy approach to literacy instruction. Six of these teachers like to use a combination of basal readers and trade books, while two teachers prefer using trade books. Interactive white boards are the most popular form of classroom technology, and are used frequently by four teachers, and less than once per week by one teacher. Educational computer games are also used at least once per week by four of the teachers.

Participants

The final diverse sample consisted of 1,480 kindergarten, 1st, and 4th grade students from eight sites, in six states, located in different regions of the US.

The final sample is comprised of students and teachers from 82 diverse elementary grade classrooms in nine schools distributed across six different states (i.e., AZ, CO, MA, MT, OH, and WA). The final study sample consisted of 472 kindergarten (i.e., Reading Street = 269, comparison = 203), 491 1st grade (i.e., Reading Street = 261, comparison = 230) and 517 4th grade (i.e., Reading Street = 265, comparison = 252) students. Eighty-three percent of the kindergarten students tested at baseline remained in the final study sample (i.e., Reading Street = 89%, comparison = 95%). Likewise, 86% of the 1st grade (i.e., Reading Street = 87%, comparison = 86%) and 90% of the 4th grade (i.e., Reading Street = 87%, comparison = 91%) students tested at baseline remained.

The data in Table 3 provides the demographic breakdown of the final study sample. The study schools demonstrated considerable variation in reading achievement and ethnicity, as well as percentage of students eligible for free or reduced-priced lunch. Overall, 30% of the final study sample was eligible to receive free or reduced-priced lunch, 6% were not English proficient, 11% were Hispanic, and 4% were African American. On average, the 4th graders scored at the 59th percentile at baseline on the GRADE. Two schools, the Arizona and the first Colorado school, had the majority of the Hispanic and African-American students represented in the study sample. The Arizona's school's population was mostly Hispanic and the vast majority of the students were eligible to receive free or reduced-priced lunch.

Table 3		Gatti Evaluation Reading Street RCT Sample Demographic Information							
Group	Grade	[1]Student Count	[2]Percent In Low Achieving Group	Percent Not English Proficient	Percent Reduced Lunch	Percent Caucasian	Percent Hispanic/ Native American	Percent African American /Caribbean	Other Ethnicity or No Information
Arizona District									
RS	K	35 (83%)	83%	54%	91%	6%	69%	11%	14%
Comparison		18 (95%)	56%	0%	72%	11%	61%	17%	11%
RS	1	20 (71%)	45%	35%	80%	15%	60%	20%	5%
Comparison		22 (76%)	59%	23%	91%	14%	68%	9%	9%
RS	4	18 (78%)	100%	89%	94%	6%	72%	0%	22%
Comparison		22 (76%)	36%	18%	86%	9%	73%	18%	0%
[3]Colorado District 1									
RS	K	29 (52%)	76%	0%	62%	55%	14%	21%	10%
Comparison		19 (86%)	63%	11%	79%	53%	26%	21%	0%
RS	1	39 (70%)	59%	5%	33%	64%	13%	15%	8%
Comparison		22 (88%)	59%	23%	50%	55%	32%	9%	4%
RS	4	32 (78%)	38%	9%	44%	44%	16%	28%	12%
Comparison		26 (67%)	31%	4%	31%	54%	19%	27%	0%
[4]Colorado District 2									
RS	K	23 (96%)	30%	17%	30%	70%	26%	0%	4%
Comparison		22 (88%)	27%	5%	45%	91%	5%	0%	4%
RS	1	13 (81%)	23%	0%	46%	85%	0%	0%	15%
Comparison		7 (100%)[3]	29%	0%	0%	100%	0%	0%	0%
RS	4	20 (79%)	5%	0%	35%	80%	5%	5%	10%
Comparison		20 (80%)	35%	0%	45%	80%	20%	0%	0%
Massachusetts District 1									
RS	K	53(100%)	9%	2%	4%	91%	4%	2%	3%
Comparison		36 (97%)	17%	3%	8%	83%	0%	0%	17%
RS	1	56 (98%)	20%	0%	2%	89%	0%	4%	7%
Comparison		59 (98%)	19%	0%	5%	97%	0%	2%	1%
RS	4	48 (96%)	4%	0%	8%	92%	2%	0%	6%
Comparison		48 (96%)	10%	4%	2%	96%	2%	0%	2%

Massachusetts District 2									
RS	K	49 (92%)	27%	12%	4%	90%	8%	0%	2%
Comparison		43(100%)	12%	5%	7%	88%	5%	0%	7%
RS	1	48 (94%)	25%	8%	6%	90%	4%	2%	4%
Comparison		51 (98%)	25%	0%	6%	94%	4%	0%	2%
RS	4	63 (94%)	5%	2%	3%	94%	3%	0%	3%
Comparison		68 (99%)	4%	0%	6%	93%	1%	1%	5%
Montana District									
RS	K	59 (94%)	34%	0%	44%	88%	3%	2%	7%
Comparison		42 (98%)	31%	2%	29%	83%	2%	5%	10%
RS	1	47 (92%)	36%	0%	23%	89%	2%	0%	9%
Comparison		30 (88%)	37%	0%	27%	77%	3%	3%	17%
RS	4	46 (92%)	7%	0%	22%	91%	2%	2%	5%
Comparison		48 (98%)	21%	0%	21%	85%	6%	0%	9%
Ohio District									
RS	1	16 (89%)	63%	0%	50%	88%	0%	6%	6%
Comparison		14(100%)	50%	0%	64%	100%	0%	0%	0%
RS	4	17(100%)	6%	0%	35%	94%	0%	0%	6%
Comparison		14 (88%)	29%	0%	64%	100%	0%	0%	0%
[5]Washington District									
RS	K	21 (91%)	38%	0%	76%	76%	5%	5%	14%
Comparison		23 (96%)	43%	0%	35%	100%	0%	0%	0%
RS	1	22 (92%)	41%	0%	73%	95%	0%	0%	5%
Comparison		25(100%)	28%	0%	28%	92%	4%	0%	4%
RS	4	21 (91%)	14%	0%	48%	81%	19%	0%	0%
Comparison[2]		6 (60%)[4]	50%	0%	100%	67%	17%	0%	16%

1. Percents within parentheses next to student counts indicate the percent of students tested at baseline that were also tested at the end of the school year.

2. The lower achieving kindergarten sample constituted those scoring at baseline within the first stanine of the norming sample (i.e., lowest 4%), 1st graders within the first, second, and third stanine at baseline (i.e., lowest 23%), and those 4th graders scoring at or below the 3.0 grade equivalent at baseline.

3. The first Colorado site could not provide meal program status for individual students. Participation in the meal program for each student was estimated by choosing the most likely participants as determined from all available known information.

4. The 1st grade Colorado District 2 comparison group lost 56% of their one classroom due to a lack of demographic information.

5. The 4th grade Washington comparison students came from a combination grade 4/5 classroom and were not numerous.

The tables below provide evidence the students from the two study conditions were similar in their demographic makeup.

Kindergarten	Reading Street	comparison
low achieving	39%	31%
not English proficient	11%	3%
free/reduced lunch	38%	32%
Hispanic	16%	10%
African American	5%	4%

1st Grade	Reading Street	comparison
low achieving	36%	33%
not English proficient	5%	4%
free/reduced lunch	28%	27%
Hispanic	8%	11%
African American	5%	3%

4th Grade	Reading Street	comparison
low achieving	16%	19%
not English proficient	8%	3%
free/reduced lunch	26%	26%
Hispanic	10%	12%
African American	4%	5%

The Ohio kindergarten sample was dropped from the final study sample due to the lack of a comparison group. As reflected in Table 3, the first Colorado district had relatively high attrition. This is due to the school serving a highly mobile community of military families. The first Colorado site could not provide meal program status for individual students. Participation in the meal program for each student was estimated by choosing the most likely participants as determined from all available known information. The 1st grade Colorado district two comparison group lost 56% of the students from one classroom due to a lack of demographic information. The 4th grade Washington comparison students came from a combination grade 4/5 classroom intended to handle a larger than expected student population, and therefore their numbers were not as large as their Reading Street peers.

Data Analysis Procedures

Statistical analyses were performed on students' end-of-year GRADE total and subtests, DIBELS scales, as well as, reading academic attitude and ERAS (a.k.a. Garfield) survey raw scores for each grade level. Results looked at baseline to end of year growth for Reading Street users and compared the Reading Street users to a comparison group. Reading Street users and comparison students comprised the two study conditions. Students in classrooms randomly assigned to use Reading Street received the program for basal reading/language arts instruction while students in the comparison classrooms received reading/language arts instruction from district adopted programs and/or those materials and methods preferred by their classroom teachers.

Results were also broken out and analyzed for separate levels of four key demographic variables (i.e., English proficiency, ethnicity, gender, meal status[6]). In addition, results were calculated for a group of lower achieving students at each grade level. No single set benchmark for the norming sample provided an adequately sized group of lower achieving students from the study sample. The lower achieving kindergarten sample constituted those scoring at baseline within the first stanine of the norming sample (i.e., lowest 4%), 1st graders within the third stanine at baseline (i.e., lowest 23%), and those 4th graders scoring at or below the 3.0 grade equivalent at baseline. These cutoffs allocated 35%, 35%, and 18% of the kindergarten, 1st, and 4th grade study samples to the lower achieving groups respectively.

Statistical analyses were performed on students' end-of-year GRADE total and subtests, DIBELS scales, as well as, academic attitude and reading survey scores for the three grade levels. Results were also broken out and analyzed for key sub populations of students.

Rigorous research design dictates that all characteristics of the study participants and their environmental influences that may impact the results must be equated across study groups. This is advised even when classrooms of students are randomly assigned to study groups. Random assignment can only probabilistically equate study groups prior to the start of the study. The statistical equating of confounding factors and maintenance of a controlled and consistent environment for the study participants ensures that differences found in the study groups on outcomes of interest may more confidently be attributed to the study conditions assigned to these groups.

Comparisons were made between study groups (i.e., comparison vs. Reading Street) using model adjusted group mean differences. Model adjusted group mean differences were calculated holding all covariates constant in an attempt to statistically equate the study groups on those constructs and remove their influence from the study group effect. Covariates included baseline scores, student demographic,[7] and 2009-2010 school year classroom environment indicators.[8] It

[6] The first Colorado site could not provide meal program status for individual students. This site did, however, provide the percent of students receiving free or reduced priced lunch in each classroom. Participation in the meal program for each student was estimated by choosing the most likely participants as determined via the EM algorithm using all available known student and classroom level information.
[7] gender, meal program status, ethnicity, English proficiency, special education designation

should be noted that the estimated group mean differences are expected to change depending on the covariates included in the model. The same covariates were used in all models. When results are broken out by a demographic variable or a grouping indicator the group mean difference is no longer adjusted by that variable along with the remaining model covariates, rather, these differences are separated by the levels of that variable.

A random intercepts model was employed to estimate and test model adjusted group mean differences. While students were the unit of analysis, the school districts were the independent units. The hierarchical nature of the data (i.e., students nested within classrooms, classrooms nested within schools, schools nested within districts) has the effect of reducing the amount of independent information available in the sample, therefore decreasing the precision of estimates and the power of hypothesis tests to find these estimates statistically significant.[9] A naïve covariance structure,[10] grouped by grade, within a robust empirical standard error formulation was used to calculate confidence intervals for estimated effects. This procedure results in group mean differences that are unbiased and statistical hypothesis tests that are consistent[11,12] despite the nested nature of the data.

All statistical significance tests are two-tailed with a Type I error rate of 0.05. Statistically significant estimates are ones in which the probability of sampling scores that result in a group mean difference that much greater than zero when it is in fact null, is p = 0.05 or 1 in 20 samples. Significance implies that the samples are likely drawn from two separate populations or that the group averages are unlikely to be the same in the population.

Standardized effect size estimates (i.e., effect size = estimated adjusted group difference / comparison sample standard deviation) are computed for statistically significant model adjusted group mean differences using the sample standard deviation for the comparison group's end-of year-scores.[13] The statistical models were able to find small to moderate effect sizes statistically significant. The average minimal detectable effect size for the statistical tests of adjusted group mean differences across three grade levels on all outcome measures was 0.20 standard deviations (i.e., range of 0.02 to 0.43). Effect sizes as large as 0.20 standard deviations are most likely of practical significance. The careful review of efficacy studies for educational materials[14] indicate that the average adjusted group mean difference for studies with large samples (i.e., more than 250 students) is only 0.13 standard deviations.

---

[8] teacher education and experience, classroom assistance, classroom demographics, class size, baseline classroom achievement and variation, testing time span, days using program, teaching philosophy and most recent curricular choices, basal curricula and years using basal curricula, pullout frequency and minutes

[9] Donnar, A. & Klar, N. (2000) *Design and analysis of cluster randomization trials in health research*. Arnold Publishers, London.

[10] Initially a compound symmetric structure was assumed for the error variances but the extra parameter was not statistically significant for any of the statistical models.

[11] Liang, N. M. & and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73*, pp. 13-22.

[12] SAS's Mixed procedure was used to analyze the data, see SAS Institute Inc. (2008) Online documentation 9.2. A linear model was defined with all fixed effects, full degrees of freedom (i.e., N-2), using the sandwich estimator for all standard errors with districts set as the independent level of nesting, and a naïve independent working covariance structure.

[13] Hedges, L. V. & Olkin, I. (1985). *Statistics methods for meta-analysis*. Academic Press, NY.

[14] Slavin, R. & Smith, D. (2009). The relationship between sample sizes and effect sizes in systematic reviews in education. *Educational Evaluation and Policy Analysis, 31*(*4*) pp. 500-506.

III. RESULTS

Report section III summarizes the results of data analyses, including statistical and qualitative results, and group comparisons at baseline. The first subsection demonstrates the equivalence of the samples on the quantitative outcome measures at baseline. The second subsection addresses research question one, demonstrating the reading achievement and attitude gains over the school year. The third subsection addresses research question two, comparing achievement for the Reading Street group to that of the comparison group. The fourth subsection addresses research question three, do Reading Street students demonstrate more positive attitudes toward reading and reading instruction. Finally, section five summarizes feedback collected from Reading Street teachers during focus groups interviews, addressing research question five.

Baseline Group Equivalence

Tables 4 through 8 present both the simple sample[15] and model adjusted[16] baseline group mean differences for each measure of achievement and attitude for kindergarten, 1st and 4th grade classrooms. These tables also show statistical significance test results and effect size measures for the baseline group mean differences. All baseline achievement measures are statistically equivalent for both 1st and 4th grade samples, and the magnitude of actual differences are very small.

Table 4	Kindergarten Baseline GRADE Score Study Group Comparisons						
Measure	Sample Size SM/CP	Sample Difference	Sample p-value	Sample Effect Size	Adjusted Difference	Adjusted p-value	Adjusted Effect Size
GRADE Overall	472	-2.5666	0.0418	-0.23	-1.2140	0.3479	-0.11
Word Study	472	-2.4811	0.0098	-0.26	-1.7027	0.0779	-0.18
Listening Comprehension	472	-0.0856	0.8364	-0.03	0.0908	0.8277	0.03
Phoneme Segmentation Fluency	463	-1.1036	0.5906	-0.07	-0.0001	0.9999	0.00
Nonsense Word Fluency	463	-4.8805	<.0001	-0.22	-3.5695	0.0009	-0.16

Adjusted baseline group mean differences are estimated holding student demographic variables constant across groups. Sample group mean differences are estimated allowing student demographics to vary as they were sampled and randomly assigned. Kindergarten baseline Phoneme Segmentation and Nonsense Word Fluency were measured at mid-year.

The kindergarten comparison students outperformed the Reading Street students on the Word Study GRADE subtest, and therefore the GRADE total score, at baseline. When adjusted for student demographic variables (i.e., English proficiency, ethnicity, gender, meal status) this difference was no longer statistically significant. Likewise, the kindergarten comparison

[15] Sample group mean differences are estimated allowing student demographics to vary as they were sampled and randomly assigned.
[16] Adjusted baseline group mean differences are estimated holding student demographic variables constant across groups.

students outperformed the Reading Street students on the DIBELS Nonsense Word Fluency scale. However, this difference remained statistically significant when adjusted for student demographics. The magnitude of the difference is still rather small.

Table 5	1st Grade Baseline GRADE Score Study Group Comparisons						
Measure	Sample Size SM/CP	Sample Difference	Sample p-value	Sample Effect Size	Adjusted Difference	Adjusted p-value	Adjusted Effect Size
GRADE Overall	491	-1.4250	0.3616	-0.07	-1.4642	0.3503	-0.08
Word Study	491	-1.3510	0.0825	-0.12	-1.3589	0.0771	-0.12
Reading Comprehension	491	-0.0741	0.9310	-0.01	-0.1031	0.9057	-0.01
Listening Comprehension	490	-0.3610	0.2776	-0.13	-0.3557	0.2792	-0.13
Phoneme Segmentation Fluency	488	-0.4792	0.7312	-0.03	-0.5901	0.6557	-0.04
Nonsense Word Fluency	488	-0.4651	0.7790	-0.02	-0.4994	0.7594	-0.02
Adjusted baseline group mean differences are estimated holding student demographic variables constant across groups. Sample group mean differences are estimated allowing student demographics to vary as they were sampled and randomly assigned.							

Table 6	4th Grade Baseline GRADE Score Study Group Comparisons						
Measure	Sample Size SM/CP	Sample Difference	Sample p-value	Sample Effect Size	Adjusted Difference	Adjusted p-value	Adjusted Effect Size
GRADE Overall	517	0.4633	0.7850	0.03	0.3430	0.8358	0.02
Word Study	517	-0.0211	0.9693	0.00	-0.0587	0.9155	-0.01
Reading Comprehension	517	0.4844	0.6857	0.05	0.4838	0.6746	0.05
Listening Comprehension	513	-0.1411	0.2870	-0.09	-0.1147	0.2976	-0.07
Oral Reading Fluency	514	0.2681	0.9384	0.01	0.1248	0.9725	0.00
Adjusted baseline group mean differences are estimated holding student demographic variables constant across groups. Sample group mean differences are estimated allowing student demographics to vary as they were sampled and randomly assigned.							

The kindergarten comparison students also statistically outperformed the Reading Street students on the academic attitude survey at baseline. When adjusted for student demographic variables, this difference was no longer statistically significant. In addition, the 1st grade comparison students outperformed the Reading Street students on the ERAS (Garfield) survey. This difference straddled statistical significance, however the magnitude of this effect is also on the small side.

Table 7	Baseline Reading Academic Attitude Survey Score Comparisons						
Grade	Sample Size SM/CP	Sample Difference	Sample p-value	Sample Effect Size	Adjusted Difference	Adjusted p-value	Adjusted Effect Size
Kindergarten	426	-0.9823	0.0421	-0.21	-0.8742	0.0634	-0.19
1st Grade	471	-0.6082	0.1944	-0.12	-0.5941	0.2056	-0.11
4th Grade	513	0.0276	0.9482	0.00	-0.0096	0.9815	0.00
Adjusted baseline group mean differences are estimated holding student demographic variables constant across groups. Sample group mean differences are estimated allowing student demographics to vary as they were sampled and randomly assigned.							

Table 8	Baseline Garfield Reading Attitude Survey Score Comparisons						
Grade	Sample Size SM/CP	Sample Difference	Sample p-value	Sample Effect Size	Adjusted Difference	Adjusted p-value	Adjusted Effect Size
Kindergarten	454	-0.4904	0.8305	-0.03	-0.7226	0.7520	-0.05
1st Grade	480	-2.4102	0.0548	-0.18	-2.5292	0.0452	-0.19
4th Grade	512	-0.2731	0.7941	-0.02	-0.3722	0.7199	-0.03
Adjusted baseline group mean differences are estimated holding student demographic variables constant across groups. Sample group mean differences are estimated allowing student demographics to vary as they were sampled and randomly assigned.							

Reading Street Student Gains

The GRADE achievement data indicates clearly that students can experience substantial achievement growth when using the Reading Street program during the first year of implementation.

This subsection attempts to answer research question one, or

Do students using the Reading Street curriculum demonstrate significant gains in reading/language arts achievement during the first year of curriculum implementation?

The GRADE achievement data indicates clearly that students can experience substantial achievement growth when using the Reading Street program during the first year of implementation. As evidenced by GRADE total score, growth for all three grade levels was very large and statistically significant (1.74, 2.09, and 0.74 standard deviations for kindergarten, 1st, and 4th grade respectively). The gains across the GRADE subtests were also large. In addition,

at-risk (i.e., low achieving, free/reduced priced lunch eligible, not English proficient, non-Caucasian) students at all three grade levels gained about as much or more than their counterparts. Achievement gains for all subpopulations can be found in Appendix 1.

The gains in GRADE scores for all three grade levels were very large and statistically significant.

Kindergarten Scale	GRADE Effect Size[1]
GRADE Total	1.74
Word Study	1.72
Listening Comprehension	0.83
1. effect size = baseline to EOY Reading Street group gain / comparison sample standard deviation	

1st Grade Scale	GRADE Effect Size[1]
GRADE Total	2.09
Word Study	1.54
Reading Comprehension	1.53
Listening Comprehension	0.65
1. effect size = baseline to EOY Reading Street group gain / comparison sample standard deviation	

4th Grade Scale	GRADE Effect Size[1]
GRADE Total	0.74
Word Study	0.70
Reading Comprehension	0.50
1. effect size = baseline to EOY Reading Street group gain / comparison sample standard deviation	

Group Comparisons of Achievement Gains

This section will attempt to answer research question two:

How does reading performance differ for students using the Reading Street curriculum as compared to their peers using other elementary reading curricula?

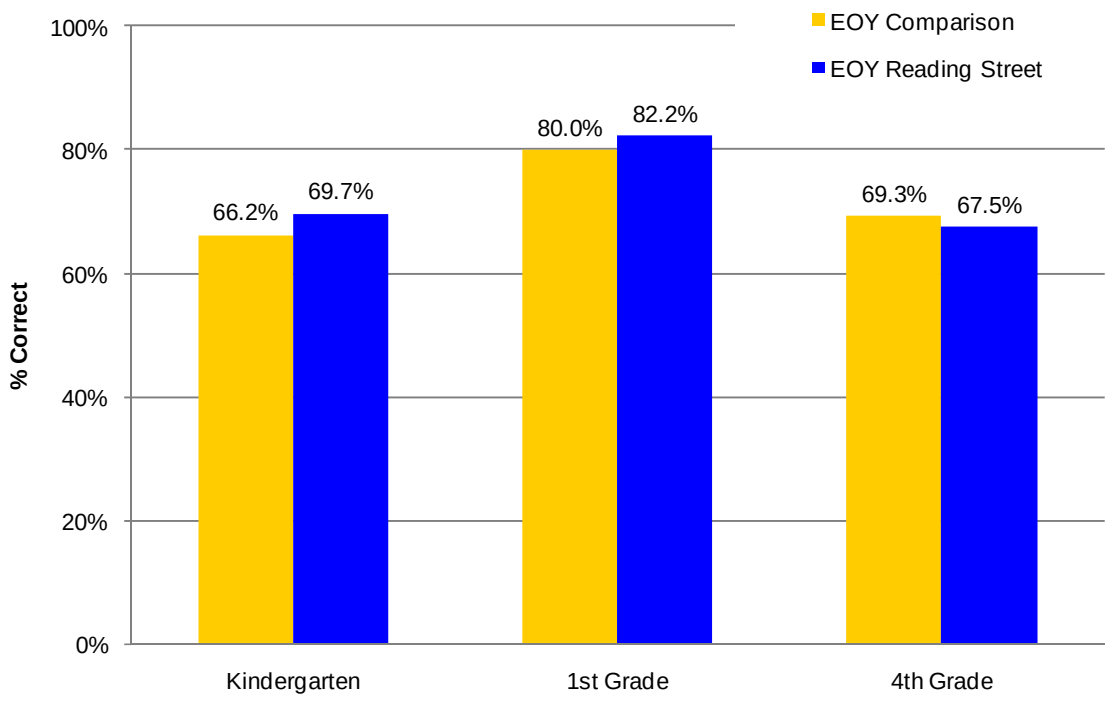
Results reported in this section compared the performance of the Reading Street users to that of a comparison group of students. Reading Street users and comparison students comprised the two study conditions. Students in classrooms randomly assigned to use Reading Street received the

program for basal reading/language arts instruction while students in the comparison classrooms received reading/language arts instruction from district adopted programs and/or those materials and methods preferred by their classroom teachers.

Figures 1-7 present the Reading Street and comparison model adjusted group means for GRADE total and subtest scores, as well as DIBELS scales for each grade level. Figures 2-7 may be found in Appendix 2. Adjusted group mean differences are calculated holding the effects of confounding factors (i.e., baseline scores, student demographic information, and classroom environment indicators) constant for both groups effectively removing their influence from the results. In addition to the initial random assignment, the equating of confounding factors and maintaining consistent implementation ensures that differences found in the study groups may more confidently be attributed to the study conditions.

Reading Street students outperformed the comparison group at kindergarten and 1st grade. Reading Street students statistically outperformed their comparison group peers at kindergarten by 0.38 standard deviations or 15% in percentile rank. Although not statistically significant, 1st grade Reading Street users also had greater gains in achievement than their comparison group peers by 0.13 standard deviations or 5% in percentile rank. Fourth grade results were not significant, indicating statistically similar growth in student achievement (i.e., 0.10 standard deviations or 4% in percentile rank).

Figure 1: GRADE Total Score



After adjusting for student and classroom characteristics, Kindergarten & 1st grade Reading Street students outscored their comparison group counterparts by 3.5% (SE=0.79%) and 2.2% (SE=2.32%) respectively. The 4th grade comparison group students outscored their Reading Street counterparts by 1.8% (SE=1.63%).

Kindergarten students using Reading Street statistically significantly outperformed their peers on the GRADE total score. They also significantly outperformed their peers on the GRADE Listening Comprehension subtest, DIBELS Phoneme Segmentation Fluency, and the DIBELS Nonsense Word Fluency. Kindergarten students using Reading Street performed similarly to their comparison peers on the GRADE Word Study subtest.

Kindergarten Scale	Effect Size[1,2]
GRADE Total	0.38
Word Study	***
Listening Comprehension	0.22
Phoneme Segmentation Fluency	0.45
Nonsense Word Fluency	0.15

*** Indicates group means are not statistically significantly different

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

First grade students using Reading Street performed similarly to their peers on the GRADE in total and on the Word Study and Reading Comprehension subtests. Reading Street students statistically significantly outperformed their peers on the GRADE Listening Comprehension subtest and DIBELS Oral Reading Fluency test. Conversely, 1st grade students using other reading programs statistically significantly outscored the Reading Street students on the DIBELS Phoneme Segmentation Fluency and the DIBELS Nonsense Word Fluency.

1st Grade Scale	Effect Size[1,2]
GRADE Total	***
Word Study	***
Reading Comprehension	***
Listening Comprehension	0.16
Phoneme Segmentation Fluency	-0.62
Nonsense Word Fluency	-0.46
Oral Reading Fluency	0.29

*** Indicates group means are not statistically significantly different

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

Fourth grade students using the Reading Street program performed similarly to their peers on the GRADE in total. They also performed similarly to the comparison group on the GRADE Word Study subtest. The 4th grade comparison group statistically significantly outperformed the Reading Street group on the GRADE Reading Comprehension subtest and the DIBELS Oral Reading Fluency.

4th Grade Scale	Effect Size[1,2]
GRADE Total	***
Word Study	***
Reading Comprehension	-0.35
Oral Reading Fluency	-0.12

*** Indicates group means are not statistically significantly different

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

Results by Subpopulations

In this section, the performance of Reading Street students from a specific subpopulation (ex., those Reading Street students eligible for reduced priced lunch) is statistically compared to that of their comparison group peers from the same population (i.e., students eligible for reduced priced lunch from classrooms that did not switch their curricula and methods). Results were broken out and analyzed for lower achieving students at each grade level as well as separate levels of four key demographic variables (i.e., English proficiency, ethnicity, gender, meal status).

The comparison students did not statistically outscore the Reading Street students on the GRADE for any of the kindergarten subpopulations. Kindergarten Reading Street students who were classified as male, female, eligible for reduced price lunch, eligible for full-priced lunch, English proficient, or Caucasian statistically significantly outperformed their comparison group peers on the GRADE total score. Low achieving, not English proficient, African American, and Hispanic kindergarten students performed statistically similarly between study groups.

Kindergarten Subpopulation	GRADE Effect Size[1,2]
Low achieving	***
Male	0.41
Female	0.36
Reduced priced lunch	0.36
Full priced lunch	0.25
Not English proficient	***
English proficient	0.33
African American	***
Hispanic	***
Caucasian	0.29

*** Indicates group means are not statistically significantly different.

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

Again, the comparison students did not statistically outscore the Reading Street students on the GRADE for any of the 1st grade subpopulations. Reading Street students classified as eligible for reduced price lunch statistically significantly outperformed their reduced price lunch comparison group peer on the GRADE. The remainder of the subpopulations performed statistically similarly across study conditions.

1st Grade Subpopulation	GRADE Effect Size[1,2]
Low achieving	***
Male	***
Female	***
Reduced priced lunch	0.34
Full priced lunch	***
Not English proficient	***
English proficient	***
African American	***
Hispanic	***
Caucasian	***

*** Indicates group means are not statistically significantly different.

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

Male comparison students statistically outperformed their male Reading Street peers on the GRADE. The remainder of the subpopulations performed statistically similarly across study conditions.

4th Grade Subpopulation	GRADE Effect Size[1,2]
Low achieving	***
Male	-0.17
Female	***
Reduced priced lunch	***
Full priced lunch	***
Not English proficient	***
English proficient	***
African American	***
Hispanic	***

Caucasian	***

*** Indicates group means are not statistically significantly different.

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

All subpopulations of kindergarten students performed statistically similarly across study conditions in Word Study.

Kindergarten Subpopulation	Word Study Effect Size[1,2]
Low achieving	***
Male	***
Female	***
Reduced priced lunch	***
Full priced lunch	***
Not English proficient	***
English proficient	***
African American	***
Hispanic	***
Caucasian	***

*** Indicates group means are not statistically significantly different.

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

When analyzing the Listening Comprehension GRADE subtest, male, eligible for free/reduced priced lunch, not English proficient, and English proficient kindergarten Reading Street students statistically outperformed their comparison group peers. The remainder of the subpopulations performed statistically similarly across study conditions.

Kindergarten Subpopulation	Listening Comprehension Effect Size[1,2]
Low achieving	***
Male	0.30
Female	***
Reduced priced lunch	0.18
Full priced lunch	***
Not English proficient	0.70

English proficient	0.18
African American	***
Hispanic	***
Caucasian	***

*** Indicates group means are not statistically significantly different.

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

Low-achieving, male, eligible for reduced-price lunch and full priced lunch, English proficient, Hispanic and Caucasian kindergarten Reading Street students statistically outperformed their comparison group peers on the DIBELS Phoneme Segmentation Fluency subtest. The remaining three subpopulations performed statistically similarly across study conditions.

Kindergarten Subpopulation	Phoneme Segmentation Fluency Effect Size[1,2]
Low achieving	0.71
Male	0.52
Female	***
Reduced priced lunch	0.58
Full priced lunch	0.41
Not English proficient	***
English proficient	0.46
African American	***
Hispanic	0.37
Caucasian	0.45

*** Indicates group means are not statistically significantly different.

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

Lastly, male, eligible for reduced-price lunch, English proficient, and Hispanic kindergarten Reading Street students statistically outperformed their comparison group peers on the DIBELS Nonsense Word Fluency subtest. The remaining subpopulations performed statistically similarly across study conditions.

Kindergarten Subpopulation	Nonsense Word Fluency Effect Size[1,2]
Low achieving	***
Male	0.16

Female	***
Reduced priced lunch	0.33
Full priced lunch	***
Not English proficient	***
English proficient	0.15
African American	***
Hispanic	0.21
Caucasian	***

*** Indicates group means are not statistically significantly different.

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

At first grade, Reading Street students who were classified as eligible for reduced price lunch statistically significantly outperformed their reduced price lunch comparison group peers on the GRADE Word Study subtest. The remainder of the subpopulations performed statistically similarly across study conditions.

1st Grade Subpopulation	Word Study Effect Size[1,2]
Low achieving	***
Male	***
Female	***
Reduced priced lunch	0.39
Full priced lunch	***
Not English proficient	***
English proficient	***
African American	***
Hispanic	***
Caucasian	***

*** Indicates group means are not statistically significantly different.

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

All subpopulations of 1st grade students performed statistically similarly across study conditions on the GRADE Reading Comprehension subtest.

1st Grade Subpopulation	Reading Comprehension Effect Size[1,2]
Low achieving	***
Male	***
Female	***
Reduced priced lunch	***
Full priced lunch	***
Not English proficient	***
English proficient	***
African American	***
Hispanic	***
Caucasian	***

*** Indicates group means are not statistically significantly different.

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

Male, full priced lunch and Caucasian 1st grade Reading Street students statistically outperformed their comparison group peers on the GRADE Listening Comprehension subtest. The remaining subpopulations performed statistically similarly across study conditions.

1st Grade Subpopulation	Listening Comprehension Effect Size[1,2]
Low achieving	***
Male	0.31
Female	***
Reduced priced lunch	***
Full priced lunch	0.20
Not English proficient	***
English proficient	***
African American	***
Hispanic	***
Caucasian	0.17

*** Indicates group means are not statistically significantly different.

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

First grade comparison students from all subpopulations statistically outperformed their Reading Street peers on the DIBELS Phoneme Segmentation Fluency subtest. Further, with the exception of the African American students, the remaining first grade comparison students statistically outperformed their Reading Street peers on the DIBELS Nonsense Word Fluency subtest.

1st Grade Subpopulation	Phoneme Segmentation Fluency Effect Size[1,2]
Low achieving	-0.56
Male	-0.66
Female	-0.59
Reduced priced lunch	-0.54
Full priced lunch	-0.65
Not English proficient	-0.64
English proficient	-0.62
African American	-1.08
Hispanic	-0.68
Caucasian	-0.57

*** Indicates group means are not statistically significantly different.

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

1st Grade Subpopulation	Nonsense Word Fluency Effect Size[1,2]
Low achieving	-0.35
Male	-0.52
Female	-0.41
Reduced priced lunch	-0.45
Full priced lunch	-0.47
Not English proficient	-1.01
English proficient	-0.43
African American	***
Hispanic	-0.65
Caucasian	-0.42

*** Indicates group means are not statistically significantly different.

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

Conversely, Reading Street students statistically significantly outperformed their comparison group peers on the DIBELS Oral Reading Fluency subtest for most of the 1st grade subpopulations. This includes low achieving, male, female, eligible for reduced priced or full priced lunch, English proficient, and Caucasian students. The remaining three subpopulations performed statistically similarly across study conditions.

1st Grade Subpopulation	Oral Reading Fluency Effect Size[1,2]
Low achieving	0.30
Male	0.26
Female	0.32
Reduced priced lunch	0.31
Full priced lunch	0.28
Not English proficient	***
English proficient	0.30
African American	***
Hispanic	***
Caucasian	0.37

*** Indicates group means are not statistically significantly different.

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

At 4th grade only one subpopulation of students produced a statistically significant effect on the GRADE Word Study subtest. The not English proficient comparison group outperformed the Reading Street group, the remaining subpopulations performed statistically similarly across study conditions.

4th Grade Subpopulation	Word Study Effect Size[1,2]
Low achieving	***
Male	***
Female	***
Reduced priced lunch	***
Full priced lunch	***
Not English proficient	-0.67
English proficient	***
African American	***

Hispanic	***
Caucasian	***

*** Indicates group means are not statistically significantly different.

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

Male, female, eligible for reduced-priced lunch and African American 4th grade comparison group student statistically outperformed their Reading Street peers on the GRADE Reading Comprehension subtest. The remaining subpopulations performed statistically similarly across study conditions.

4th Grade Subpopulation	Reading Comprehension Effect Size[1,2]
Low achieving	***
Male	-0.43
Female	-0.27
Reduced priced lunch	-0.49
Full priced lunch	***
Not English proficient	***
English proficient	***
African American	-0.64
Hispanic	***
Caucasian	***

*** Indicates group means are not statistically significantly different.

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

Lastly, 4th grade male, female, eligible for reduced-priced and full-priced lunch, not English proficient and English proficient, as well as Caucasian comparison group students statistically outperformed the Reading Street group on the DIBELS Oral Reading Fluency scale.

4th Grade Subpopulation	Oral Reading Fluency Effect Size[1,2]
Low achieving	***
Male	-0.11
Female	-0.12
Reduced priced lunch	-0.14

Full priced lunch	-0.10
Not English proficient	-0.28
English proficient	-0.10
African American	***
Hispanic	***
Caucasian	-0.12

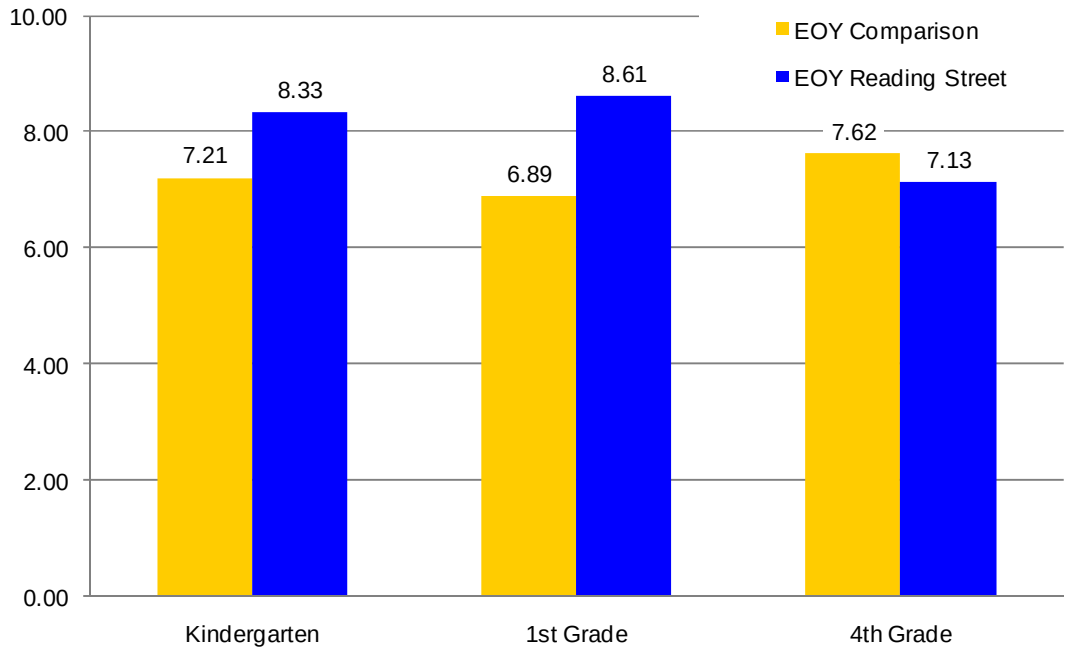
*** Indicates group means are not statistically significantly different.

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

Student Academic Attitudes

Figure 8: Reading Academic Attitude Survey



After adjusting for student and classroom characteristics, Kindergarten & 1st grade Reading Street students outscored their comparison group counterparts by 1.12 (SE=1.07) and 1.73 (SE=0.66) respectively . The 4th grade comparison group students outscored their Reading Street counterparts by 0.49 (SE=1.00).

This section will attempt to answer research question three:

Do Reading Street students demonstrate more positive attitudes toward reading and reading instruction as their peers using other elementary reading programs?

Reading Academic Attitude Survey

The reading academic attitude survey was administered to students to examine general reading attitude, confidence, motivation, self-perceived aptitude, vocabulary and comprehension, as well as, recreational and academic reading. Kindergarten and 1st grade Reading Street students had higher reading academic attitudes than their comparison group peers, although only 1st grade was statistically significant. Fourth grade Reading Street students had slightly lower reading academic attitudes than their comparison peers, but it was not statistically significant. Figure 8 shows the adjusted study group means on the reading academic attitude scale.

Student Reading Attitude	Effect Size[1,2]
Kindergarten	***
1st Grade	0.34
4th Grade	***

*** Indicates group means are not statistically significantly different

1. effect size = estimated adjusted group difference / comparison sample standard deviation
2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

Analyses of subpopulation results yielded some significant comparative effects. Specifically, Reading Street kindergarten students who were eligible for reduced-priced lunch or African-American had significantly higher attitudes than their peers using comparison programs. Reading Street students in 1st grade who were classified as low achieving, Hispanic, and not English proficient outperformed their comparison group peers by large margins. Finally, 4th grade African-American students using Reading Street had significantly higher attitudes than their African-American peers using other elementary reading programs. The only instance the comparison group students statistically outperformed the Reading Street group on the academic attitude survey was with 4th grade not English proficient students.

Kindergarten Subpopulation	Student Reading Attitude Effect Size[1,2]
Low achieving	***
Male	***
Female	***
Reduced priced lunch	0.35
Full priced lunch	***
Not English proficient	***
English proficient	***
African American	0.89
Hispanic	***
Caucasian	***

*** Indicates group means are not statistically significantly different

1. Cohen's *d* effect size = estimated group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

1st Grade Subpopulation	Student Reading Attitude Effect Size[1,2]
Low achieving	0.54
Male	0.56
Female	***
Reduced priced lunch	***
Full priced lunch	0.37
Not English proficient	0.46
English proficient	0.35
African American	***
Hispanic	0.61
Caucasian	0.37

*** Indicates group means are not statistically significantly different

1. Cohen's *d* effect size = estimated group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

4th Grade Subpopulation	Student Reading Attitude Effect Size[1,2]
Low achieving	***
Male	***
Female	***
Reduced priced lunch	***
Full priced lunch	***
Not English proficient	-0.39
English proficient	***
African American	0.22
Hispanic	***
Caucasian	***

*** Indicates group means are not statistically significantly different

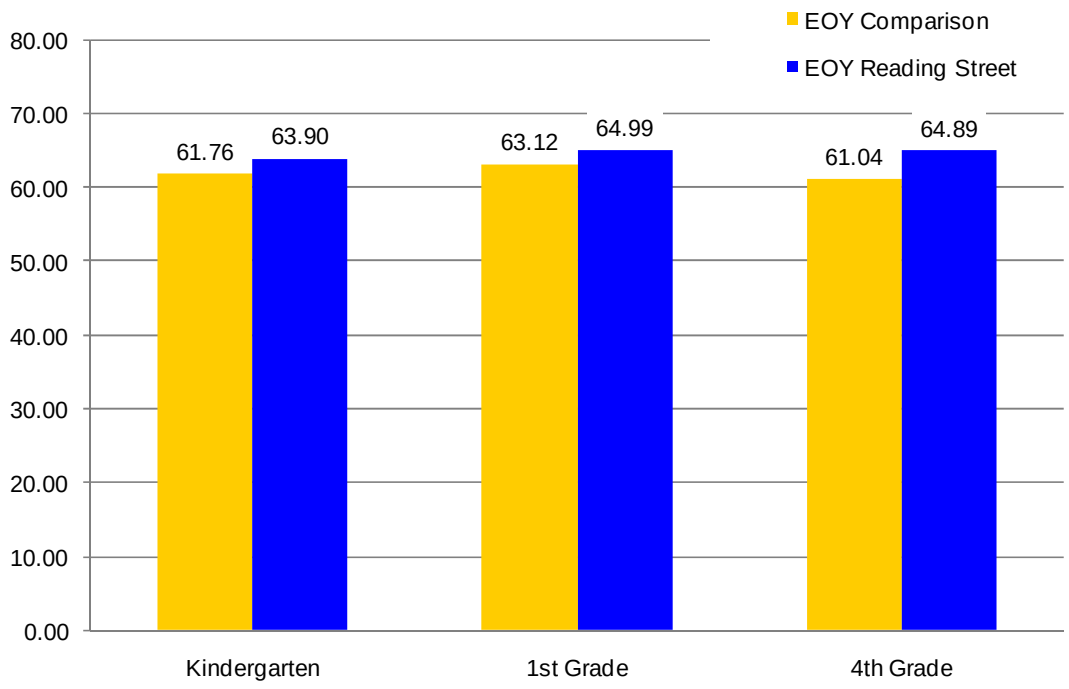
1. Cohen's *d* effect size = estimated group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

Early Reading Attitude Survey

The ERAS (i.e., Garfield) survey was also administered to measure student recreational and academic reading attitudes. The ERAS survey was administered because of its wide usage and recognition in the education field. The results from this survey indicate Reading Street students had more positive attitudes than their comparison group peers at all grade levels, although only the 4th grade difference was statistically significant. Figure 9 shows the differences in reading attitudes.

Figure 9: Elementary Reading Attitude (Garfield) Survey



After adjusting for student and classroom characteristics, Kindergarten, 1st & 4th grade Reading Street students outscored their comparison group counterparts by 2.14 (SE=1.22), 1.86 (SE=1.70) and 3.84 (SE=1.87) respectively.

Garfield Reading Survey	Effect Size[1,2]
Kindergarten	***
1st Grade	***
4th Grade	0.32

*** Indicates group means are not statistically significantly different

1. effect size = estimated adjusted group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

The ERAS results were also analyzed by the subpopulations. Reading Street kindergarten students who were eligible for reduced-priced lunch significantly outperformed their comparison peers. There were no other significant differences at kindergarten. There were also no significant comparative effects at 1st grade. However, at 4th grade, Reading Street students who were classified as lower-achieving, female, eligible for full-priced lunch, English proficient, and African-American outperformed their comparison group peers.

Kindergarten Subpopulation	Garfield Reading Survey Effect Size[1,2]
Low achieving	***
Male	***
Female	***
Reduced priced lunch	0.47
Full priced lunch	***
Not English proficient	***
English proficient	***
African American	***
Hispanic	***
Caucasian	***

*** Indicates group means are not statistically significantly different

1. Cohen's *d* effect size = estimated group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

1st Grade Subpopulation	Garfield Reading Survey Effect Size[1,2]
Low achieving	***
Male	***
Female	***
Reduced priced lunch	***
Full priced lunch	***
Not English proficient	***
English proficient	***
African American	***
Hispanic	***
Caucasian	***

*** Indicates group means are not statistically significantly different

1. Cohen's *d* effect size = estimated group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

4th Grade Subpopulation	Garfield Reading Survey Effect Size[1,2]
Low achieving	0.46
Male	***
Female	0.38
Reduced priced lunch	***
Full priced lunch	0.38
Not English proficient	***
English proficient	0.32
African American	0.81
Hispanic	***
Caucasian	***

*** Indicates group means are not statistically significantly different

1. Cohen's *d* effect size = estimated group difference / comparison sample standard deviation

2. The average effect size for studies with large samples (i.e., more than 250 students) has been recently estimated at 0.13 standard deviations.

Teacher and Student Reading Street Opinions

This section addresses research question five:

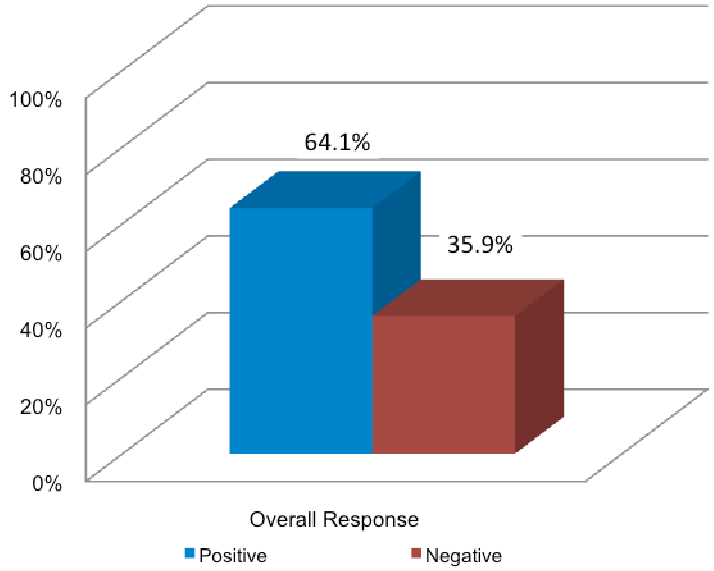
How did teachers and students react to the Reading Street program?

Opinions about the Reading Street program were systematically collected from teachers during focus group sessions. Focus groups were conducted at each school during site observations between April and mid-May. These sessions provided a forum for teachers and administrators to answer specific questions, as well as, express their professional and personal opinions regarding the program. The teachers were encouraged to speak without hesitation or inhibition and to be as candid as possible.

Each focus group session followed a standard format and was moderated by a member of the research team. Teachers were asked about their general opinions toward Reading Street, as well as their reactions to specific components of the program (e.g. Get Ready to Read, Read and Comprehend, Language Arts, etc.). Efforts were made to minimize response bias by avoiding leading questions, asking for both positive and negative information, and maintaining good rapport with participants.

A total of 913 comments were coded from the focus group sessions; 512 were positive or indicated perceived strengths of the program and 287 were negative or indicated perceived weaknesses of the program. The remaining 114 comments were neutral, or did not indicate perceived strengths or weaknesses of program. The Reading Street teachers presented an overall positive attitude about the program, with 64% of their focus group comments positive in nature.

Overall Teacher Response to Reading Street Program



When commenting on the program, teachers typically expressed satisfaction with the program.

Kindergarten Teacher: "I love it!"

Grade 1 Teacher: "I've really enjoyed it- it has closed some gaps for us."

Grade 4 Teacher: "There are many aspects that I like."

Grade 4 Teacher: "Just compared to last year's class, [this class] is far advanced."

Teacher response was mixed regarding the Teacher Edition and 5-Day Planner. Some found the layout difficult to follow, yet most agreed that the guide became more user-friendly as the months progressed. According to one 4th grade teacher, *"I was overwhelmed by the 12 teacher editions at first. Now it makes sense."* Most teachers felt that the teacher edition allowed for easy planning.

Grade 4 teacher: "When I am planning, everything is within the TE."

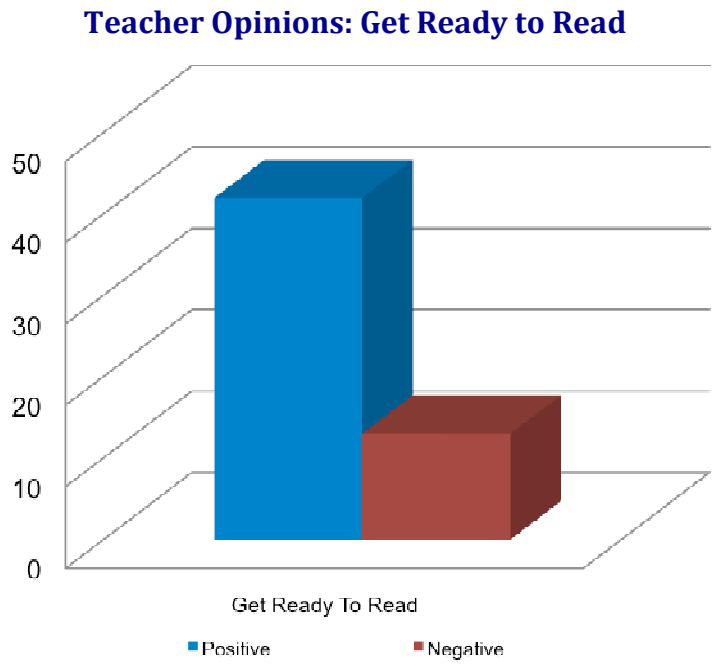
Get Ready to Read

The *Get Ready to Read* portion of the program encompasses concept development and vocabulary. This whole group instructional piece includes ELL poster activities, comprehension skills, reading, research, fluency, word analysis, genre study, and a weekly assessment. The

majority of teachers were very receptive to the design and content of this segment of the program.

Teachers responded most positively to the vocabulary component:

Kindergarten Teacher: "The kids like the name 'Amazing' words – think something with sparkles on it. It makes it [vocabulary] special."



Grade 1 Teacher: "The kids love the fact that they've learned so many amazing words."

Grade 4 Teacher: "The vocabulary piece has been really strong—I see a lot of progress with that and they really enjoy that."

Kindergarten Teacher: "If someone says them [Amazing Words] during the day, they love that. They like to pick them out in stories."

Many teachers also felt that the activities in the Get Ready to Read section provided ongoing scaffolding and practice.

Kindergarten teacher: "The kids like to hear the amazing words and vocabulary words throughout the week. Even as they hear it in Read Aloud book, big book, trade book they're quick to pick up on it, they're very excited. 'Look it's our amazing word.' That's a really good benefit that I haven't seen happen in previous schools."

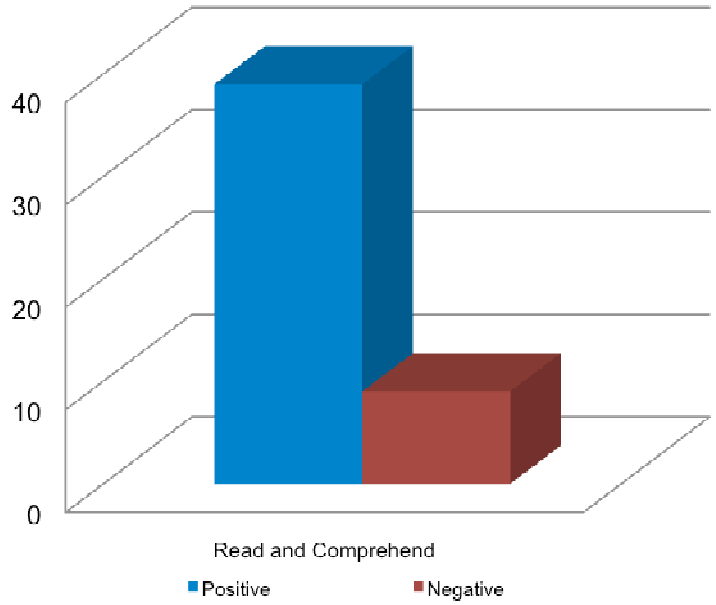
Grade 4 Teacher: "[The students are] very confident in their vocabulary, and I rarely get a student who gets more than one or two wrong on a vocabulary test. It goes in and stays there because we've used it so many times throughout the week."

Some teachers expressed a desire for more phonemic awareness. Others commented that the Amazing Words were either too easy or too difficult for their students. The majority of teachers, however, had very few negative things to say.

Read and Comprehend

The *Read and Comprehend* section includes a multi-day reading selection. Teachers responded positively to the Read and Comprehend portion of the program, particularly the stories. One 1st grade teacher reported, *"I really like the stories and the way the 2nd time we read it, the kids can partner up. It's very organized regarding the retell and showing the pictures again rather than the kids having to go back and remember what was happening."* In addition to the positive teacher feedback, students generally responded well to the Read and Comprehend section. According to one 4th grade teacher, *"[The students] really enjoy discussion around question of the week."*

Teacher Opinions: Read and Comprehend



Teachers appreciated that the read aloud and sing along components help students gain broad-based background knowledge and discover commonalities among themes, texts, authors, characters, and conflicts. According to one kindergarten teacher, *"The program as a whole is very good. I'm seeing some very good results from it. I appreciate the comprehension piece."*

Writing Portions

The writing component was met with mixed reviews on behalf of the teachers. Lower grade teachers appeared to be in consensus that the sequencing of writing skills was different from traditional methods, expressing opinions such as, *"The sequence is not right. I would like to see the list and thank you note after they have mastered paragraphs."*

In the higher grades, concerns with the writing section primarily focused on concerns about state testing. According to one 4th grade teacher, *"The writing doesn't work well for us. On our test we write an expository and a narrative. And to get it to the state standard it takes us half a year*

to do narratives and half a year to do expository. We don't sit and write everyday for an hour. In Reading Street, they might have good starters, but to get us to the level that we need, it would take a lot longer than a week to do a state standard type of job." Some teachers' also expressed concerns about learning and adopting a new methodology. One 4th grade teacher commented: *"The writing pieces are really hard for our kids. There aren't many mini lessons before it. It just assumes kids know how to do these things. 'So now, go write an invitation or now, go write a poem'. To us it feels a little disconnected."*

It appears that pedagogy and attitudes play an important role in the success of the writing portion. Teachers that responded well to the writing components generally considered the content and expectations as positively challenging material rather than difficult. One 4th grade teacher indicated: *"I was very concerned about writing at the beginning of year because we are responsible for long composition in March every year, and teaching our children length of writing. The positive [about Reading Street] is, children aren't bored with writing by the time March comes. They are learning to write and use all of the tools of writing on the smaller pieces such as advertisements and writing a myth. So they're doing some fun things. It really wasn't fun before. They do the expository and persuasive writing that we're eventually going to have to be responsible for, so I'm thrilled with that piece of the writing – it's a whole different outlook on it."*

Another teacher liked the *Look back and Write* portion of the weekly assessments.

Grade 4 teacher: "[The writing piece] keeps asking kids to look back and write – respond to text, and prompts them. That's a skill our kids need to do. The look back and write piece is really important for us."

In reviewing the focus group transcripts, it is clear that teachers were divided in their opinions of the writing portion of the program. Opinions stemmed from the abilities students brought to class, expectations based in traditional teaching styles, and open-mindedness on behalf of the teachers themselves.

Progress Monitoring and Assessment

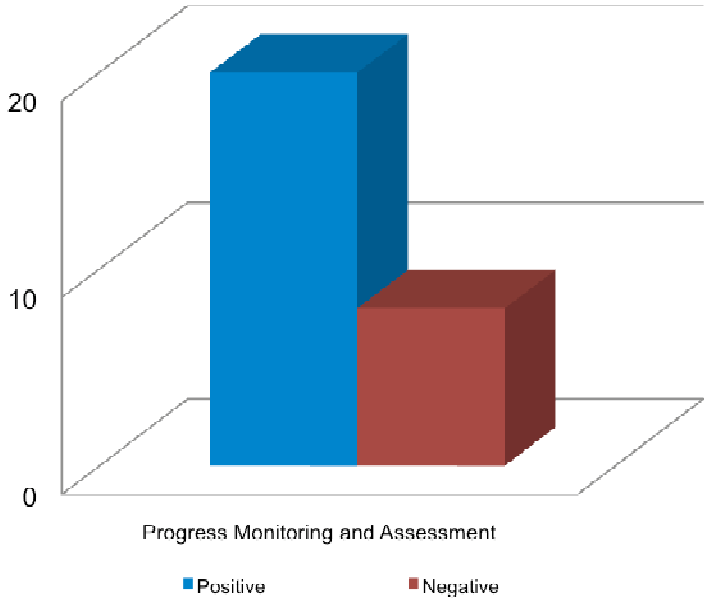
Teachers generally liked the progress monitoring and assessment components of the program. In particular, teachers working with a group of students that required differentiated instruction appreciated the ongoing attention paid to the needs of all students. According to one kindergarten teacher, *"It allows me to group my kids differently".* Teachers also appreciate showing parents and administrators progress made on behalf of students.

Grade 4 teacher: "At conference time I had a pile of tests I could show parents. You can see growth with that."

Despite the overall positive nature of responses, a small number of teachers felt that the assessment piece was not suitable for students performing below grade level. According to one kindergarten teacher: *"The weekly test does not help us with our low kids because it is 100% reading, and our SI kids can't even attempt it, so they get 0s every week. It doesn't show us any growth. And they are growing, but they are still below the test. I would like to see more SI questions, so they can at least get two right."* Such comments should not be considered entirely reflective of the efficacy of the monitoring and assessment components, particularly at the

kindergarten level. Students often arrive at school with varying educational experience, particularly at the beginning of the school year, requiring special accommodations.

Teacher Opinions: Progress Monitoring



Teachers generally agreed, however, that the progress monitoring and assessment components kept them adequately informed of student growth. These components often informed instruction and assisted with preparation.

Grade 1 teacher: "I like the end of the week, I know where they are. I continue to use end of week assessment with my SI students, because that helps direct my teaching."

Grade 1 teacher: "It's constantly on your mind, 'don't wait till Friday.'"

Student Response to Program

Teachers were overwhelmingly positive about their students' interactions with the program. Of the 219 recorded comments, 80% were positive in nature. Several teachers commented on the increased motivation, participation, and energy in their classrooms and students responding well to reading instruction, *"They do get engaged, I've gotten wonderful responses from it."*

In addition, teachers indicated growth in skill as a direct result of the increase in student involvement and structure offered by the program.

Kindergarten teacher: "My [students] love hearing a story read by me and they can actually follow along in the book. And now they can get it out by themselves. Now they can find the page. They love listening to the stories and following along, and they love the retell and being able to put them in order. That it's colorful helps them a lot. They're empowered now."

Kindergarten teacher: "My kids love the Skills Buddy!"

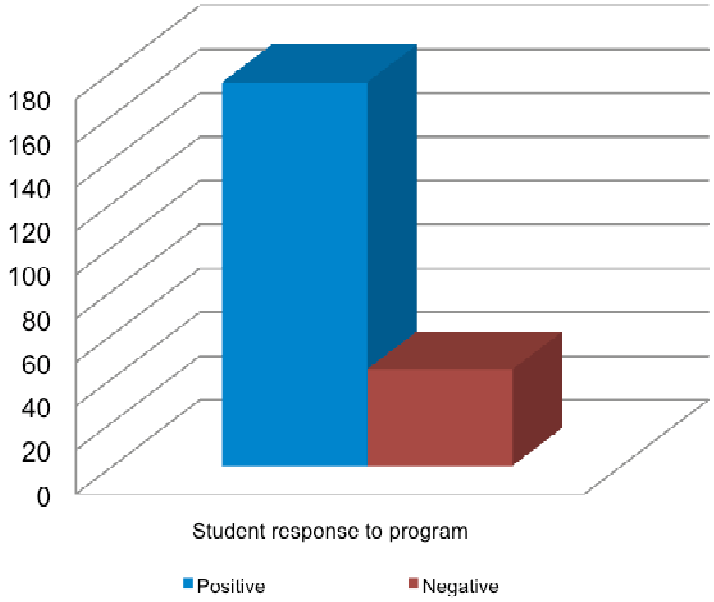
Kindergarten teacher: "They want to read to each other and they feel so accomplished."

Kindergarten teacher: "On Wednesday they know they are going to do sentences, and they say 'yay!' I like that on Friday they know they are going to have a test, they look forward to it."

Grade 1 teacher: "They don't whine and they're into it now and even when we're on the 2nd read they will flip right to it."

Grade 1 teacher: "They all seem to enjoy the program and are actively engaged in every aspect."

Overall Student Response to Reading Street



Students responded well to small group time, particularly at the kindergarten level. *"When we get the colored readers, they are like Yeah!"* said one kindergarten teacher. Another indicated, *"My students love the small groups. And I like it! Last week I was doing excessive DRA, so we went whole group, and they really missed it, and I missed it! They enjoy that small time with me."* Echoing most sentiments, another kindergarten teacher said, with regard to small group time: *"My kids are doing really well. I think because it is leveled, they can feel successful and feel really good about themselves, where as they may not feel that with whole group. And my high kids too, That is the one point where they are challenged."*

Materials, Content, Alignment, and Accessibility

Many teachers recognized that although the program appears to consist of an overwhelming amount of material at first, the plethora of material is beneficial when ensuring alignment with expectations, standards, and ability levels. As one 4th grade teacher commented, *"I've never had a shortage of things for them to do."*

Grade 1 teacher: "It's so differentiated. We are all doing simple machines this week, all three levels have a book about simples machines they can use."

Others felt that the content would benefit from supplementation with additional resources. According to one 1st grade teacher, *"If they're reading only 4 pieces, you've got to supplement with your own reading materials—they need to read more."*

Most teachers, however, found the content of the Reading Street program to be sufficient in meeting the needs of state and school assessments. According to one Reading Specialist, *"The Content is meeting state standards."* Similarly, another 4th grade teacher commented that the content *"Is wonderful. Everyone agrees in general it's meeting standards."*

In general, teachers appreciated *"how nice it was to have things to pick and choose from."*

Resoundingly, teachers and students responded positively to the leveled and decodable readers. As one 1st grade teacher stated, *"I like the leveled readers, that the language, and the writing is all in one place."* Teachers often applauded the readers' ability to combine skills taught throughout lessons. One teacher also appreciated "*the connection of leveled readers to main story."*

Grade 1 teacher: "The decodable readers really help my lower lever students."

Grade 4 teacher: "The leveled readers work well for all 3 levels, and students enjoy them. They've gotten a lot of practice which is great."

Focus Group Summary

Teachers and students alike had positive experiences with the Reading Street program. Teachers appreciated the program's components that support differentiated and small group instruction (ex., leveled and decodable readers), the ongoing progress monitoring and assessment, vocabulary, and multi-day reading selections. Teachers also very much liked that lessons are organized around central themes and that the program adds structure to the weekly reading/ELA instruction. Teachers also experienced some drawbacks. They felt that the pacing took some time to master and that the concept talk can run long or "*be wordy."* Many teachers also felt the consumable notebook could be improved in its worksheet content and usability, and that the layout of the teacher's edition could be improved. Many teachers also had initial reservations about the writing portions being too difficult or meeting content standards. Teachers firmly believe that the program increases motivation, participation, and energy in their classrooms. Students responded particularly well to small group time and the structure offered by the program.

IV. DISCUSSION

Though this section summarizes the findings from school year one of a two year longitudinal evaluation, information gathered from year one provide some interesting trends. Teachers and students alike had positive experiences with the Reading Street program. Teachers appreciated the program's components that support differentiated and small group instruction, the ongoing progress monitoring and assessment, vocabulary, and multi-day reading selections. Teachers also very much liked that lessons are organized around central themes and that the program adds structure to the weekly literacy instruction. Teachers also experienced some drawbacks. They felt that the pacing took some time to master and that the concept talk can run long or "*be wordy.*"

Teachers were overwhelmingly positive about their students' interactions with the program and firmly believe that the program increases motivation, participation, and energy in their classrooms. Students responded particularly well to small group time and the structure offered by the program. There was some evidence to support the Reading Street teachers' assertions as the 1st grade Reading Street students had statistically higher reading academic attitudes than those of their comparison group counterparts by a moderate amount. Some at-risk Reading Street populations, at all three grade levels, had statistically higher attitudes.

Teachers were overwhelmingly positive about their students' interactions with the program and firmly believe that the program increases motivation, participation, and energy in their classrooms.

The initial product training session, as well as three additional follow-up observations and training sessions, were well received by the Reading Street teachers. Reading Street teachers unanimously reported that instruction felt natural to them by winter break. A few months of use and discovery were necessary for the teachers to become fully comfortable with pacing and flow of the program. The same was true for their students falling into the structured routine offered by the Reading Street program. Average ratings from the final training session indicated that Reading Street teachers were implementing the required components well with good pacing. On a 1-10 scale with a rating of five 5 indicating average implementers, kindergarten = 7.4, 1st grade = 7.4, and 4th grade = 7.3. Seventy-seven percent of teachers received a rating of seven or higher.

The final sample consisted of 1,480 kindergarten, 1st, and 4th grade students from eight sites, in six states, located in different regions of the US. Overall, 30% of the final study sample was eligible to receive free/reduced priced lunch, 6% were not English proficient, 11% were Hispanic, and 4% were African American. Attrition at school year one was not a major concern as 86% of those students tested at baseline remained in the final study sample. The comparison group teachers used several published literacy programs along with a variety of methods and supplemental materials. Students received an average of 95 minutes of daily reading instruction (Reading Street = 93 minutes, comparison = 96 minutes).

When compared to similar students from the same school that continued to receive established programs and methods, students receiving the Reading Street program in their first school year performed equivalently or higher on the GRADE achievement test across the three grade levels.

The GRADE achievement data indicates clearly that students can experience substantial achievement growth when using the program in their first school year, even when schools and classroom teachers are implementing the program for the first time. The gains in GRADE scores for all three grade levels were very large and statistically significant. When compared to similar students from the same school that continued to receive established programs and methods, and after random assignment and statistical equating, the Reading Street students outperformed the comparison group at kindergarten and 1st grade. Reading Street students statistically outperformed their comparison group peers at kindergarten by 0.38 standard deviations or 15% in percentile rank. Although not statistically significant, 1st grade Reading Street users also had slightly greater gains in achievement than their comparison group peers by 0.13 standard deviations or 5% in percentile rank. Fourth grade results were not significant, indicating statistically similar growth in student achievement (i.e., 0.10 standard deviations or 4% in percentile rank).

The results for the DIBELS scales were statistically significant in the favor of Reading Street at kindergarten (i.e., PSF 17% and NWF 6% in percentile rank) and mixed at 1st grade with Phoneme Segmentation and Nonsense Word Fluency in favor of the comparison group (i.e., PSF -23% and NWF -18% in percentile rank) and Oral Reading Fluency in favor of Reading Street (i.e., 0.29 standard deviations or 11% in percentile rank). At 4th grade the results for Oral Reading Fluency were statistically significant and slightly in favor of the comparison group (i.e., -0.12 standard deviations or 5% in percentile rank). GRADE and DIBELS results for subpopulations of students were consistent with those seen for the whole sample.

A.1 Reading Street Gains by Subpopulation

Appendix 1 lists the standardized baseline to end-of-year average achievement score gain for all recorded subpopulations.

Kindergarten Subpopulation	GRADE Effect Size[1]
Low achieving	2.53
Male	1.78
Female	1.71
Reduced priced lunch	2.08
Full priced lunch	1.53
Not English proficient	2.24
English proficient	1.68
African American	2.09
Hispanic	2.17
Caucasian	1.60

1. effect size = baseline to EOY Reading Street group gain / comparison sample standard deviation

Kindergarten Subpopulation	Word Study Effect Size[1]
Low achieving	2.42
Male	1.74
Female	1.69
Reduced priced lunch	2.00
Full priced lunch	1.54
Not English proficient	1.98
English proficient	1.68
African American	2.06
Hispanic	2.05
Caucasian	1.60

1. effect size = baseline to EOY Reading Street group gain / comparison sample standard deviation

Kindergarten Subpopulation	Listening Comprehension Effect Size[1]
Low achieving	1.41
Male	0.88
Female	0.78
Reduced priced lunch	1.12
Full priced lunch	0.65
Not English proficient	1.74
English proficient	0.71
African American	1.00
Hispanic	1.26
Caucasian	0.70

1. effect size = baseline to EOY Reading Street group gain / comparison sample standard deviation

1st Grade Subpopulation	GRADE Effect Size[1]
Low achieving	2.54
Male	2.09
Female	2.08
Reduced priced lunch	1.99
Full priced lunch	2.13
Not English proficient	2.04
English proficient	2.09
African American	1.97
Hispanic	2.15
Caucasian	2.07

1. effect size = baseline to EOY Reading Street group gain / comparison sample standard deviation

1st Grade Subpopulation	Word Study Effect Size[1]
Low achieving	2.47
Male	1.59
Female	1.50
Reduced priced lunch	1.83
Full priced lunch	1.43
Not English proficient	1.60
English proficient	1.54
African American	1.52
Hispanic	1.66
Caucasian	1.52

1. effect size = baseline to EOY Reading Street group gain / comparison sample standard deviation

1st Grade Subpopulation	Reading Comprehension Effect Size[1]
Low achieving	1.34
Male	1.49
Female	1.57
Reduced priced lunch	1.15
Full priced lunch	1.69
Not English proficient	1.42
English proficient	1.54
African American	1.39
Hispanic	1.51
Caucasian	1.53

1. effect size = baseline to EOY Reading Street group gain / comparison sample standard deviation

1st Grade Subpopulation	Listening Comprehension Effect Size[1]
Low achieving	1.07
Male	0.76
Female	0.54
Reduced priced lunch	1.00
Full priced lunch	0.51
Not English proficient	0.47
English proficient	0.66
African American	***
Hispanic	0.56
Caucasian	0.65

*** Indicates gain was not statistically significantly different from zero

1. effect size = baseline to EOY Reading Street group gain / comparison sample standard deviation

4th Grade Subpopulation	GRADE Effect Size[1]
Low achieving	1.06
Male	0.67
Female	0.81
Reduced priced lunch	0.75
Full priced lunch	0.73
Not English proficient	0.69
English proficient	0.74
African American	0.74
Hispanic	0.70
Caucasian	0.74

1. effect size = baseline to EOY Reading Street group gain / comparison sample standard deviation

4th Grade Subpopulation	Word Study Effect Size[1]
Low achieving	0.74
Male	0.68
Female	0.71
Reduced priced lunch	0.66
Full priced lunch	0.71
Not English proficient	***
English proficient	0.73
African American	0.54
Hispanic	0.47
Caucasian	0.75

*** Indicates gain was not statistically significantly different from zero

1. effect size = baseline to EOY Reading Street group gain / comparison sample standard deviation

4th Grade Subpopulation	Reading Comprehension Effect Size[1]
Low achieving	0.92
Male	0.41
Female	0.60
Reduced priced lunch	0.55
Full priced lunch	0.48
Not English proficient	0.77
English proficient	0.48
African American	0.63
Hispanic	0.62
Caucasian	0.46

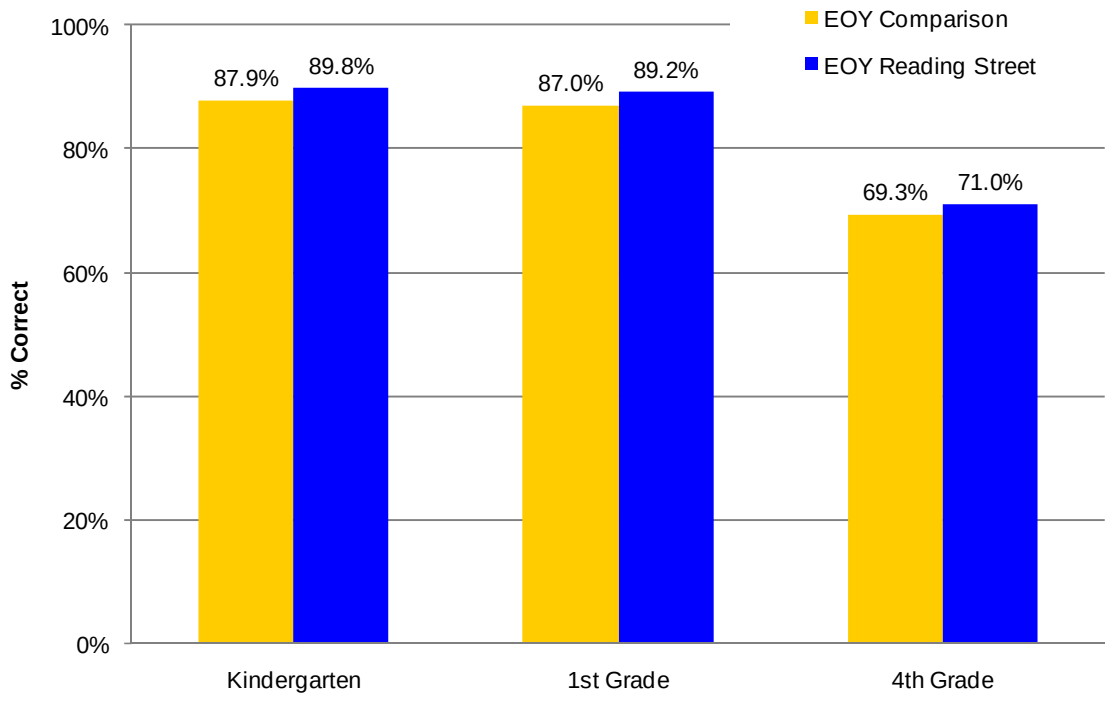
1. effect size = baseline to EOY Reading Street group gain / comparison sample standard deviation

A.2 Figures 2 Through 7

Appendix 2 presents Figures 2-7 which depict the Reading Street and comparison model adjusted group means for GRADE subtest scores, as well as DIBELS scales.

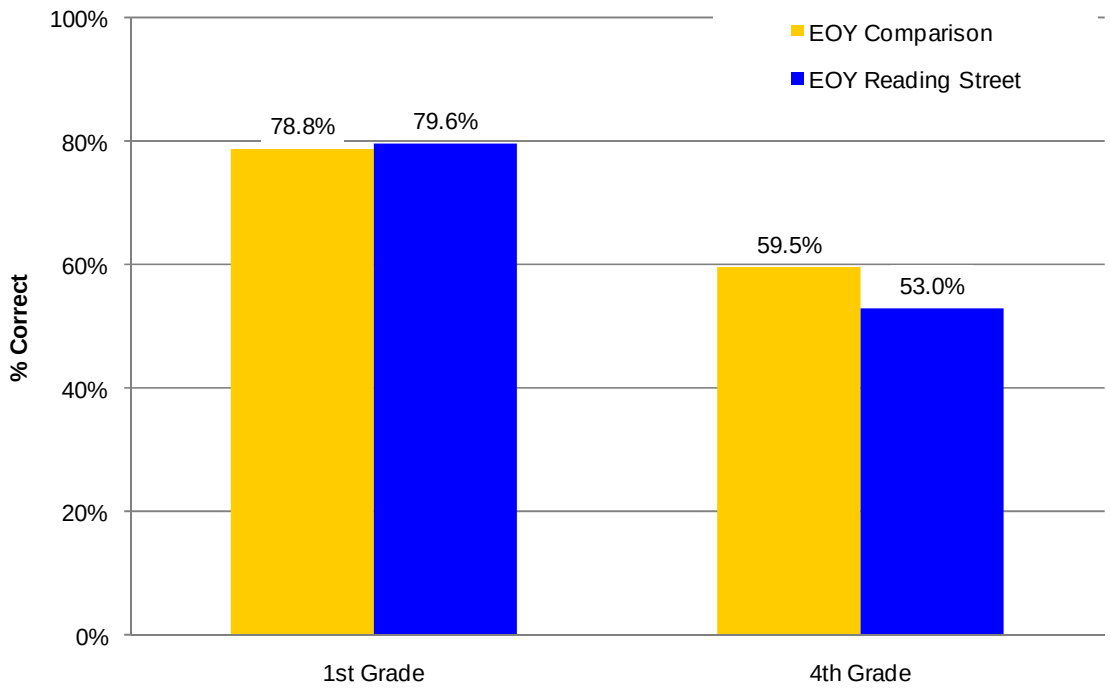
It should be noted here that the estimated adjusted group mean differences are the most meaningful statistic in these figures. Adjusted group mean differences are calculated holding the effects of confounding factors (i.e., baseline scores, student demographic information, and classroom environment indicators) constant for both groups effectively removing their influence from the results. In addition to the initial random assignment, the equating of confounding factors and maintaining consistent implementation ensures that differences found in the study groups may more confidently be attributed to the study conditions. The bars for the comparison group represent the expected score after setting each continuous covariate to its sample mean and equally weighting the levels for each categorical covariate. This expected score is not necessarily more valid than the sample mean or some other calculated score since actual weights and means for a specific target population are not included and the sample itself is not randomly drawn from any specified population.

Figure 2: GRADE Word Study



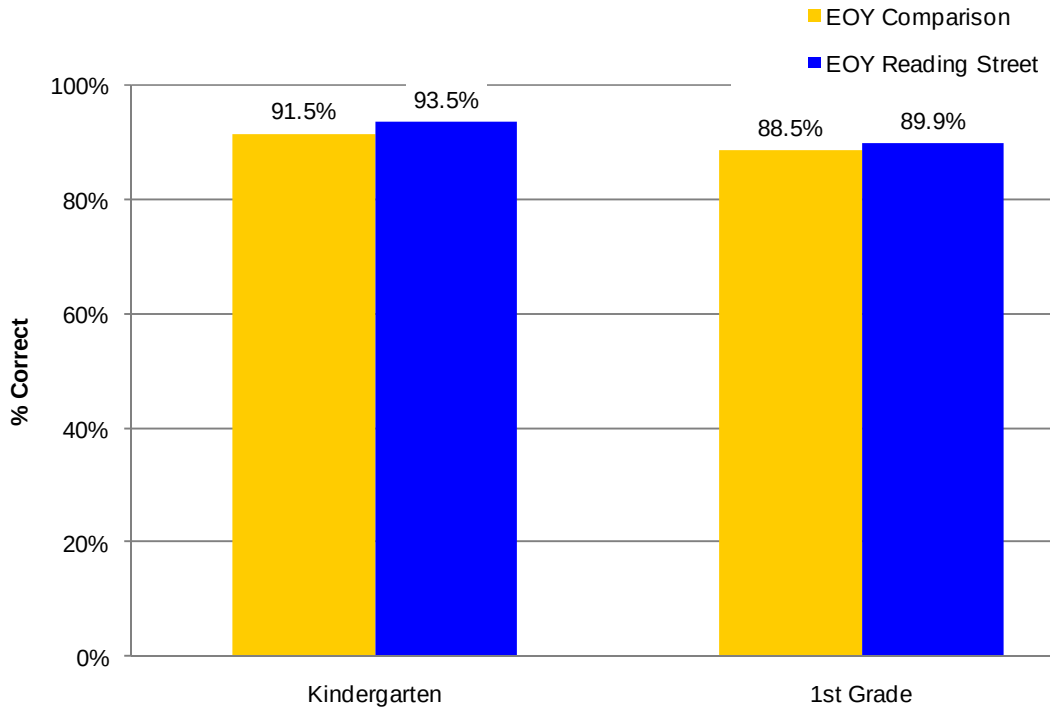
After adjusting for student and classroom characteristics, Kindergarten, 1st and 4th grade Reading Street students outscored their comparison group counterparts by 1.9% (SE=1.32%), 2.2% (SE=1.61%) and 1.7% (SE=1.94%) respectively.

Figure 3: GRADE Reading Comprehension



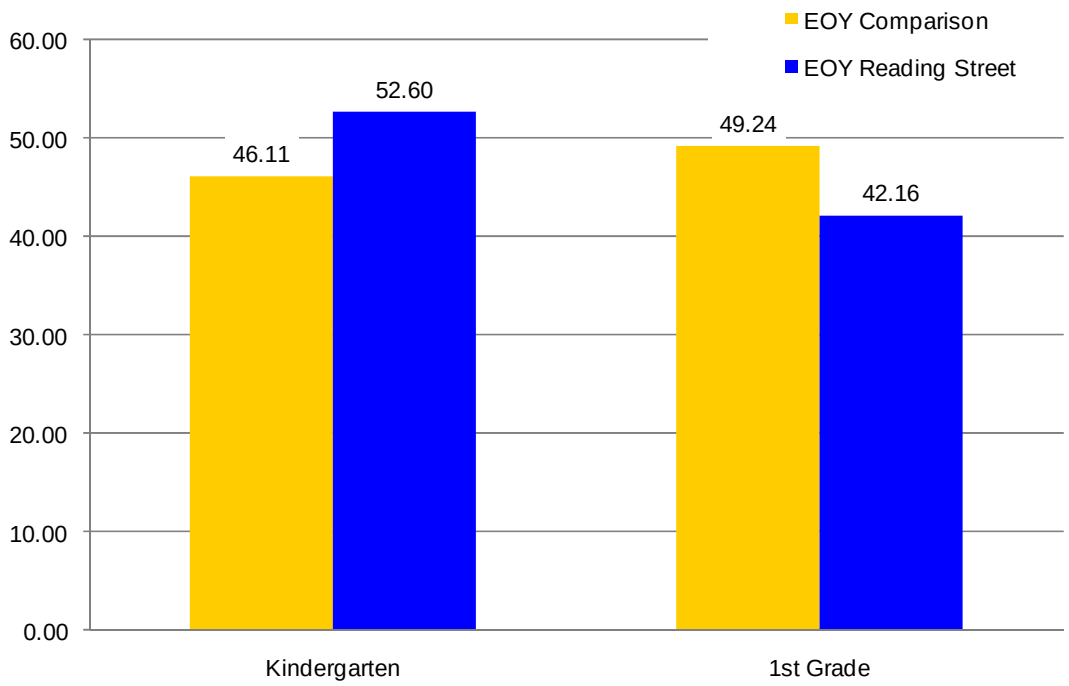
After adjusting for student and classroom characteristics, 1st grade Reading Street students outscored their comparison group counterparts by 0.8% (SE=3.81%). The 4th grade comparison group students outscored their Reading Street counterparts by 6.5% (SE=2.72%).

Figure 4: GRADE Listening Comprehension



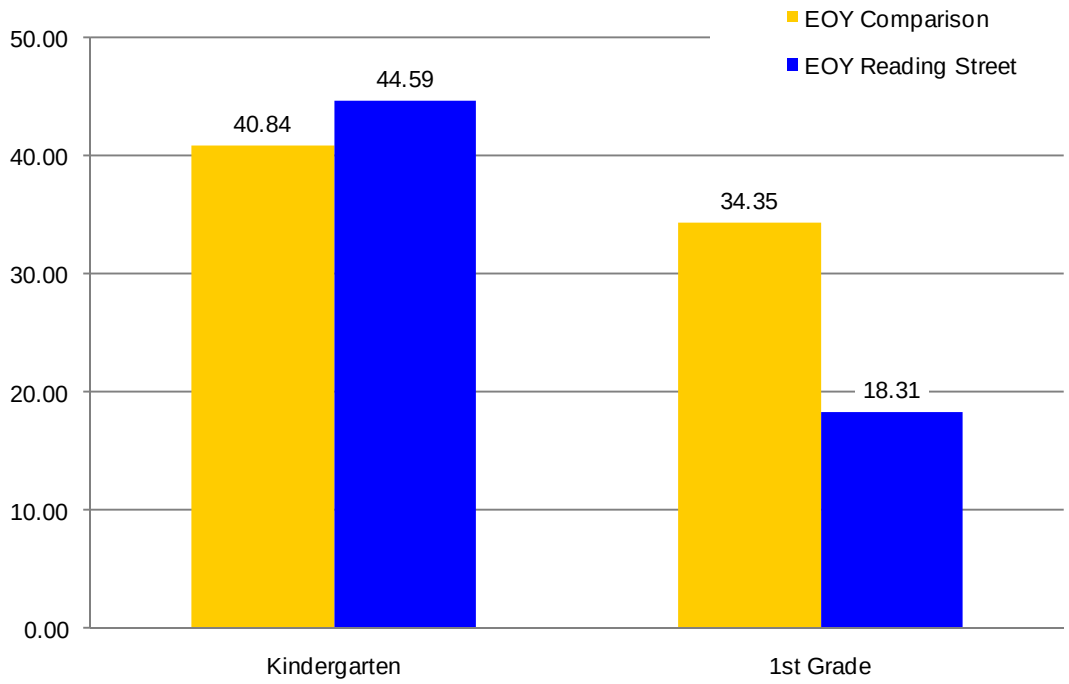
After adjusting for student and classroom characteristics, Kindergarten & 1st grade Reading Street students outscored their comparison group counterparts by 2.0% (SE=0.71%) and 1.4% (SE=0.65%) respectively.

Figure 5: Phoneme Segmentation Fluency



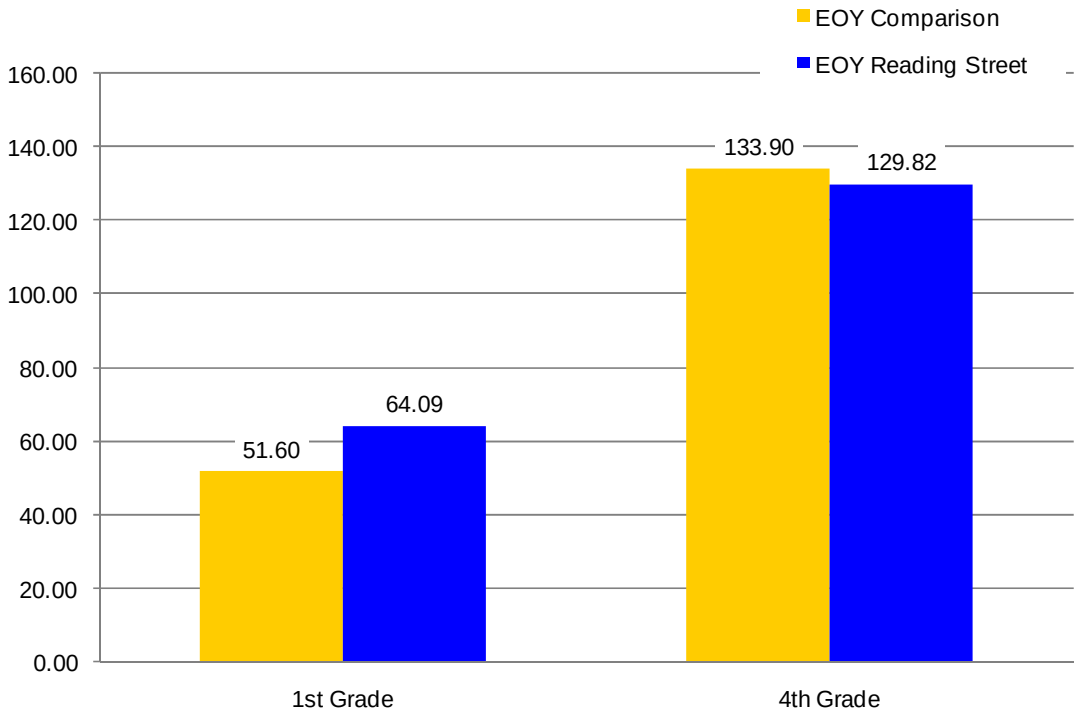
After adjusting for student and classroom characteristics, Kindergarten Reading Street students outscored their comparison group counterparts by 6.49 (SE=2.33). The 1st grade comparison group students outscored their Reading Street counterparts by 7.08 (SE=0.27).

Figure 6: Nonsense Word Fluency



After adjusting for student and classroom characteristics, Kindergarten Reading Street students outscored their comparison group counterparts by 3.75 (SE=1.85). The 1st grade comparison group students outscored their Reading Street counterparts by 16.04 (SE=1.57).

Figure 7: Oral Reading Fluency



After adjusting for student and classroom characteristics, 1st grade Reading Street students outscored their comparison group counterparts by 12.49 (SE=0.52). The 4th grade comparison group students outscored their Reading Street counterparts by 4.08 (SE=0.76).