June 30, 2012

Washington Striving Readers

Year 1 Evaluation Report

education northwest
CREATING STRONG SCHOOLS & COMMUNITIES

Washington Striving Readers
Year 1 Evaluation Report

June 30, 2012

Theresa Deussen, Ph.D.
Caitlin Scott, Ph.D.
Kari Nelsestuen
Angela Roccograndi
Ann Davis

education northwest
101 SW Main Street, Suite 500
Portland, OR 97204
educationnorthwest.org

About Education Northwest

Education Northwest (formerly Northwest Regional Educational Laboratory) was founded more than 40 years ago as a nonprofit corporation. Our mission is to improve learning by building strong schools, families, and communities. We draw on many years of experience designing and conducting educational and social research, as well as providing consultation for a broad array of research and development efforts. One of our particular areas of focus is the evaluation of literacy initiatives.

We are located in downtown Portland, Oregon, but much of our work takes us around the five state Northwest region (Alaska, Idaho, Montana, Oregon, and Washington). We also conduct work in other states and on national projects.

Contact

Education Northwest
101 SW Main, Ste. 500
Portland, OR. 97204
http://educationnorthwest.org
503.275.9500

Project Lead

Theresa Deussen

Cover photo copyright: Dee Dixon

Executive Summary

In 2009, the U.S. Department of Education conducted a competition for a second round of Striving Readers grants. Its dual purpose was to:

- Raise middle and high school students' literacy levels in Title I–eligible schools with significant numbers of students reading below grade level.
- Build a strong, scientific research base for identifying and replicating strategies that improve adolescent literacy skills through a required experimental study design.

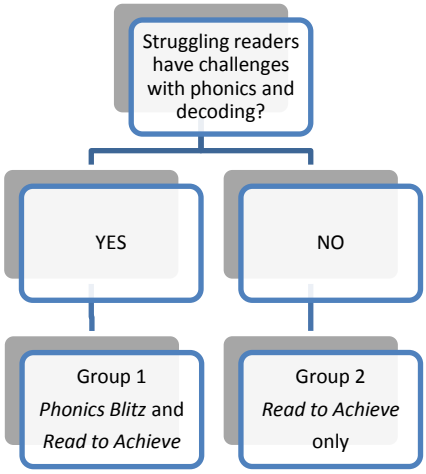
The competition invited states to adopt an intervention program designed to improve the reading of struggling adolescent readers. It required that the intervention be implemented in 10 or fewer middle or high schools and evaluated using an experimental design. The Office of Superintendent of Public Instruction (OSPI), Washington's state education agency, joined together with evaluators at Education Northwest to submit a proposal for the competition. Washington state was one of just eight states to be awarded Striving Readers grants in the second round.

The grant originally included a planning year, followed by three years of implementation in selected schools. However, Congress eliminated the funding for the program in spring 2011, three-quarters of the way through the first year of implementation. Existing funding was sufficient to complete the first year of program implementation and data collection, but the second and third years of implementation did not take place. Therefore, this Year 1 evaluation report is the only report about the program's implementation and outcomes.

The Washington Striving Readers Intervention

Each state that won a Striving Readers grant had to select an intervention to serve struggling readers (defined as students reading at least two years below grade level). In Washington, state project staff decided to focus on middle school and designed a program that took one of two different forms, depending on students' reading skills and specific challenges. As illustrated in Figure 1, Group 1 consisted of students who had difficulty with phonics and decoding, where they spent the first part of the year working in the *Phonics Blitz* program (Really Great Reading, 2010). They then moved into the *Read to Achieve* program (Marchand-Martella & Martella, 2010), a program that concentrated on vocabulary and reading comprehension strategies. Group 2 consisted of students who did not need phonics intervention; they spent the entire year working in *Read to Achieve.*

Figure 1



Eligibility of Struggling Readers for One of Two Forms of the Striving Readers Intervention

Evaluation Design

As the evaluators, we worked closely with state project staff and intervention program authors in the planning year to design the study and ensure its fit with program goals. This included evaluating multiple components of the program's *implementation*: the degree to which teachers received the intended professional development and in-class support to deliver the two reading programs; fidelity of implementation (whether teachers taught the programs as the program authors intended); and lesson completion (a measure of the amount of material covered). It also included an evaluation of the *impact* of the Washington Striving Readers program using experimental methodology. This meant that students eligible for the intervention were randomly assigned to either receive the intervention or to participate in a control condition. State project staff deliberately selected schools that were not already offering reading interventions to struggling students, so that being in the control condition was equivalent to the same experience students would have had in the absence of the Striving Readers grant. Students in the control condition took a study hall or elective, while students in the intervention group received a Striving Readers class, either a Group 1 or Group 2 class depending on their needs.

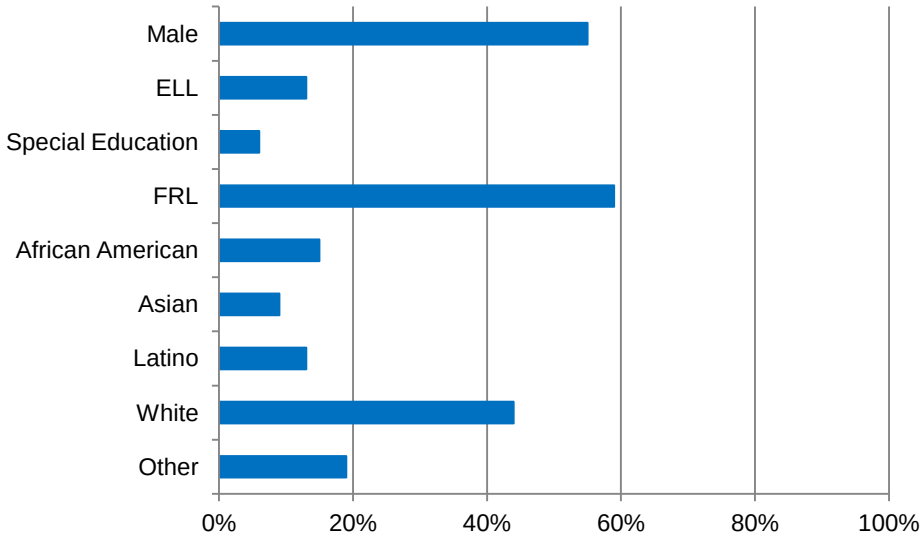
Participating Schools and Students

Five schools from three districts in Western Washington participated in Washington Striving Readers. All of the schools were eligible for Title I funding and served students living in poverty (between 45 and 64 percent of students were eligible for free/reduced-price lunch). In each school, the intervention was offered as a reading class in addition to students' regular English language arts class. All classes were taught by certificated teachers who were hired and trained specifically for Striving Readers. Four of the schools had one intervention teacher each, while the fifth school had two teachers. Class sizes were very small—9 or fewer students in Group 1 classes and 12 or fewer students in Group 2 classes.

Across the five schools, a total of 203 students began the intervention in fall 2010. Another 212 students were in the control condition. Since some students moved away over the course of the school year, there were 358 students by the time of the posttest (176 in the treatment condition and 182 in the control condition).

As Figure 2 illustrates, over half of the students in the study were male and eligible for free/reduced-price lunch (FRL, an indicator of socio-economic status). English language learners (ELLs) made up 13 percent of the sample, and 6 percent of the sample was eligible for special education (but did not have an Individualized Education Plan, or IEP, in reading). Figure 2 summarizes this information as well as data on the ethnicity of participating students.

Figure 2



Demographic Characteristics of Students in the Washington Striving Readers Impact Study (Treatment and Control Conditions Combined)

Implementation Findings

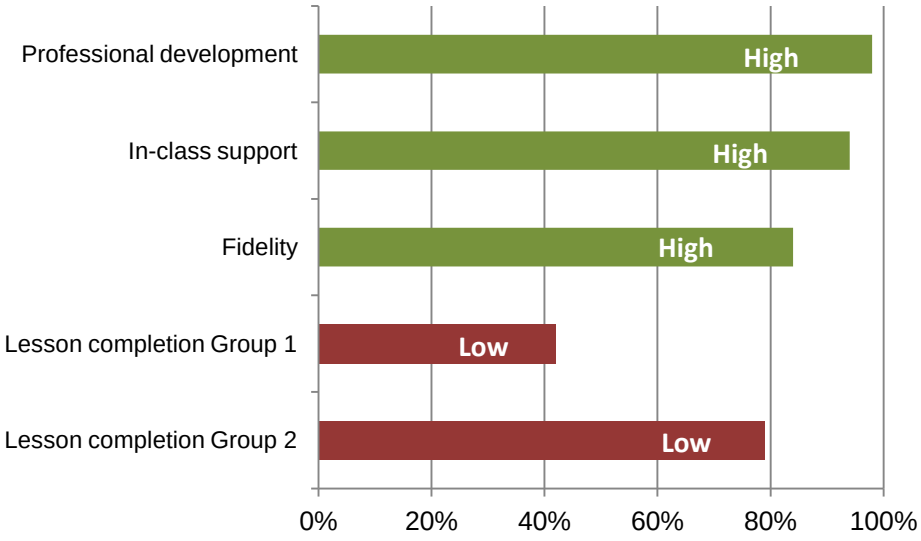
The program offered 70 hours of professional development for teachers and all teachers participated in at least 90 percent of these offerings. All teachers also received the intended amount of in-class support, defined as at least 12 visits from a project coach with each visit lasting at least one hour.

Fidelity of implementation, which was strongly encouraged by state project staff, was high for both the *Phonics Blitz* and *Read to Achieve* programs. We observed multiple classes taught by each teacher twice during the year, using program-specific protocols that we developed and piloted in consultation with the program authors. While some individual observations of teachers did not meet the standard for high fidelity, the overall fidelity average was 84 percent, which constituted high implementation.

As Figure 3 indicates, lesson completion was the only aspect of implementation that was not consistently high. No teacher was able to complete the 50 *Phonics Blitz* lessons in the first 12 weeks of school as was expected; in some classes, it took twice that long to finish the program. This also reduced the number of *Read to Achieve* lessons that could be completed in the remaining weeks of the school year for Group 1 classes. In Group 1 classes, students finished only 42 percent of the *Read to Achieve* material by the end of

the year, which corresponded to a "low" rating for lesson completion. In Group 2, where only *Read to Achieve* was taught, lesson completion varied more and averaged 79 percent, also a "low" rating. In some cases, this may have been due to overly challenging or even unrealistic pacing schedules. In some cases, however, it was at least partly due to teachers' misunderstandings of how much time they were supposed to devote to particular activities.

Figure 3



Implementation of Key Components of Washington Striving Readers

Impact on Student Achievement

We used three different assessments to measure impact. We used the *Gates-MacGinitie Reading Test* to measure reading comprehension. We used two subtests from the *Woodcock Reading Mastery* assessment—the word attack and word identification subtests—to measure decoding (alphabetics). Finally, we included scores from the *Measure of Student Progress*, or *MSP*, Washington's state reading assessment, as a measure of general literacy skills.

We examined the overall impact of Washington Striving Readers using a fixed effects regression model that accounted for the random assignment of students within schools and groups (Group 1 or Group 2). As Table 1 illustrates, we found statistically significant results only on the *MSP,* where students in the treatment condition made greater gains than those in the control condition. The effect size (Glass' *delta,* a measure of the magnitude of the impact on student learning) was 0.16. This is not a large impact but is comparable to the impacts found in a number of first-round Striving Readers sites (e.g., Faddis et al., 2010; Hamilton et al., 2011) and suggests that students in the treatment condition made some improvement in their literacy skills. Even with the improvement, however, on average students did not attain proficiency on the *MSP*.

On the *Gates-MacGinitie* and the two *Woodcock Reading Mastery* subtests, the small differences we found were not statistically significant.

In addition to examining the overall results, we also looked separately at results for Group 1 and for Group 2. Students in Group 1, who received the *Phonics Blitz* and *Read to Achieve* combined intervention, scored higher than students in the control condition on the *Gates-MacGinitie* and the *MSP*, but results were not statistically significant (effect size 0.13). We also found nonsignificant results for the *Woodcock Reading Mastery* subtests, although the effect size for the *Woodcock Reading Mastery* word attack subtest was larger (0.33). It is important to note that this was an especially small group, making it more difficult to find statistically significant findings.

Students in Group 2 received the *Read to Achieve* intervention for the entire year. Scores of students in the treatment condition on the *Gates-MacGinitie*, *Woodcock Reading Mastery* word attack, and the *MSP* were not significantly different from those of students in the control condition.

Table 1
Results of Washington Striving Readers Impact Study

Group	Sample size (n)	Effect size	Significant at p>=0.5?
Overall			
Gates-MacGinitie	358	0.03	No
Woodcock word ID	357	-0.04	No
Woodcock word attack	357	0.08	No
MSP	401	0.16	**Yes**
Group 1			
Gates-MacGinitie	63	0.13	No
Woodcock word ID	63	0.14	No
Woodcock word attack	63	0.33	No
MSP	76	0.11	No
Group 2			
Gates-MacGinitie	295	0.02	No
Woodcock word ID	294	-0.03	No
Woodcock word attack	294	0.07	No
MSP	325	0.16	No

Summary

The Washington Striving Readers program provided intensive in-school reading intervention to 176 middle school students who read significantly below grade level. The teachers who provided the intervention received the intended professional development and in-class coaching, and they delivered the intervention the way it was intended, with one exception: fewer lessons were completed than intended, meaning that students did not receive all of the content they were supposed to receive. This was particularly true for students in Group 1, who started the year with difficulty decoding.

The study was designed to combine results from three years in order to have a larger sample size and be able to detect effects of the intervention. Because Congress eliminated funding after the first year of implementation, our sample size was smaller than planned, making it less likely we would find significant effects. For the most part we found no significant differences between the scores of students in the treatment and in the control conditions. There was, however, a significant positive impact on the *MSP.* As noted earlier, the size of the impact was similar to that found in other Striving Readers programs. We also know that the effect size of the average annual gain of middle school students in

reading ranges is about 0.25 (Hill, Bloom, Black, & Lipsey, 2007), so that the gain made by students in the treatment condition was comparable to about five months' growth.[1] This improvement reduced the gap between low-performing readers and their peers who read at grade level, but did not close that gap. Students in the treatment condition still had average *MSP* scores that put them below the cut point to be considered "proficient" readers.

We also noted a much larger effect size for Group 1 on the *Woodcock Reading Mastery* word attack subtest. This finding, although not significant, is promising, and the impact of the Group 1 treatment (*Phonics Blitz* and *Read to Achieve*) on students' decoding skills may merit further study.

Originally the study was intended to continue for two additional years. We had hoped that those additional years would allow us to learn more about the impact on students as well as explore whether implementation changed and lesson completion improved when teachers had more experience with the programs. Cutting the study short meant that we were not able to learn everything we had hoped to about the Washington Striving Readers intervention. Nevertheless, there are meaningful lessons from this one-year study that can have important implications for those implementing similar interventions in the future.

For example, we learned that it is possible for teachers to attain a high level of implementation, even when teaching two new programs, within a few months of their introduction to the program. We also found, however, that it is important to attend not only to the fidelity of program implementation but to the amount of material taught during the year. When teaching new programs, teachers may need additional support to ensure appropriate pacing.

The findings also demonstrated that it is possible to make a statistically significant difference in struggling students' overall literacy achievement in the course of one school year. Students in the Washington Striving Readers intervention performed better on the state reading assessment than did students in the control condition, who did not receive any supplemental reading support. The gains made, however, were not sufficient to bring middle school students who read substantially below grade level up to a proficient level. In light of these and other findings (Vaughn et al., 2011), it may be that these students need more than a one-year intervention. A summer program and/or a second year in intervention might help students make additional progress.

[1] Hill, Bloom, Black, & Lipsey (2007) report an average annual gain in effect size of 0.32 for grade 5-6, 0.23 for grade 6-7, and 0.26 for grade 7-8, or an average of 0.27 across the three years. An effect size of 0.16 represents 59 percent of that gain, or about 5 months of a 9-month school year.

Acknowledgments

Many people came together to make this study possible. We'd like to thank Cheryl Young and Sarah Rich, both formerly of the Office of Superintendent of Public Instruction in Washington state, for their hard work and thoughtfulness in designing and then implementing the Washington Striving Readers project. From an early stage, program developers Linda Farrell and Michael Hunter (*Phonics Blitz)* and Nancy Marchand-Martella and Ronald Martella (*Read to Achieve*) gave graciously of their time and attention to make sure we understood their programs and created an adequate tool to measure implementation. Evelyn Probert and Pam Cavanee, project coaches, were patient with our repeated queries into what they were doing and how they were doing it.

We'd also like to thank Carolyn Moilanen and Jane Arkes, who were on site at the schools to ensure that the system to assess hundreds of middle-schoolers in a short period of time ran smoothly without missing anyone. Of course, that incredible feat would not have been possible without the logistical mastermind of Ann Davis at Education Northwest and the cheerful and tolerant support of the principals, teachers, instructional aides, and librarians at the five schools where the study took place.

Makoto Hanita at Education Northwest, Anne Wolf from Abt Associates, and Ryoko Yamaguchi at Plus Alpha Consulting (formerly at Abt Associates) all provided crucial feedback on the design and analysis of the impact study. We are grateful for their methodological expertise. Denise Crabtree and Helen Davis at Education Northwest provided patient assistance with graphics and formatting.

Above all, our thanks go to the six Striving Readers teachers whose work lives we intruded upon for an entire school year. They let us watch them at professional development, observe them teach, assess their students, and ask them question upon question. They demonstrated a good-natured willingness to be thoroughly inconvenienced in order that we all could learn more about how to help adolescents become better readers.

CONTENTS

LIST OF TABLES

List of Figures

Chapter 1
Introduction and Program Description

In 2009, the U.S. Department of Education conducted a competition for a second round of Striving Readers grants. The first round, funded in 2006, had provided districts with funding to strengthen comprehensive, schoolwide approaches to adolescent literacy in schools with significant numbers of students reading below grade level. In addition, these grants included funding for interventions for struggling readers. For the 2009 competition, the grants funded only the intensive interventions, and not the comprehensive, schoolwide approaches.

There were also two other ways in which the 2009 Striving Readers competition differed from its predecessor in 2006. First, while the 2006 competition awarded grants to districts, the 2009 competition was only open to state education agencies. Second, while the 2006 competition funded five-year Striving Readers projects, the 2009 competition provided for four years of funding: a planning year, followed by three years of implementation.

According to the Department of Education, the second cohort of Striving Readers had two purposes:

- Raise middle and high school students' literacy levels in Title I–eligible schools with significant numbers of students reading below grade levels
- Use an experimental study design to build a strong, scientific research base for identifying and replicating strategies that improve adolescent literacy skills (http://www2.ed.gov/programs/strivingreaders/index.html)

Because building the research base was as important a purpose as improving adolescent reading, the competition required that applicants partner with a research organization that would conduct and use an experimental design to evaluate the program. The Office of Superintendent of Public Instruction (OSPI), Washington's state education agency, partnered with evaluators at Education Northwest to submit a proposal for the competition. Eight states were awarded grants in the second cohort, including Washington state.

Although the second round was originally supposed to be implemented in schools for three years, the funding for the program was eliminated by Congress in spring 2011, three-quarters of the way through the first year of implementation. Existing funding was sufficient to complete the first year and collect end-of-year data, but the second and third years never took place and the evaluation was discontinued. That is why this evaluation report, with its focus on the first year of Washington Striving Readers, is the only report about the program's implementation and outcomes.

In this chapter, we provide a detailed description of the Washington Striving Readers program—the intervention programs, professional development and in-class support for teachers, anticipated class sizes, and lesson completion. Our logic model depicts the contribution of the different program components. At the end of the chapter, we provide a short overview of the study. Later chapters of the report describe the methods in greater detail.

Differentiated Intervention

Washington Striving Readers was designed to provide differentiated intervention to struggling readers and drew on two different intervention programs, *Phonics Blitz*, second edition (Farrell & Hunter, 2007) and *Read to Achieve* (Marchand-Martella & Martella, 2010). State project staff intentionally designed the overall program so that eligible students could be assessed and then matched with an intervention that best met their needs. Both interventions were implemented by six teachers in five middle and junior high Title I schools in Western Washington in 2010–2011.

Phonics Blitz. *Phonics Blitz* includes explicit instruction in phonemic awareness, phonics, and fluency for students who have fallen behind grade level in these skills. The second edition of *Phonics Blitz* includes 50 teacher-led lessons with sequenced activities in three areas:

- **Phonemic awareness**. Students learn to identify and segment each of the phonemes in spoken one-syllable words. Students practice segmenting phonemes orally in each lesson, a skill that becomes the basis for learning to decode. Phonemic awareness instruction also teaches students to explicitly identify and categorize vowel sounds, regardless of their spelling.
- **Phonics**. Students first learn to read single syllable words starting with short vowels in closed syllables and quickly proceed to read multisyllable words with closed syllables and schwa. Students then move to words with consonant-le, r-controlled vowels, open syllables, silent e, and vowel teams. Lessons explicitly teach the three sounds of suffix –ed and hard and soft c and g. Spelling conventions are also taught.
- **Fluency**. At the beginning of each lesson, students read a nondecodable passage aloud for 1 minute while other students mark their error(s). The majority of passages are expository. Readers track their accuracy and words correct per minute on a tracking chart. Once students read consistently with 98% accuracy, they are encouraged to increase their reading rate while maintaining accuracy.

Because it is primarily a decoding program, there is no explicit vocabulary instruction in *Phonics Blitz*, although the lessons do include content-area vocabulary such as "continental," "subtropical," and "octagon." At the beginning of each lesson, up to five vocabulary words are previewed with student-friendly definitions.

Similarly, *Phonics Blitz* does not focus on comprehension skills, although it does include literal comprehension questions about the passages students read where the students must find the answers to the questions in the passage. Students' writing assignments in this program are in response to these comprehension questions.

Teachers using *Phonics Blitz* are guided by a teacher's edition and accompanying materials. Students work from two "*Blitz*" books, a fluency passage book, and hands-on manipulatives, which include large and small letter tiles, "syllaboards," and magnetic white boards.

Each lesson has up to seven activities: oral reading, phonemic awareness, phonics concepts, word sort, detective work, words to read, and sentences to read. All students participate in all of these activities as a whole group, with opportunities for individual responses and some partner work. *Phonics Blitz* emphasizes a fast pace throughout the lesson to keep students involved. The program also offers suggested hand motions and specific language for teachers to use when they demonstrate lesson components in order to build consistency for students and to minimize long explanations from the teacher. The program directs teachers to use positive error correction when students make a mistake,

meaning that the teacher tells the students what they did correctly before guiding them to correct the mistake.

Read to Achieve. *Read to Achieve* emphasizes comprehension strategies, vocabulary strategies, fluency strategies, and higher order thinking skills. There are two components to its curriculum, content-area reading (25 units of 5 lessons each) and narrative reading (15 units with 5 lessons each). The three main emphases in both components are:

- **Comprehension**. The program uses five approaches to building students' comprehension skills:
  - Text connections such as identifying topics, purpose for reading, and prior knowledge
  - Text structures for various expository and narrative texts
  - Comprehension monitoring strategies such as rereading and pace
  - Note-taking strategies such as "SQ3R" (survey, question, read, recite, review)
  - Metacognitive strategies such as think-pair-share activities, graphic organizers, summary, and other strategies
- **Vocabulary**. Instruction includes decoding multipart words and word-learning strategies. Students learn to focus on specific words that are bold and highlighted in text and learn to use dictionaries, the glossary, and, when available, online tools.
- **Fluency.** Students engage in both oral and silent fluency reading and monitor their own progress. This includes cold and hot fluency timings where students record their words correct per minute (wcpm). Between the cold and hot timings are opportunities to practice reading the passage, work on multisyllabic words within the passage, answer questions about the passage, and write about or illustrate what they have learned.

Read to Achieve also includes opportunities for extended discussion of text meaning and interpretation, moving from teacher-led to student-led discussions over time. Questions posed to the group, partners, or individuals give students opportunities to discuss the text. For example, a question in the content-area program is, "How do metamorphic rock forms help scientists understand geological change on Earth?" A suggested question in the narrative program is, "Why was Gage weakened by the medicine Harlow gave him to bring his body into balance?"

In each unit, lessons move from activities in which teachers provide strong support to activities with more moderate levels of teacher support and, eventually, activities that students engage in independently or with a partner and without teacher support. In later units, activities that were first introduced with strong teacher support are revisited with lower levels of support, so that students have many opportunities to practice. The program incorporates student self-assessment, small group collaboration, and opportunities for both group and individual responses. Teacher materials are written with a "soft script." Teachers are expected to follow the intent of each step of the lesson but may not need to read the instructions word-for-word to students.

Washington Striving Readers Logic Model

Washington Striving Readers had a well-delineated logic model (Figure 1.1). The first program input is a systematic placement of students into two groups: Group1 students need specific help with decoding skills and are placed in *Phonics Blitz* before beginning *Read to Achieve*. Group 2 students receive a full year of instruction in *Read to Achieve* only. Chapter 4 describes student eligibility and placement in detail.

The logic model (see page 5) also specifies plans for teacher professional development and in-class support, class size, and lesson completion. These aspects of Washington Striving Readers are described in detail below.

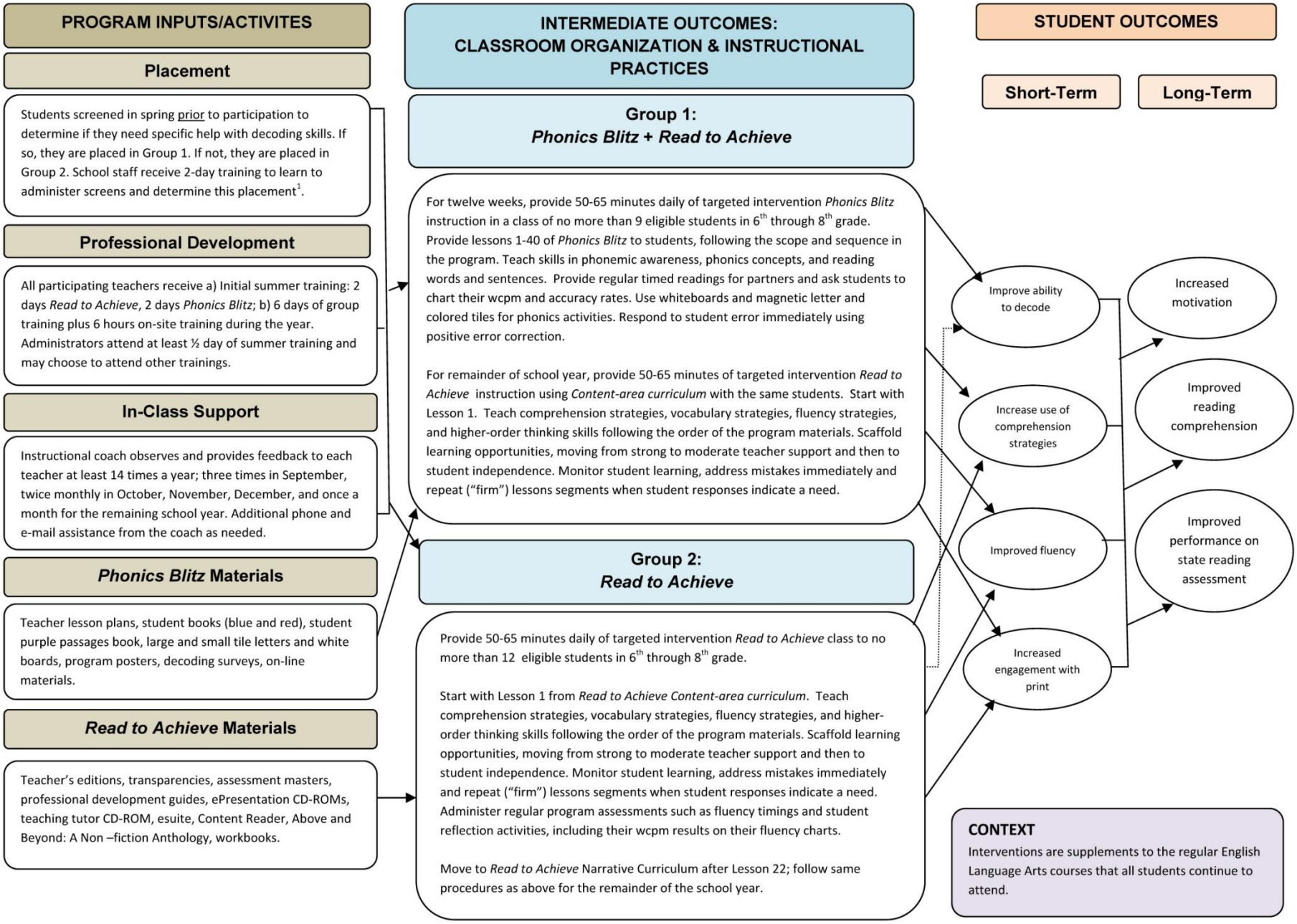
Professional development for teachers. The logic model describes an initial summer training for teachers, six additional days of group training during the year, plus one six-hour on-site training. Striving Readers training for the teachers began with a four-day (28 hour) institute in August 2010. In the 2010–2011 school year, this model was executed as planned.

At a four-day summer training, teachers were introduced to the Striving Readers grant, its purpose and requirements, and then trained on the two programs. The program training was provided by two of the program developers: Linda Farrell for *Phonics Blitz* and Nancy Marchand-Martella for *Read to Achieve*. During the school year, there were six additional days, or 42 hours, of professional development. These trainings included one day dedicated to each program and one day to learn and understand how to incorporate AIMSweb assessment data in the classroom. Three days were devoted to *Language Essentials for Teachers of Reading and Spelling (LETRS)* training, a professional development program created by Louisa Moats that provides teachers with an understanding of how students learn to read and write and the instructional strategies best supported by research. Table 1.1 summarizes the professional development for Striving Readers teachers. In addition to the above group trainings, all teachers also received on-site training from the developer of *Phonics Blitz* in October 2010, in which she observed each teacher and provided immediate individual feedback.

Table 1.1
Professional Development Offered to Washington Striving Readers Teachers

Date	Hours	Training content
Aug. 9, 2010	7	Summer Institute: Overview of project and study (2 hours) from Project Director and *Phonics Blitz* training from program author (5 hours)
Aug. 10, 2010	7	Summer Institute: Continued *Phonics Blitz* training from program author
Aug. 11, 2010	7	Summer Institute: *Read to Achieve* training from program author: background and research
Aug. 12, 2010	7	Summer Institute: *Read to Achieve* training from program author: content area and narrative program
Oct. 7, 2010	7	*Phonics Blitz* training from program author (model lessons, review)
Nov. 17, 2010	7	AIMSweb assessment and data training from REACH educational consulting
Dec. 7, 2010	7	*Read To Achieve* training by program author
Feb. 8, 2011	7	*Language Essentials for Teachers of Reading and Spelling* (*LETRS*) training, Part I from Lisa Thompson
Feb. 9, 2011	7	*LETRS* training, Part 2
Apr. 26, 2011	7	*LETRS* training, Part 3
Total hours	**70**	

Figure 1.1



[1]Placement in Group 1 or Group 2 is part of the overall Washington Striving Readers model, and would be essential to the design even without a study that calls for random assignment.

Administrators were encouraged but not required to attend the first half-day of the summer training and were welcome to attend other trainings as they desired.

In-class support. Due to concerns that the group training, by itself, might not be sufficient to ensure the high levels of fidelity to the two programs that state project staff hoped to see, Washington Striving Readers also created ongoing implementation support in the form of coaching. Each teacher was supposed to meet with one of the two part-time coaches 14 times over the course of the year, with the support heavily concentrated in the early part of the year (three times in September; twice monthly in October, November, and December; and once a month for the rest of the school year). Coaches were expected to tailor their services to meet the needs of the teachers. They could, for example, observe lessons, model lessons, problem-solve with the teacher, or offer individualized assistance. Coaches were also available by phone and e-mail as needed.

In-class support was provided by two coaches with substantial prior experience in literacy. One of the coaches had previously taught *Phonics Blitz.* The other learned the program for the first time at the August 2010 training. Neither coach had previous experience with *Read to Achieve.* Both coaches also had more than five years of previous coaching experience, primarily with K–3 teachers implementing literacy programs.

The training and support for teachers was built on their existing expertise. All six teachers in the study were certificated teachers who were hired specifically to teach Striving Readers classes. Districts were provided with hiring guidance (e.g., preferred experience in scientifically based reading instruction with adolescents), but ultimately districts made their own decisions about which teachers to hire. Teachers, hired in spring or summer of 2010, knew when they accepted the position that they would be participating in a rigorous evaluation of the intervention.

Class size. Class sizes were designed to be small. Group 1 classes (*Phonics* Blitz followed by *Read to Achieve*) were intended to have up to nine students per class. Group 2 classes (*Read to Achieve* only) could have up to 12 students. In practice, all but one of the 31 classrooms adhered to these class size guidelines. One Group 2 classroom had 13 instead of 12 students. Some classes were very small; a few classrooms served only three or four students.

Most classes mixed students by grade level, with students in the sixth, seventh, or eighth grades receiving the same instruction.

Lesson completion. Striving Readers classes were designed to meet daily for the full school year. The program established pacing guides for each district, which teachers were expected to follow. According to the pacing guide, Group 1 students were supposed to complete the 50 lessons of *Phonics Blitz* in approximately the first 12 weeks of school before moving to *Read to Achieve* content curriculum. They were expected to reach unit 21 (of a possible 25) by the end of the year.

Group 2 classes were expected to cover the first 22 units of the content-area curriculum and the first nine units of the narrative curriculum by the end of the year. All estimates included time for assessments, teacher professional development days, and other activities that might disrupt the flow of instruction. (See Chapter 2 Methods for the Evaluation of Implementation for more details about pacing.) Four schools had traditional class periods, which met every day for about 45 minutes. One school with two teachers had a block schedule, where students came every other day for approximately 85 minutes.

Planned experiences for control students during intervention period. Students who were eligible for Striving Readers and assigned to the control group were to have a study hall or an elective class instead of the Striving Readers class. They were not slated to receive an additional reading class or tutoring.

Overview of the Study Design

The evaluation used an experimental design to test the impact of Washington Striving Readers on students' phonics, reading comprehension, and general literacy achievement. We randomly assigned eligible students in the sixth, seventh, and eighth grades to treatment or control conditions. Students in treatment conditions enrolled in a Washington Striving Readers class, which could take one of two possible forms, depending on whether they were in Group 1 or Group 2. They remained in the class for the entire school year. Students in the control condition enrolled in an elective or study hall.

We administered pretests to students within the first few weeks of school and posttests the following May. We used four outcome assessments across three reading domains: the *Gates-MacGinitie* reading comprehension test (comprehension domain) and the *Woodcock Reading Mastery* word attack and word identification subtests (alphabetics domain). We also analyzed student performance on the *Measure of Student Progress* (*MSP*), the Washington state reading assessment (general literacy achievement domain). With the exception of the state reading assessment, which was administered by standard district procedures, all outcome assessments were administered by the evaluation teams.

To understand the implementation of Washington Striving Readers, we measured both the delivery of the intended professional development to teachers, as well as the delivery of the intended interventions to students in the treatment condition.

To evaluate the delivery of professional development, we recorded teacher attendance at summer and school-year professional development sessions and calculated the percentage of total possible hours actually attended. We also documented the content of professional development, which is detailed later in this report. In addition, we collected information about the percentage of coaching sessions that teachers actually participated in, along with descriptions of how coaches worked with teachers.

To evaluate the delivery of the intervention, trained observers from the evaluation team conducted multiple classroom observations in each intervention classroom at two different points in the school year. In consultation with the program developers, we developed a separate observation protocol for each of the intervention programs. In addition, we collected information on the number of lessons from *Phonics Blitz* and *Read to Achieve* that were completed during the year and used those as an indicator of the amount of intended material to which students were exposed.

Organization of This Report

Chapter 2 describes in detail the development and use of measures to evaluate the implementation of Washington Striving Readers (the instruments themselves are provided in Appendix A). Chapter 3 then summarizes the results of the implementation evaluation. Next, Chapter 4 describes the methods for evaluating the impact of the program. Chapter 5 describes the analytic sample, provides the CONSORT flow diagrams, as well as the results of the impact evaluation. Chapter 6 summarizes conclusions and identifies lessons that can be drawn from this study.

Appendix A provides the instruments we used to evaluate implementation of the program.

We deliberately wrote this report with a practitioner audience in mind. We have tried to ensure that our discussion of the quantitative methods and analysis results will make sense to people who do not read technical reports on a daily basis. At the same time, we wanted to include sufficient detail for those readers who want them. For this reason, we moved some technical information out of the body of the report and into the appendices. Appendix B provides information about the baseline equivalence of groups—that is, data showing that randomization worked, and that students in the treatment and control groups were not significantly different from one another. Appendix C provides detailed statistics from the multilevel regression models we used to analyze the impact of Washington Striving Readers.

Chapter 2
Methods for the Evaluation of Program Implementation

Over the past decade, the field of education has seen a notable increase in the number of randomized controlled trials (RCTs) of educational programs and interventions. Yet, the investment in this time-consuming and expensive research has proven largely disappointing, as many studies have failed to find significant impact even of programs that seem initially promising. Recent studies of federally funded programs in a range of disciplines, including adolescent reading (Feldman, Schenck, Coffey, & Feighan, 2010), mathematics (Agodini et al., 2009), and technology use (Dynarski et al., 2007) found very small positive impact or no impact at all. The lack of more impressive findings has brought increased attention to the question of implementation, as educators ask whether the tested programs were really implemented as intended (Mahoney & Zigler, 2006; Penuel, Frank, Fishman, Sabelli, & Cheng, 2009). This is not always an easy question to answer since programs often include many components, and even the program designers cannot necessarily say which components are most crucial.

Recognizing the importance of measuring multiple components of implementation, we designed our implementation evaluation to address four primary research questions:

1. To what extent did teachers participate in professional development activities?
2. To what extent did teachers receive in-class support for implementation?
3. To what degree did teachers implement *Read to Achieve* and *Phonics Blitz* with fidelity?
4. To what extent did teachers complete all of the required lessons in *Read to Achieve* and *Phonics Blitz*?

Details about the measures of implementation used to address each of the four questions follow. We have included observation and interview protocols in Appendix A.

Measuring Professional Development

To address the first question about participation in professional development, we compared the number of professional development hours offered to the number of hours attended. We also observed much of the professional development, examined meeting materials, and interviewed teachers.

We documented the number of hours of professional development *offered* to teachers according to meeting agendas. To document the number of hours of professional development teachers *received* we collected teachers' sign-in records from each professional development session and recorded the number of attendance hours, by teacher and event, in a database. We then divided the total number of hours of professional development each teacher received by the total number of hours offered. As seen in Table 2.1, we translated the resulting percentage of professional development hours attended into ratings of high (≥90%), medium (70–89%), or low (< 70%). We set the cut points for these ratings in consultation with state project staff members.

Table 2.1
Professional Development Implementation Ratings

High implementation	Medium implementation	Low implementation
Teacher attended ≥90% of professional development hours	Teacher attended 70-89% of professional development hours	Teacher attended < 70% of professional development hours

To provide a more detailed description of the content of professional development sessions, we collected agendas and handouts from each session, and a member of our team attended most sessions.

Finally, we asked teachers to report their perceptions of the professional development they received. Two in-person teacher interviews were conducted by evaluators trained to use a semistructured interview protocol. Questions were developed in consultation with project staff members and in response to the project's professional development plan for the year. Interview data were analyzed using a content analysis process—identifying concepts found in interviewee responses. Responses were coded inductively based on emergent themes, drawing together common interpretations that yield a framework for interpreting responses (Creswell, 1998).

Measuring In-class Support for Teachers

Our second research question asked, "To what extent did teachers receive in-class support for implementation?" To answer this question, we compared the amount of in-class support time from coaches Striving Readers *intended* to provide and the amount of time *actually* provided. We also interviewed teachers to gather their perceptions of the quality of the in-class support.

To collect data, we used password-protected, online coaching logs, which the two coaches completed regularly. The logs included the date of each visit and the amount of time spent with each teacher. A "visit" was defined as a contact focused on instructional issues and/or student data for at least one hour. Coaches also indicated in the log which of nine possible activities they conducted during their visits (e.g., modeling instruction, analyzing data together).

We calculated the percentage of in-class support teachers received by dividing the total number of visits each teacher received from the coach by 14 (the minimum number of intended visits). We then turned the percentage into high, medium, and low ratings as shown in Table 2.2. These ratings were determined in consultation with program staff members and reflect only the quantity of in-class support received.

Table 2.2
In-class Support Implementation Ratings

High implementation	Medium implementation	Low implementation
Teacher visited by the coach at least 12 times, or ≥86% of intended coaching	Teacher visited by the coach 8–11 times, or 57–85% of intended coaching	Teacher visited by the coach <7 times, or <57% of intended coaching

Teacher interviews, described previously, included questions about teachers' perceptions of the quality of in-class supports based on the intended model.

Measuring Fidelity of Implementation

Our third research question was, "To what degree did teachers implement *Read to Achieve* and *Phonics Blitz* with fidelity?" To examine this question, we developed classroom observation protocols, conducted observations, and calculated fidelity ratings.

Developing classroom observation protocols. Our first step in developing classroom observations was to identify the critical components of each program. We did this through interviews with program authors and a review of program materials, including teacher manuals, student materials, and training materials. The first protocol drafts were piloted in four schools (two *Read to Achieve* and two *Phonics Blitz* schools). The program authors accompanied us to those schools and told us whether the lessons we observed had strong, medium, or weak implementation. We compared their rating with the numeric score we gave the same observation using our protocol. This testing led to further revisions and another round of pilot testing. It also provided us with an empirical basis for our eventual decisions about cut points for high, medium, and low ratings.

The final *Phonics Blitz* observation protocol included 50 descriptors across seven program components: oral reading, phonemic awareness, phonics, word sort, detective work, words to read, and sentences to read. The descriptors were program-specific operations such as: teacher uses correct error procedures; students always use fingers when stretching sounds; teacher states objective. The rating scale for each descriptor was (1) not very true of this lesson; (2) somewhat true of this lesson; and (3) very true of this lesson.

We calculated rater agreement during fall observations; five *Phonics Blitz* classes were scored by two separate observers. There was 100 percent agreement on the overall ratings of high, medium, or low fidelity.

The *Read to Achieve* protocol used three holistic rubrics to score fidelity of teacher activities and routines (ratings of 1–5), level of support (ratings of 1–3), and error correction (ratings of 1–3). These rubrics were applied to all lesson components: comprehension, vocabulary, comprehension with vocabulary, fluency, higher order thinking, and beyond the book. There were also four descriptive rubrics for rating hot and cold fluency timings.

We calculated rater agreement during fall observations; 12 *Read to Achieve* classes were scored by two separate observers. The agreement for *Read to Achieve* was 75 percent. Because we wanted higher levels of agreement, the four observers met to discuss what was most problematic about the protocol. Our consensus was to remove a measure called "firming" and to use three-point rubrics instead of five-point rubrics for level of support and error correction, resulting in the protocol described above.

In addition to program fidelity, we wanted to measure some of the overall classroom characteristics of both *Read to Achieve* and *Phonics Blitz* classrooms, such as student engagement and classroom climate. To develop these measures, we were influenced by other validated and widely used rubrics, such as the CLASS and ELLCO (Pianta, Karen, LaParo & Hamre, 2008; Smith, Brady, & Anastasopoulos, 2008). We developed seven rubrics, measured on a scale of 1–4, to examine the following characteristics: classroom climate, organization of materials, classroom routines, student engagement, addressing behavior problems, lesson pacing, and teacher monitoring. These measures were not designed to be part of the fidelity score, but rather to add context to the fidelity outcomes.

Conducting classroom observations. The four observers participated in a two-day training to prepare for classroom observations. The training included a detailed review of the program components, followed by a study of the protocol and application of the protocol to written lessons.

Observers visited each classroom twice, in October 2010 and February 2011, for a total of 46 observations. There were 13 observations of *Phonics Blitz* (8 in the fall and 5 in the winter) and 33 of *Read to Achieve* (17 *Read to Achieve* in the fall and 16 in the winter). Observations lasted the entire class period.

Calculating fidelity ratings. To calculate the fidelity ratings, we divided the number of total points given to the observation by the number of possible points. The percentages were then translated into a rating of high (≥75%), medium (50–75%) or low (<50%) fidelity. As noted above, we set the cut points for these ratings in consultation with program authors and trainers (see Table 2.3).

Table 2.3
Fidelity of Implementation Ratings

High implementation	Medium implementation	Low implementation
≥ 75% implementation as measured by observations	50–75% implementation as measured by observations	<50% implementation as measured by observations

Each teacher had between five and nine observations during the school year. To calculate an overall fidelity score for each classroom, we averaged each teacher's fidelity ratings across all of their observations and applied the same ratings as shown above in Table 2.3.

Measuring Lesson Completion

Our final question about implementation was, "To what extent did teachers complete all of the required lessons in *Read to Achieve* and *Phonics Blitz*?" To examine this question, we compared the number of lessons teachers reported completing to the number of lessons teachers were expected to complete.

Data about lesson completion were collected twice. At the end of week 12, the point when teachers were supposed to be finished with *Phonics Blitz* in Group 1 classrooms, teachers reported through e-mail what lesson number(s) they had reached. At the end of the year, Striving Readers coaches reported this information for each teacher.

We calculated lesson completion ratings by dividing the lessons teachers completed by the intended number of lessons. Group 1 *Phonics Blitz* teachers were expected to complete 50 lessons in 12 weeks, or just over four lessons per week. This was the standard for "high" lesson completion. For "medium" lesson completion, the cutoff was at least 43 lessons in 12 weeks, while "low" completion was anything less. By the end of the year, "high" implementation in *Read to Achieve* for Group 1 meant completing 21 units from the content-area curriculum, "medium" was 17–20 units, and "low" was fewer than 17 units.

For *Group 2 (Read to Achieve* only*)* a "high" was defined as completing 30 units (21 units from the content-area curriculum and the first 9 units from the narrative curriculum). According to the program developers, this matched a typical pace of four lessons per week. A "medium" level of completion was set at 25 units—21 from the content-area curriculum and at least 4 from the narrative curriculum. We scored the completion of fewer than 25 units as a "low" level of completion (see Table 2.4).

Table 2.4
Lesson Completion Ratings

	Intended completion rates	**High**	**Medium**	**Low**
Group 1, Part A (Phonics Blitz)	50 lessons	100% of lessons (50 lessons in 12 weeks)	86–99% of lessons (43–49 lessons in 12 weeks)	<86% of lessons (<43 lessons in 12 weeks)
Group 1, Part B (Read to Achieve)	21 units*	100% of lessons (21 units by year end)	81%–99% (17-20 units by year end)	<81% (<17 units by year end)
Group 2: Read to Achieve	30 units**	100% of lessons (30 units by year end)	83%-99% (25–29 units by year end)	<83% (<25 units by year end)

*The content-area curriculum includes 25 lessons, but the pacing guide for Group 1 made it only possible to reach Unit 21 by the end of the year.
**30 units include 21 units from the content-area curriculum, plus 9 units from the narrative curriculum.

In addition to these data, the in-person interviews in the fall and winter asked teachers about the pacing of the program, how it worked for them, and what they found challenging. As previously described, we conducted a content analysis of the interview data (Creswell, 1998).

Chapter 3
Results of the Implementation Evaluation

In Chapter 2 we laid out the methods we used to evaluate the level of implementation. This chapter reports on the results of the implementation evaluation for the 2010–2011 school year. Specifically, we report on the level of professional development received, the amount of coaching support, the level of fidelity to the instructional programs, and the rate of lesson completion. In addition, we summarize teacher comments about the challenges involved in implementing the two reading programs, and what supports they found helpful. We also briefly describe the experience of students in the control group Overall, teachers had high levels of participation in Striving Readers professional development and received high levels of in-class support from state coaches. Teachers also implemented both programs with high levels of fidelity. However, lesson completion rates (a measure of the amount of material covered) was low for Group 1 and varied from low to high for Group 2 classes.

Professional Development

Altogether, there were 70 hours of professional development offered to the Striving Readers teachers in Year 1. As described in Chapter 2, in our planning work with program developers and state project staff, we had previously determined that receiving at least 90 percent of those hours (63 hours) constituted a "high" level of implementation. All six teachers had high levels of participation in this professional development. Specifically, five of the six teachers attended all 70 hours of professional development that were offered; one teacher missed the first day of the summer institute but attended all other trainings (Table 3.1).

Table 3.1
Hours of Striving Readers Professional Development Received by Teachers

	Summer Training 2010			2010–2011	Total hours received	Percentage of PD received	Level of participation
	Overview	*Phonics Blitz* training	*Read to Achieve* training	Other professional development			
Possible hours	*2*	*12*	*14*	*42*	*70*	--	--
Teacher A	2	12	14	42	70	100%	High
Teacher B	2	12	14	42	70	100%	High
Teacher C	0	7	14	42	63	90%	High
Teacher D	2	12	14	42	70	100%	High
Teacher E	2	12	14	42	70	100%	High
Teacher F	2	12	14	42	70	100%	High

In addition to the participating teachers, the project director and two Striving Readers coaches attended all trainings, and often the Striving Readers coordinator for each school was present. A few trainings were also attended by building administrators, especially the first day of training in the summer.

In-class Support

To support implementation, a project coach was supposed to visit each teacher at least 14 times during the school year. In discussions with the project director, we defined "high" implementation of this component as teachers receiving at least 12 of those 14 visits.

In fact, all teachers received at least 12 visits, and some received more (Table 3.2). Support visits, which averaged 2 hours in length, amounted to between 23 and 28 hours of coaching per teacher during the school year. Visits were more frequent in the first half of the school year than in the second half.

Table 3.2
In-class Support Received by Washington Striving Readers Teachers

	Number of visits from coach	Total number of on-site coaching hours	Implementation level
Teacher A	16	28	High
Teacher B	13	28	High
Teacher C	14	23	High
Teacher D	14	23	High
Teacher E	12	29	High
Teacher F	12	27	High
OVERALL	81	158	High
Average per teacher	13.5	26.3	

During their visits to teachers, coaches reported their most frequent activities were observing *Read to Achieve* (31% of visits) and providing feedback on *Read to Achieve* (32%). Working with the teachers on data occurred in 22 percent of visits and providing "other information" occurred during 23 percent of visits. Coaches reported observing *Phonics Blitz* and providing feedback in just 16 percent of their visits. This lower percentage makes sense because *Phonics Blitz* was not slated to last for the full year. Coaches rarely reported modeling instruction for either program (2 percent of *Phonics Blitz* visits and 6 percent of *Read to Achieve* visits). Coaches were available by phone and e-mail, as well as present at all Striving Readers trainings, although these activities were not recorded as part of on-site support.

Fidelity of Implementation

In close consultation with program authors, we determined "high" fidelity of classroom implementation of *Read to Achieve* or *Phonics Blitz* meant scores of 75 percent or higher on observations. Teachers were strongly encouraged by the project staff to demonstrate high levels of implementation and make few, if any, modifications to the programs.

All teachers implemented both programs with high fidelity. As shown in Table 3.3, teachers had average fidelity scores between 77 and 92 percent; this was above the 75 percent cutoff established at the beginning of the program. These findings are based on 46 classroom observations during the 2010–2011 school year.

Table 3.3
Overall Teacher-level Fidelity of Implementation

	Number of *Phonics Blitz* observations	Number of *Read to Achieve* observations	Total number of observations	Average fidelity	Range in fidelity scores	Overall fidelity of implementation
Teacher A	4	3	7	88%	88–95%	High
Teacher B	3	5	8	88%	76–95%	High
Teacher C	0	5	5	78%	65–85%	High
Teacher D	1	8	9	92%	73–100%	High
Teacher E	1	7	8	83%	79–94%	High
Teacher F	4	5	9	77%	50–93%	High
OVERALL	13	33	46	84%	50–100%	High

We also calculated the fidelity levels for each program separately. For *Phonics Blitz*, the overall level of fidelity of implementation was high for all teachers (average 88 percent fidelity, Table 3.4). Each of the 13 observations, which received ratings between 76 and 96 percent, reached high fidelity.

Table 3.4
Teacher-level Fidelity of Implementation for Phonics Blitz

	Number of observations	Fidelity average	Range	*Phonics Blitz* fidelity level
Teacher A	4	91%	88–94%	High
Teacher B	3	80%	76–83%	High
Teacher D	1	96%	–	High
Teacher E	1	82%	–	High
Teacher F	4	89%	83–93%	High
OVERALL	13	88%	76–96%	High

Within the observations of *Phonics Blitz* we also examined each of the seven lesson components. Oral reading, a component in which students read a fluency passage and record their words correct per minute and accuracy rate, had the highest average fidelity rating (95%). The average fidelity rating for both the phonemic awareness and phonics components was 86 percent. For the remaining components of the lessons, the average fidelity rating fell between 63 and 85 percent. Only the sentences to read component had less than a high fidelity rating (see Table 3.5).

Table 3.5
Fidelity Ratings for Seven Components of Phonics Blitz

Lesson component	Number of observations	Fidelity average	Range	Fidelity level
Oral reading	10	95%	71–100%	High
Phonemic awareness	10	86%	75–98%	High
Phonics	12	86%	67–100%	High
Word sort	6	85%	67–100%	High
Detective work	4	83%	67–100%	High
Words to read	3	82%	67–100%	High
Sentences to read	3	63%	56–67%	Medium

Levels of fidelity of implementation for *Read to Achieve* were also high for five of the six teachers; the sixth teacher had an average rating of 67 percent, which translates to medium fidelity. The lowest fidelity score for any single observation was 50 percent, while the highest was 100 percent (see Table 3.6).

Table 3.6
Teacher-level Fidelity for Read to Achieve

Teacher	Number of observations	Fidelity average	Range	***Read to Achieve*** **fidelity level**
Teacher A	3	84%	63 to 95	High
Teacher B	5	93%	85 to 97	High
Teacher C	5	78%	65 to 89	High
Teacher D	8	92%	73 to 100	High
Teacher E	7	83%	76 to 94	High
Teacher F	5	67%	50 to 80	Medium
OVERALL	33	82%	50 to 100	High

Lesson Completion

The sole aspect of implementation that was rated "low" was lesson completion: a measure that compares the amount of material that teachers were intended to cover and the amount of material they actually covered during the year. To reach "high" implementation, teachers were expected to cover all of the intended material. That is, Group 1 teachers would cover all 50 *Phonics Blitz* lessons in 12 weeks and reach *Read to Achieve* unit 21 by the end of the year, while Group 2 teachers would reach *Read to Achieve* unit 30 by year's end.

In Group 1 classes, all teachers received "low" lesson completion ratings. *Phonics Blitz*, which was designed to be covered in 12 weeks, took up to twice as long to teach. As shown in Table 3.7, none of the teachers was able to complete the program by week 12, and on average, they had completed only half of the intended lessons by week 12. After week 12, they continued teaching *Phonics Blitz* until all 50 lessons were taught, but this left them fewer remaining weeks to teach *Read to Achieve*. At the end of the year, teachers had completed between 29 and 57 percent of the *Read to Achieve* material that the program intended for them to cover in their Group 1 classes.

Table 3.7
Group 1 Teacher-level Lesson Completion

	Completion rate of ***Phonics Blitz*** at 12 weeks	Completion rate of ***Read to Achieve*** at year end	Overall level of lesson completion rating
Teacher A	36%	29%	Low
Teacher B	60%	57%	Low
Teacher D	54%	29%	Low
Teacher E	64%	57%	Low
Teacher F	38%	38%	Low
OVERALL	50%	42%	Low

Note: Percentages are the amount of material teachers covered, divided by the amount of material they were expected to cover. One teacher did not teach any Group 1 classes and so is not included in this table.

Among Group 2 classrooms where *Read to Achieve* was the only program taught all year, lesson completion rates varied. Two teachers had "high" lesson completion rates, covering all of the intended material by the last week of school. One teacher had "medium" implementation, covering 90 percent of the material. Three teachers covered 63 to 81 percent of the material ("low" lesson completion), as shown in Table 3.8.

Table 3.8
Group 2 Teacher-level Lesson Completion

	Completion of *Read to Achieve* at year end	**Overall level of lesson completion rating**
Teacher A	63%	Low
Teacher B	100%	High
Teacher C	81%	Low
Teacher D	81%	Low
Teacher E	100%	High
Teacher F	90%	Medium
OVERALL	79%	Low

Overall Implementation

Table 3.9 summarizes the level of implementation across the program components. Professional development, in-class support, and the fidelity of both *Phonics Blitz* and *Read to Achieve* received high overall levels of implementation. Lesson completion for Group 1 and 2 was low.

Table 3.9
Summary of Implementation Levels by Teacher

	Professional development	**In-class support**	**Fidelity, *Phonics Blitz***	**Fidelity, *Read to Achieve***	**Lesson completion, Group 1**	**Lesson completion, Group 2**
Teacher A	High	High	High	High	Low	Low
Teacher B	High	High	High	High	Low	High
Teacher C	High	High	--	High	–	Low
Teacher D	High	High	High	High	Low	Low
Teacher E	High	High	High	High	Low	High
Teacher F	High	High	High	Medium	Low	Medium
OVERALL	High	High	High	High	Low	Low

In the remainder of the chapter, we report on implementation challenges and supports.

Phonics Blitz Implementation Challenges and Supports

Teachers' initial experiences and perceptions of *Phonics Blitz* varied. After the initial summer training, only two of the six teachers reported feeling prepared to teach the program. The other four teachers reported that it was only after they saw a full demonstration of the program being used with students in mid-October that they really understood how the program worked. They suggested that any future implementation include that kind of demonstration for teachers.

For two teachers, their sense of being initially under-prepared translated into reporting that the program was challenging or very challenging to teach at the beginning. By the second month of the program, the project director decided one of these teachers should not teach *Phonics Blitz* at all and reassigned her to teach only Group 2 classrooms, where she would only instruct from the *Read to Achieve* program. In contrast, another teacher with substantial elementary teaching experience and a deep background in phonics found the program easy to use. The other teachers fell somewhere in between, reporting in the fall that the program was "fairly easy" or "getting easier" to implement.

The follow-up *Phonics Blitz* training, conducted by one of the program developers in October 2010, was well-received by all of the participating teachers. All teachers rated it a "9" or "10" on a 10-point usefulness scale. Teachers explained it was useful because the developer had observed them teaching and provided specific feedback afterwards, and because they got to see an entire lesson for the first time.

All of the teachers knew they were behind in the *Phonics Blitz* pacing schedule. Several teachers asserted that lessons took 1.5 to 2 days rather than the one class period that was intended. This was also evidenced during observations; the full lesson was only completed in 2 of 13 observations. In one case, the oral reading component of a lesson took the entire class period, even though it was designed to take about 12 minutes. Other teachers said their pacing was behind schedule because the year started off slowly due to changing class rosters and the need to get all kids "on board." And finally, teachers reported that the actual number of instructional days was fewer than anticipated due to school scheduling and special school events that interfered with regular lessons.

While *Phonics Blitz* presented some challenges to teachers, students appeared to like the program. Observers recorded high levels of student engagement in *Phonics Blitz* lessons, giving an average rating of 3.6 on a 4.0 scale. We also observed high levels of material organization (3.9) and teacher monitoring of student work (3.8).

Read to Achieve Implementation Challenges and Supports

In contrast to their experience with *Phonics Blitz*, all six teachers reported that *Read to Achieve* was easy to implement and that they felt prepared to teach the program after the initial summer training since it was "straightforward" and the teacher materials were "teacher friendly" and "easy to follow." Like *Phonics Blitz*, observers gave high ratings to the material organization (3.8 out of 4.0), although slightly lower ratings for teacher monitoring of student work (3.1).

Although they said the program was easy to teach, teachers reported being frustrated by the inability to modify *Read to Achieve* to increase student engagement. Four of the six teachers said their students were "not engaged," "bored," and that the program was "too repetitive." Observation data also showed low levels of student engagement during *Read to Achieve* instruction with an average student engagement score of 2.8 on a 4.0 scale. Despite their frustrations with the program, teachers did not modify the content or format, since project staff had conveyed an expectation of strict fidelity to the content and format of the lessons. This was an important source of tension for several teachers.

Support from Project Coaches

In general, teachers were positive about the in-class support they received from the two coaches. They said coaches helped create a feeling of community, gave the program legitimacy in their schools, and were "helpful" in general. Some teachers said coaches provided useful suggestions for improving their instruction, such as encouraging more partner work. Two teachers, however, said the coaches had not really changed their instruction in any way.

Experiences for Control Students During Intervention Period

Besides describing what the Striving Readers intervention looked like, it was also important for us to know something about the counterfactual (i.e., the experience of the students in the control group). According to the study design, students in the control condition were not supposed to receive any supplemental reading instruction, although as with students in the treatment condition, they still participated in their regular English language arts class. In fact, the participating schools were initially selected in part because they reported that they did not provide any supplemental reading instruction to their struggling readers.

Between the time the project proposal was submitted and Striving Readers was actually implemented in the schools, more than a year had passed. During that time, three of the five grantee schools began offering some type of reading intervention. In two of the schools, principals agreed that eligible students who were assigned to the control group would not receive any intervention during the study period (some of these students were also eligible for math interventions and received those instead). We monitored class schedules of students in the control group the first week of school, in mid-November, and again in late February to ensure that those students were not in any supplemental reading class. A few students were assigned in either fall or spring semester to a reading class, but when we called that to the attention of the school, those students were moved to a study hall or an elective instead.

In the third school with reading interventions, however, a subset of students eligible for Striving Readers were placed in "LAP"classes, reading classes funded by the state Learning Assistance Program. After consultation with the school and state project staff, we removed a total of 23 students (12 assigned to the control and 11 to the treatment condition) from the study because of their involvement in LAP classes. Chapter 5 of this report on the random assignment of students provides more detail about the numbers of students removed from the study due to this other reading program.

In the end, no student in either the treatment or control group received any other supplemental reading class during the year of the study. It is possible that some of them, in either group, received after-school assistance or tutoring, but we do not know since we did not measure this.

We also explored the question of whether any of the instructional strategies and materials from Striving Readers might have been used in other classes. The Striving Readers teachers themselves only taught Striving Readers classes, so they did not deliver the intervention to any students outside the treatment condition. In interviews, we also asked the Striving Readers teachers if they had shared information about *Phonics Blitz* or *Read to Achieve* with other teachers. In all cases, Striving Readers teachers said they had not shared any information with the exception of very general comments such as "the students are doing well" or "we use AIMSweb testing." Given the amount of professional development involved in learning the two programs and the cost of the materials, we think it unlikely that students in the control condition had any exposure to either of the two programs.

Chapter 4
Methods for the Evaluation of Program Impact

To investigate the impact of Striving Readers on student learning, we used a randomized controlled trial or experimental design. In this design, students are randomly assigned to either the treatment condition, which receives the intervention being tested, or the control condition, which does not. Because assignment is random and not based on any student characteristics or preferences, we are then able to attribute any difference in outcomes to the intervention.

In this chapter, we first describe how the program identified eligible students and decided which form of the intervention (*Phonics Blitz* and *Read to Achieve,* or *Read to Achieve* only) was most appropriate. We also describe reasons some students were found ineligible. Next, we describe the plan for random assignment, the different outcome measures, and the procedures we used to collect outcome data. We also describe the analytical model we decided to use.

Identification of Eligible Students

The original grant competition in 2009 required states to design their Striving Readers programs to serve students who read two or more years below grade level. Our task, therefore, was to find an operational definition of "two or more years below grade level." For this, we turned to staff in the assessment department at OSPI, who determined that a score of 390 or below on the *Washington Assessment of Student Learning* (WASL, the state reading assessment at the time) was roughly equivalent to two years below grade level. This then became the initial basis for eligibility for Washington Striving Readers.

In practice, OSPI added two other eligibility requirements:

1) Students would not be on an Individualized Education Plan (IEP) for reading, because this would mean they must receive some sort of reading intervention and, therefore, could not be assigned to the control group.
2) Students would not be beginning (Level 1) ELLs, because developers of the intervention programs felt that a basic level of English was necessary in order to benefit from the intervention programs.

During the planning year, other eligibility requirements were added based on conversations with the developers of the two intervention programs. These revealed that both programs were designed for students with at least a minimum level of reading ability. Project staff at OSPI decided to set floors on student reading level in order to ensure that students were able to benefit from the intervention. Specifically, to benefit from *Read to Achieve,* developers explained that students should be able to read more than 100 words correct per minute (wcpm) with at least 90 percent accuracy. To benefit from *Phonics Blitz,* students could read more slowly (at or below 100 wcpm) but not below 71 wcpm for incoming sixth-graders and 75 wcpm for incoming seventh- and eighth-graders). They also needed to read with at least 88 percent accuracy. According to the developers, students who read with lower levels of accuracy needed a more basic decoding intervention that included support for reading "cvc words" (short words spelled with a consonant, vowel, and another consonant); *Phonics Blitz* presumes that students have already mastered this spelling pattern.

Putting all of these eligibility considerations into practice required a complex, multistage process of determining eligibility. In March 2010, we identified a pool of *potentially eligible* students. Students were potentially eligible for Striving Readers if they had a score of 390 or below on the 2009 WASL reading test (2010 results were not available in time to be used); if they did not have an IEP in reading; and if they were not a Level 1 ELL. We also considered students potentially eligible if they were missing a 2009 WASL reading score. At this point, we provided schools with information letters for parents of all potentially eligible students. The letters told them about the study and gave them the opportunity to opt out on behalf of their students. They could make this choice at any time, but we encouraged them to let us know, if possible, prior to screening or random assignment.

One concern we had was that students might have improved substantially between the 2009 state assessment and spring of 2010 and would no longer need interventions. It was also possible that students without a 2009 WASL reading score, perhaps because they had recently moved to Washington state or had missed the previous year's assessment, might not need intervention. Therefore, we confirmed the eligibility of potentially eligible students and checked the eligibility of students with missing scores using results from two screening instruments administered in spring 2009. The first was the *AIMSweb Curriculum-Based Measurement Reading Maze* (hereafter referred to as "Maze"), a test consisting of a paragraph in which every seventh word is omitted. Students are asked to choose which word, out of four options provided, makes the most sense to fill in the blanks. Students work individually to complete this test; the number of correct answers can then be transformed into a percentile ranking. The second was the AIMSweb reading *Curriculum-Based Measurement* (CBM), an assessment of oral reading fluency. The Maze yields a percentile ranking, while the CBM results in a raw score, a score for words read correctly per minute, an accuracy score, and a percentile ranking. We used the screening results both to confirm eligibility and to determine whether students would be in Group 1 or Group 2.

Thus, there were several purposes to screening:

- For students who had been two years below grade level on the 2009 state reading assessment, to ensure that they had not made such large gains in reading during the 2009–2010 school year that they no longer required intervention
- For students with no 2009 state reading assessment, to use the Maze as a substitute measure and thereby determine who read two years below grade level (defined as the 32nd percentile on the Maze)
- To exclude students who scored below the "floor"—below 70 wcpm (incoming sixth-grade students) or 75 wcpm (incoming seventh- and eighth-grade students) on the CBM—or whose accuracy was below 88 percent on the CBM, because they were considered to lack sufficient reading skills to benefit from the intervention
- For all eligible students, to determine whether they required decoding instruction and should be assigned to Group 1 or did not require it and could immediately move into *Read to Achieve* (Group 2)

These criteria are summarized in Table 4.1. These rather complex eligibility and assignment criteria were developed in an effort to respond to the publishers' description of the type and level of student for whom the two programs were most appropriate.

In spring 2010, the evaluation team trained teams of state, district, and school educators to administer the screening assessments. Those teams then screened 771 students in late spring 2010. Although educators

administered the assessments, only evaluators made eligibility decisions, and we based all decisions exclusively on the criteria described here.

There were a few students who were on our original list of potentially eligible students who were not screened, for a variety of reasons:

- **Withdrew prior to screening**. Students who moved between the determination of potential eligibility and when screening occurred were dropped from the list of potentially eligible students.
- **Not screened due to absence.** Although multiple make-up screening sessions were held, if students were absent from the make-up sessions as well, they were dropped from the list of potentially eligible students.
- **Not screened due to suspension**. Students who were suspended during screening and make-up screening were dropped from the list of potentially eligible students.
- **ELL level 1 status discovered late.** One student on the potentially eligible list was not screened because we discovered that the student was a Level 1 ELL student and should never have been placed on the list of potentially eligible students.
- **Other.** A few students were dropped from the list of eligible students at the request of the school because they were in the gifted program, had a severe anxiety disorder, and/or were currently being home schooled.

No students had to be removed from the group of eligible students due to parental requests to remove their students from the study. However, six parents of students who otherwise would have been potentially eligible returned the letter requesting that their students not participate. Because we received the letters before screening, these students were not screened.

Table 4.1
Eligibility and Assignment Criteria for Washington Striving Readers

Potential Eligibility		
Students were *potentially eligible* for Striving Readers if: 1. They did not have an IEP in reading AND 2. They were not a Level 1 English language learner AND		
They had a score of 390 or below on the 2009 state reading assessment	OR	They were missing the 2009 state reading assessment
Eligibility		
Potentially eligible students were screened to confirm eligibility. They were *eligible* if: 1. They were above the "floor," meaning their accuracy on the CBM was 88 percent or above and their raw score on the CBM was: a. Equal to or greater than 70 for incoming 6th-graders b. Equal to or greater than 75 for incoming 7th- and 8th-graders AND 2. They were below the "ceiling," defined as:		
For those with a 2009 state reading assessment: below the 51st percentile +10wcpm on the CBM and below the 51st percentile on the Maze	OR	For those missing the 2009 state reading assessment, Maze scores were comparable to two years below grade level, at or below the 32nd percentile on the Maze
Grouping Assignment		
Eligible students were placed in one of two groups. Students were placed in Group 1 (*Phonics Blitz* plus *Read to Achieve*) if: 1. Their CBM score was 100 wcpm or below OR 2. Their CBM score was over 100 wcpm but their accuracy on the CBM assessment was 88 or 89 percent		
Eligible students were placed in Group 2 (*Read to Achieve* only) if: 1. Their CBM score was greater than 100 wcpm AND 2. Their accuracy on the CBM was 90 percent or higher		

Once screening was completed, we went through the list of potentially eligible students and excluded students who were deemed ineligible for Striving Readers because their screening score was above the ceiling or below the floor.

Table 4.2 describes the various reasons students were removed from the study during the screening stage and reports the number of students removed. In summary, 771 students were identified by schools as potentially eligible based primarily on state test scores. After screening, 221 students were eliminated from the study. This left 550 students to be randomly assigned to treatment or control conditions.

Table 4.2
Numbers of Potentially Eligible Students Found Ineligible for the Study

Reasons for Ineligibility:	School 1	School 2	School 3	School 4	School 5	Total
Screening test scores too high	26	8	18	17	7	76
Screening test scores too low	9	2	17	16	5	49
ELL (level 1)	0	0	0	1	0	1
Withdrew before random assignment	14	17	14	24	7	76
Not screened due to absence	5	1	0	3	0	9
Not screened due to suspension	0	0	0	2	0	2
Other (e.g., gifted, anxiety disorder, home schooled although enrolled)	0	2	0	4	2	8
TOTAL	54	30	49	67	21	221

Random Assignment of Students

In spring 2010, we developed a plan for the random assignment of students to treatment and control conditions. To accommodate schools' need to create their master schedule over the summer, we arranged to conduct random assignment in June 2010. We planned to use a computer program to randomly assign students to treatment or control conditions within groups (eligibility for Group 1 or Group 2) and schools. At each school and in each group, we planned to assign 50 percent of students to treatment and 50 percent to control conditions. Once enough students were assigned to fill up treatment classes (since class sizes were deliberately small, this was a real possibility), we planned to create a waitlist. When there were odd numbers of eligible students at a school in Group 1 or 2 and when there was no waitlist, we decided we would have the additional student assigned to treatment rather than control. Because we used a computer program to create the lists of students in each condition, neither we nor the schools had any influence on which students were assigned to the treatment condition and which to the control.

Outcome Measures and Collection of Data

The research questions ask about the impact of Striving Readers on comprehension, decoding, and general literacy achievement. Four different student outcome measures provide information about the domains relevant to this study. Table 4.3 matches the outcome measures to the research question and specifies how and when we collected data.

Table 4.3
Summary of Outcome Data Collection

Research questions	Data source(s)	Administered to	When administered
Does Washington Striving Readers… help struggling middle school readers improve in comprehension?	*Gates-MacGinite* comprehension assessment	Group administered to all students	Fall (in first two weeks of school) and June (within last three weeks of school)
Does Washington Striving Readers… help struggling middle school readers improve in decoding?	*Woodcock Reading Mastery* word identification subtest *Woodcock Reading Mastery* word attack subtest	Individually administered to all students	Fall (in first two weeks of school) for word attack subtest only and June (within last three weeks of school) for both subtests
Does Washington Striving Readers… improve struggling middle school readers' performance on the state reading assessment?	*Measurements of Student Progress (MSP)* state reading assessment used for accountability	Group administered to all students	May

Gates-MacGinitie Reading Test. The *Gates-MacGinitie* is a group-administered, nationally normed assessment of reading comprehension. It provides students with 11 reading passages drawn from a range of fiction and nonfiction texts across multiple content areas, and asks students to answer questions that require understanding both explicit and implicit information in the passages.

There are two forms of the test (Form S and Form T), making it appropriate for use in pre- and posttesting. These assessments were renormed in 2005–2006 on a population of 59,066 K–12 students in 43 states. Developers report generally high correlation with other reading assessments but do not report exact values in the most recent technical manual (MacGinitie, MacGinitie, Maria, & Dreyer, 2002).

Because Striving Readers students were eligible for the study precisely due to their difficulties in reading, we decided to pretest them using the *Gates-MacGinitie* assessment designed for one level lower than their grade level (for example, grade 6 students were tested with the grade 5 assessment). To score the assessments, we used the out-of-grade-level norms provided by the publisher. In the spring, students were assessed at their own grade level, in order to enhance the face validity of the final outcome assessment.

A team of testers, trained and coordinated by the evaluation team, administered the *Gates-MacGinitie*. Pretests were administered between September 8 and September 23, 2010. We held a refresher training for the testers before they administered the posttests, between May 18 and June 9, 2011. We sent the assessments to Riverside, the test publishers, for scoring.

Measure of Student Progress (MSP). The *MSP* was new in 2010 and replaced the previous state assessment, the WASL. According to the OSPI, scores on the new *MSP* are comparable to scores on the WASL, even though the new assessment is shorter (State of Washington, 2012a). The *MSP* assesses reading comprehension, analysis, and critical thinking using functional documents (such as letters or

e-mails), informational passages (such as newspaper articles or excerpts from science or social studies texts), and literary passages (such as poems or excerpts from novels). Students read these passages and respond to multiple-choice, completion, and short-answer items (State of Washington, 2012b).

The *MSP*, like the WASL earlier, is administered to all students in grades 3–8 and in grade 10 each spring (the WASL previously was administered in April of each year but *MSP* testing occurs in May each year). For middle school students, the new reading *MSP* is a 90-minute assessment administered to a whole group in a single sitting. The *MSP* was the only assessment for this study that was not administered by the evaluation team. Instead, we obtained the data directly from the districts several months after testing.

Woodcock Reading Mastery Test-Revised. The *Woodcock Reading Mastery* consists of a "comprehensive battery of tests measuring several important aspects of reading ability" (Woodcock, 1998). For this study, we used two subtests, both administered to students individually:

> ***Woodcock Reading Mastery word attack subtest.*** The word attack subtest measures students' ability to decode either nonsense words or very uncommon words. Because students are not familiar with the list of words they are asked to read, the test measures their ability to apply phonic and structural analysis skills in order to pronounce new words. We administered the word attack subtest in both fall and spring of the 2010–2011 school year.
>
> ***Woodcock Reading Mastery word identification subtest.*** The word identification subtest asks students to read aloud isolated words. This subtest measures a student's ability to recognize words on sight. We administered the word identification subtest only in the spring as a posttest measure.

The same team of testers that administered the *Gates-MacGinitie* also administered the *Woodcock Reading Mastery* at the same time. Unlike the *Gates-MacGinitie,* which was machine scored by the publisher, the *Woodcock Reading Mastery* was hand-scored by our team of test administrators.

Because of the school's concern with the amount of time taken for assessment, we calculated the total number of minutes required annually to test students in the treatment and control groups. Testing time was a major reason we decided not to administer the vocabulary component of the *Gates-MacGinitie* at all and administered the *Woodcock Reading Mastery* word identification subtest once rather than twice per year. Table 4.4 summarizes the total amount of time required for testing each student; for nearly all students, the total number of minutes was near the bottom end of the range.

Table 4.4
Annual Testing Burden per Student

Measure	Time for each administration (minutes)	Number of administrations per year	Total minutes per year
Gates-MacGinitie	45–50	2	90–100
Woodcock Reading Mastery word attack	10–30*	2	20–60
Woodcock Reading Mastery word identification	10–30*	1	10–30
TOTAL			192–262

* The *Woodcock Reading Mastery* requires between 10 and 30 minutes for each subtest, depending on students' decoding ability. Students who can perform more tasks continue the test longer.

In addition to outcome assessment data, we collected student demographic information from each district about each student in the study.

Summary of Analytic Approach to the Impact Analysis

Our purpose was to estimate the impact of the Washington Striving Readers intervention on students' reading, as measured by four outcomes: *Gates-MacGinitie* comprehension, *Woodcock Reading Mastery* word attack, *Woodcock Reading Mastery* word identification, and the *MSP*. Our null hypothesis was that participation in Washington Striving Readers made no difference in student performance on these tests. We tested each hypothesis (one for each measure) using a multilevel model with sites modeled as fixed effect clusters to adjust for the nesting of students within schools (Raudenbush & Bryk, 2002). We did not conduct separate analyses by grade level, since students in grades 6–8 were in the same intervention class using the same materials. However, we did analyze the data separately for Group 1 and Group 2, as well as for the entire sample combined.

Covariates

For each analysis, students' prior achievement was used as a covariate. For the *Gates-MacGinitie,* this was the pretest score on the *Gates-MacGinitie*. For the *Woodcock*, only the word attack subtest was given in the fall; therefore, that subtest served as the covariate for both of the subtests administered in the spring. For the *MSP*, we used students' *MSP* score from the previous year (2010) as the covariate.

To account for missing pretests, we used a dummy variable adjustment in which two variables were used to represent prior student achievement. The first was the grand-mean centered pretest with missing values coded as "0." The second was the missing pretest dummy, in which missing values are coded as "1" and nonmissing values are coded as "0." In studies with a random assignment, this approach to handling missing data enables cases with missing pretest data to be retained in the analysis without biasing the impact estimate or its standard (Puma, Olsen, Bell, & Price, 2009).

We also included dummy-coded independent variables to account for the following student demographic variables: gender (1 = male, 0 = female), special education status (1 = identified for special education in a subject other than reading, 0 = not identified for special education in a subject other than reading), ELL status (1 = identified as ELL level 2 or above, 0 = not identified as ELL level 2 or above; note that level 1 students were newcomers and were excluded from that study), low income (1 = receiving free or reduced-price lunch [FRL], 0 = not receiving FRL).

We also included student ethnicity in all our analyses as dummy-coded variables. The ethnic category "other" served as the referent. This included students who identified themselves to their districts as "American Indian," "Asian," "Hawaiian Pacific Islander," or "multiracial." Ethnicity variables included African American, Latino, and white. For these variables, "1" meant that the student had identified him/herself as belonging to this group and "0" meant that the student had not.

The Model

In our model, we estimated the impact of the Washington Striving Readers program on student reading with School-Group fixed effects. School 1 Group 2 served as the referent, and the nine blocks were School 1 Group 1, School 2 Group 1, School 2 Group 2, School 3 Group 1, School 3 Group 2, School 4 Group 1, School 4 Group 2, School 5 Group 1 and School 5 Group 2. The general model is shown below.

$$Posttest_i = \beta_0 + \beta_1 TREATMENT_i + \beta_2(Pretest_i - \overline{Pretest}) + \beta_3(MissingPretest_i - \overline{MissPretest}) + \sum_{k=1}^{k} \beta_{k+3}(Student\ Demographic\ variable_i - \overline{Student\ Demographic\ variable}) + \sum_{j=1}^{j} \beta_{j+(k+3)} SchoolGroup\ block + \varepsilon_i$$

Where posttest$_i$ is the student reading outcome for student *i*;

- β_0 is the average student reading outcome among control group students in the referent school-group
- β_1 is the impact of the Washington Striving Readers program on the student reading outcome
- β_2 is the parameter for the pretest of student *i* and cannot be interpreted because of the dummy variable adjustment
- β_3 is the parameter for the missing pretest indicator for student *i* and cannot be interpreted because of the dummy variable adjustment
- $\sum_{k=1}^{k} \beta_{k+3}$ is a vector of *k* parameters for the effects of *k* student demographic variables on the student reading outcome
- $\sum_{j=1}^{j} \beta_{j+(k+3)}$ is a vector of *j* parameters for the difference in the average reading outcome among control group students in school-group *j* compared to control group students in school-group 0 (i.e., the average reading outcome among control students in school-group *j* is $\beta_0 + \beta_{j+(k+3)}$)
- ε_i is the deviation from the average reading outcome for student *i*

All student-level demographic variables were grand-mean centered in impact analysis models. For each outcome, we initially included all student demographic variables in the model. However, in order to create the most parsimonious model possible, in the final model we excluded variables with *p* values greater than 0.20. This resulted in slightly different models for each outcome. The student demographic covariates used in each final model can be seen in the tables in Appendix C.

To examine the impact of the intervention within each group, we analyzed the data for Group 1 and Group 2 separately. After selecting only students in the group in question, we used the same general model as in the overall analysis but using only School, rather than School-Group as the blocking variable.

Chapter 5
Results of the Impact Evaluation

In the previous chapter, we described eligibility criteria, the plan for randomization, and our analytical model. In this chapter, we first describe the analytic sample (the group of students for whom we had sufficient data to conduct our analyses) in detail. This means describing what the initial group of eligible students looked like and then following them through randomization and pretesting to posttesting, documenting the reasons some of them were lost from the study. We also provide detailed demographic information about the final analytic sample. We then present the results of the analyses using the four outcome measures for the overall sample. We follow this with separate analyses of results for Groups 1 and 2.

Random Assignment of Eligible Students

As described in the previous chapter, we worked with schools to screen 771 potentially eligible students. The screening process left us with 550 eligible students across the five participating schools. Those who needed intervention in decoding were eligible for Group 1 (with *Phonics Blitz* plus *Read to Achieve* as the intervention) while those who did not were eligible for Group 2 (with *Read to Achieve* as the sole intervention).

To conduct the random assignment of these eligible students to treatment and control conditions, we followed the plan described in Chapter 4. Overall, of the 550 eligible students, 250 students were assigned to the treatment condition, 246 were assigned to the control condition, and 54 were assigned to the waitlist. Two of the schools had enough eligible students to create a waitlist. The three other schools did not have waitlists and in some cases the treatment was undersubscribed. Many more students were in Group 2 (192 students in the treatment condition, 191 in the control condition, and 52 on the waitlist) than in Group 1 (58 students in the treatment condition, 55 in the control condition, and 2 on the waitlist.)

Attrition After Randomization and Before Pretest

Once students were assigned to treatment or control conditions, we began to track attrition in earnest. Attrition, or the loss of students during the study, is a very important consideration in an experimental study. In particular, it is important to know whether attrition occurred because of something about the study itself, or whether it occurred for other, exogenous reasons. Because the reasons that students are lost to a study might affect the interpretation of findings, in this section, we report in detail on attrition in Table 5.1 and in the text that follows.

Some students who were randomly assigned to treatment or control conditions left the study before they even knew that they were participants. We refer to these students as "lost exogenously." Although randomization occurred in June, students and parents were not notified of the assignment to Striving Readers classes or to the control condition (study hall or an elective) until the beginning of the school year (August 31 to September 5, 2010, depending on the school). Some of these students moved over the summer, enrolled in a different middle school, and as a result, were never informed of their assignment. This included 43 students from the treatment condition and 33 from the control condition. Of the 43 students assigned to the treatment condition who never showed up, 10 were in Group 1 and 33 were in

Group 2. Of the 33 assigned to the control condition who never showed up in the fall, 7 students were in Group 1 and 26 were in Group 2.

In addition, we discovered after random assignment that 32 students who we thought were eligible were in fact ineligible to participate in the study because a) the students had IEPs in reading, or b) were level 1 ELL students, or c) were in the one school that had a supplemental reading program and were on the list of students who had to receive state Learning Assistance Program (LAP) services. This applied to 19 students assigned to the treatment condition (11 students from Group 1 and 8 from Group 2) and to 13 students assigned to the control condition (7 students from Group 1 and 6 from Group 2).

The issue of the LAP services only came up after school started in fall 2010. As described in Chapter 1, the schools were originally selected in part because they reported that they did not already offer interventions to their struggling readers. Once the grant was awarded, we also talked to schools individually about pre-existing reading interventions or tutoring for struggling students and were told the schools did not offer these supports. Despite these efforts, we failed to find out about the LAP services provided at several schools. LAP is Washington's state-funded program (WAC 392-162 and Chapter 28A.165 RCW) that provides additional academic support in reading, math, and/or writing to students who score below grade level on the state's assessment. Achievement on district assessments of basic skills may also be considered. Eligible students are expected to participate in at least one of the subjects for which they are eligible. Schools do not have to serve all of their eligible students if they do not receive sufficient funding, but they do have to rank order their students by test scores and serve the lowest performing students first.

Two of the schools that provided LAP services to students had previously decided that if their LAP-eligible students were also eligible for the Striving Readers intervention, regardless of their assignment to treatment or control conditions, they would not receive supplemental reading services, although they could receive services in math (the schools did not offer services in writing). A third school, however, wanted to keep providing supplemental reading classes to LAP-eligible students who were also eligible for Striving Readers. OSPI determined in September 2010 that according to state law, schools had to serve the students in some way and had the option to choose how to support their LAP-eligible students. If the school chose to offer only supplemental reading classes, then the assignment to LAP classes had to take precedence over the Striving Readers study. Therefore, students eligible for LAP reading classes at this school became ineligible for Striving Readers. All of this information came out only after random assignment. As a result, 23 LAP students at one school were removed from the study after random assignment.

One student was removed from the study after random assignment at the request of his parents.

To compensate for the loss of students in the sample, both due to students moving away and to LAP participation, we added 27 students from the waitlist to the study. Of these, 15 students were assigned to the treatment condition (1 in Group 1 and 14 in Group 2) and 12 to the control condition (all 12 in Group 2). Ultimately, after the losses and the substitutions from the waitlist, we had 415 students (203 in the treatment condition and 212 in the control condition) in our target sample (i.e., the students whom we attempted to pre- and posttest).

Table 5.1
Numbers of Randomly Assigned Students Lost or Added and Reasons by Treatment Condition

Group	Reason lost or added	Treatment	Control
Total Sample	Lost exogenously (never informed)	43	33
	Special education status discovered late	6	0
	ELL level discovered late	0	1
	LAP students removed	11	12
	Unknown school requested removal	1	0
	Parent request to remove	1	0
	Added from waitlist	15	12
Group 1	Lost exogenously (never informed)	10	7
	Special education status discovered late	4	0
	ELL level discovered late	0	1
	LAP students removed	6	6
	Unknown school requested removal	1	0
	Parent request to remove	0	0
	Added from waitlist	1	0
Group 2	Lost exogenously (never informed)	33	26
	Special education status discovered late	2	0
	ELL level discovered late	0	0
	LAP students removed	5	6
	Unknown school requested removal	0	0
	Parent request to remove	1	0
	Added from waitlist	14	12

Sample of Students Taking the Pretests and Posttests and Attrition Before Posttesting

Shortly after the 2010–2011 school year began, we administered pretests for the *Gates-MacGinitie* reading comprehension test and the *Woodcock Reading Mastery* word attack subtest. We also obtained students' 2010 state reading assessment scores. Across our three pretest assessments, we were able to obtain scores for between 92 and 98 percent of students in all the subgroups. In Table 5.2, we report the percentage and number of students in the total sample and in Groups 1 and 2 who were pretested on the *Gates-MacGinitie* and *Woodcock Reading Mastery* and for whom we were able to obtain 2010 state reading assessment scores.

Table 5.2
Percentages and Numbers of Students Completing the Pretests by Treatment Condition

Group	Test	Treatment percentage (number)	Control percentage (number)	Treatment & control percentage (number)
Total Sample	*Gates-MacGinitie*	97% (196)	94% (200)	95% (396)
	Woodcock Reading Mastery	96% (195)	97% (205)	96% (400)
	MSP (state test)	97% (196)	99% (209)	98% (405)
Group 1	*Gates-MacGinitie*	95% (36)	95% (39)	95% (75)
	Woodcock Reading Mastery	92% (35)	100% (41)	96% (76)
	MSP (state test)	92% (35)	98% (40)	95% (75)
Group 2	*Gates-MacGinitie*	97% (160)	94% (161)	96% (321)
	Woodcock Reading Mastery	97% (160)	96% (164)	96% (324)
	MSP (state test)	98% (161)	99% (169)	98% (330)

Missing pretests occurred for several reasons. Some students were absent during the testing window and on the makeup days. A few students withdrew from the school (and sometimes returned later). In a few cases, we received unusable or damaged testing materials. We also did not have access to the spring 2010 *MSP* for students who moved to Washington from out of state and entered our study districts after the *MSP*. We later used a statistical adjustment that allowed us to include students who completed end-of-year tests but had missing pretests.

In spring 2011, we attempted to administer *Gates-MacGinitie* and *Woodcock Reading Mastery* posttests to all 415 students who were randomly assigned, regardless of whether they took the pretest. Completion rates for the spring testing ranged from 76 percent to 97 percent, as shown in Table 5.3.

Table 5.3
Percentages and Numbers of Students Completing the Posttests by Treatment Condition

Group	Test	Treatment percentage (number)	Control percentage (number)	Treatment & control percentage (number)
Total Sample	*Gates-MacGinitie*	87% (176)	86% (182)	86% (358)
	Woodcock Reading Mastery	86% (175)	86% (182)	86% (357)
	MSP (state test)	97% (196)	97% (205)	97% (401)
Group 1	*Gates-MacGinitie*	84% (32)	76% (31)	80% (63)
	Woodcock Reading Mastery	84% (32)	76% (31)	80% (63)
	MSP (state test)	97% (37)	95% (39)	96% (76)
Group 2	*Gates-MacGinitie*	87% (144)	88% (151)	88% (295)
	Woodcock Reading Mastery	87% (143)	88% (151)	88% (294)
	MSP (state test)	96% (159)	97% (166)	97% (325)

Among the students lost to the study for *Gates-MacGinitie* and *Woodcock Reading Mastery*, 53 had moved out of the schools (26 from the treatment condition and 27 from the control condition). Three had unusable, damaged, or missing tests. One was identified for special education midyear and was, therefore, ineligible for the study. Finally, one student refused to take the *Woodcock Reading Mastery* but did complete the *Gates-MacGinitie.*

We collected more posttests for the *MSP* than for the other two assessments. This was because some students moved away from the Striving Readers schools during the school year so they did not take the

Gates-MacGinitie and *Woodcock Reading Mastery* posttests, but they still took the *MSP* in their new schools, and we were able to receive those scores from the state education agency. Still, we had some attrition from the *MSP* sample as well. Among the 14 students lost to the study for the *MSP*, 13 had moved out of the state or for some reason were not tested. Six of these students were in the treatment condition and seven were in the control condition. One student (the same student mentioned in the previous paragraph) was identified for special education midyear and received an IEP in reading; that student became, therefore, ineligible for the study. After these losses, we were left with our final analytic samples for the study: 358 students in the *Gates-MacGinitie* analytic sample, 357 in the *Woodcock Reading Mastery* analytic sample, and 401 in the *MSP* analytic sample.

Demographics of the Analytic Samples and Equivalence of Treatment and Control

All five schools in this study received Title I funding and served student populations with substantial percentages of students who were eligible for FRL (between 45 and 64 percent). Four of the five schools served ELLs, who made up between 5 and 13 percent of the student population in those schools. One school did not serve any ELLs because the district chose to concentrate ELLs in another middle school, which was also in our study. Ethnically, the largest proportion of students in each school was white, ranging from 41 to 51 percent, depending on the school. The remaining students were Latino (ranging from 13 to 27 percent), African American (ranging from 8 to 21 percent), or other ethnicities.[2]

The demographic characteristics of our analytic samples were similar to the overall characteristics of the schools. Table 5.4 gives these demographic characteristics in detail for students in the *Gates-MacGinitie* and *Woodcock Reading Mastery* analytic samples. (Recall that these two samples differed by a single student; where the demographic make-up of the two samples differs, both are listed in the table.) There were no statistically significant differences between treatment and control groups for any of these demographic characteristics. Demographic characteristics of the *MSP* analytic sample differed slightly from those of the other two samples, but never by more than 3 percentage points. For a full table of the demographic characteristics of the *MSP* analytic sample, see Appendix B, which also provides further details about the significance testing.

[2] Demographic information about the schools was drawn from the 2009 Washington school report cards available at http://reportcard.ospi.k12.wa.us/summary.aspx?year=2010-11

Table 5.4
Demographic Characteristics of the Gates-MacGinitie *and* Woodcock Reading Mastery *Analytic Samples* by Group*

Group	Demographic characteristics	Treatment	Control
Total Sample N = 358 *Gates-MacGinitie* N = 357 *Woodcock Reading Mastery*	Male	55%	59%
	American Indian	4%	3%
	Asian	9%	7%
	African American	15%	18%
	Pacific Islander	6%	6%
	Latino	13%	18%
	White	44%	42%
	Multiracial	9%	7%
	FRL	59%	61%
	Special Education	6%	6%
	ELL	13%	11%
Group 1 N = 63 *Gates-MacGinitie* N = 63 *Woodcock Reading Mastery*	Male	47%	52%
	American Indian	13%	3%
	Asian	3%	13%
	African American	13%	23%
	Pacific Islander	13%	3%
	Latino	13%	19%
	White	47%	39%
	Multiracial	0%	0%
	FRL	59%	58%
	Special Education	13%	7%
	ELL	28%	19%
Group 2 N = 295 *Gates-MacGinitie* N = 294 *Woodcock Reading Mastery*	Male	57%	61%
	American Indian	2%	3%
	Asian	10%	5%
	African American	16%	17%
	Pacific Islander	4%	6%
	Latino	13%	18%
	White (*Gates*)	44%	42%
	White (*Woodcock*)	43%	42%
	Multiracial (*Gates*)	10%	8%
	Multiracial (*Woodcock*)	11%	8%
	FRL	59%	62%
	Special Education	5%	5%
	ELL	10%	9%

* Percentages are identical for *Gates* and *Woodcock* unless separate percentages are given.

Although the random assignment of students to treatment and control conditions should ensure that the two groups are very similar in terms of their reading ability at the beginning of the experiment, it is possible for random assignment to yield two non-equivalent groups. In Table 5.5, we report on the baseline equivalence of the students in the treatment and control conditions on the pretests. We report the mean (average) pretest scores of students in the control condition and students in the treatment condition in the two middle columns of the table. The righthand column indicates that none of the differences in average scores are statistically significant (i.e., *p*-values are all greater than .05).

In Appendix B we provide additional details about baseline equivalence of the groups at pretest and also in terms of their gender, ethnicity, receipt of free/reduced-price lunch, receipt of special education services, and ELL status. These analyses also show no significant differences between groups.

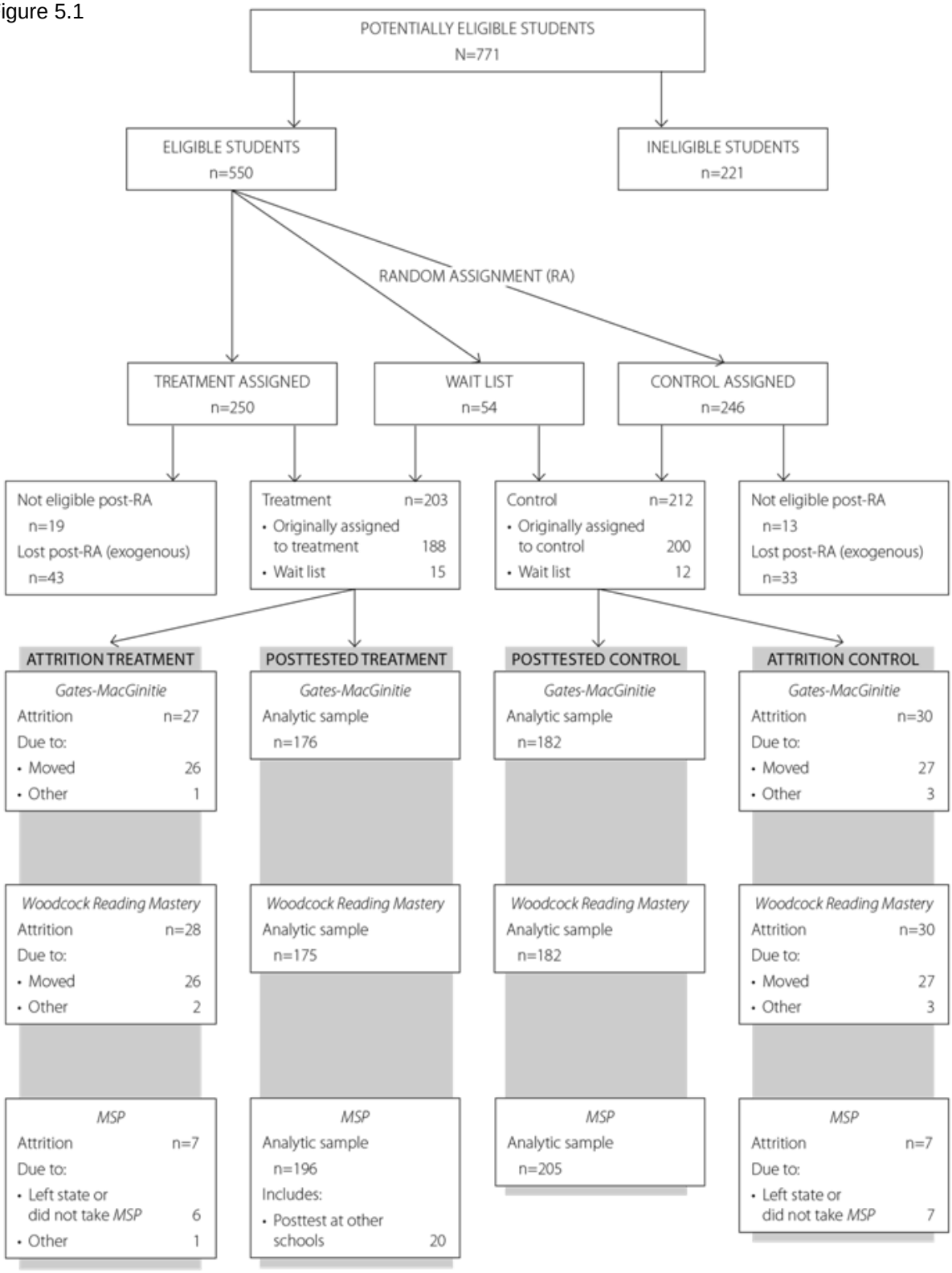
Table 5.5
Pretest Equivalence of the Analytic Sample

Total sample	N	Control mean	Treatment model-adjusted mean	*P*-value
Gates-MacGinitie	358	496.50	495.69	.792
Woodcock Reading Mastery word attack	357	98.76	98.40	.681
MSP	401	377.79	374.4	.516
Group 1				
Gates-MacGinitie	63	478.03	480.30	.765
Woodcock Reading Mastery word attack	63	92.16	91.93	.893
MSP	76	368.95	354.21	.397
Group 2				
Gates-MacGinitie	295	500.42	498.85	.647
Woodcock Reading Mastery word attack	294	100.16	99.79	.712
MSP	325	379.87	379.26	.899

CONSORT Flow Diagram of the Total Analytic Sample and Attrition Rates

Figure 5.1 on the following page is a CONSORT flow diagram (Schulz, Altman, & Moher, 2010). The term CONSORT refers to the statement on Consolidated Standards of Reporting Trials, which lays out expectations for reporting on samples and attrition in randomized controlled trials. This diagram illustrates the process that led us from the original identification of 771 students who were potentially eligible for Striving Readers to our ultimate analytic sample for each assessment.

Figure 5.1



CONSORT Flow Diagram Overall Sample

The CONSORT flow diagram essentially summarizes the information provided in chapter 4 about screening and eligibility criteria as well as the information in this chapter about the sample and attrition. Starting at the top of the diagram, we began with a group of 771 students who were potentially eligible for the Striving Readers intervention. We used the screening process in spring 2010 to confirm that students were eligible for Striving Readers and to learn whether they belonged in Group 1 or Group 2. In that first screening, we found 221 students were ineligible for Striving Readers. The remaining 550 students were randomly assigned to the treatment condition, the control condition, or in schools where there were more eligible students than spaces in the two groups, to a waitlist. This can be seen in the second and third row of blocks in the diagram.

The outer two boxes in the fourth row of the diagram report on students we lost after randomization for reasons unrelated to the study. Most of these were students who moved away from the school during the summer, before they even learned they had been assigned to the treatment or control group. Some were the students that we found out, belatedly, should never have been considered eligible.

The middle two boxes in the same fourth row in the flow diagram report the number of students we actually had in each group at the beginning of the school year, including students who were added from the waitlist. Between the beginning and the end of the school year, we lost additional students, leaving us with our final analytic sample (represented in the fifth, sixth, and seventh rows of boxes).

Table 5.6 reports the rate of attrition we experienced between pretest and posttest in our study. In calculating attrition in the total sample for the *Gates-McGinitie*, we included 447 students (222 treatment and 225 control) in the number we had hoped to keep in the study. This 447 included students who initially participated in the study (203 treatment and 212 control), as well as those who attended Striving Readers schools but were found to be ineligible after random assignment (19 treatment and 13 control). It did not include students lost to the study exogenously (i.e., students who never enrolled in the school and, therefore, never knew of their assignment to the study: 27 from the treatment and 30 from the control condition).

Of these 447 whom we hoped to keep in the study for the *Gates-MacGinitie*, we lost 89 students (20%). Of these 89 lost students, 32 were actually ineligible (19 treatment and 13 control) and 57 students (27 treatment and 30 control) left the study for other reasons most frequently because they moved away, as shown in Figure 5.1. Attrition calculations for the other tests (*Woodcock Reading Mastery* and *MSP*) were conducted in the same way. The resulting percentages are shown in Table 5.6.

Table 5.6
Attrition Rates From Pretest to Posttest for the Total Sample

	Overall attrition rate	Treatment condition attrition rate	Control condition attrition rate	Differential attrition rate
Gates-MacGinitie	19.9%	20.7%	19.1%	1.6%
Woodcock Reading Mastery	20.1%	21.2%	19.1%	2.1%
MSP	10.3%	11.7%	8.9%	2.8%

One important consideration in experimental studies is whether attrition was different for the treatment than for the control group. If it was, this could be a source of bias that could affect the findings. Differential attrition is also reported in Table 5.6 and did not exceed three percent for any of the

assessment measures. These differential attrition rates, as well as the overall attrition rates, are well within the range of what is considered acceptable attrition for an experimental study, according to What Works Clearinghouse standards (U.S. Department of Education, 2010).

Crossovers and Posttested at Other Schools

"Crossovers" occur when students from the treatment condition do not attend the Striving Readers class and thus have an experience more like students in the control condition, or when students from the control condition enroll in a Striving Readers class. We did not have any crossovers among students in the *Gates-MacGinitie* and *Woodcock Reading Mastery* analytic samples. All students for whom we had *Gates-MacGinitie* and *Woodcock Reading Mastery* posttests had experienced the condition—treatment or control—to which they were assigned.

The analytic sample used to examine impacts on the *MSP*, however, included 20 students who may have been crossovers. These 20 students were assigned to the treatment condition but transferred out of the school midyear. Because these students remained in the state, they still took the *MSP*, and we were able to analyze their scores as part of their originally assigned condition. We do not have data indicating when their transfers took place. Students who transferred near the end of the year, shortly before *MSP* testing, probably received most of the treatment. Students who transferred at the beginning of year, however, did not receive the Striving Readers intervention and therefore had a "crossover" experience. Since we cannot be sure which occurred, we labeled these students "Posttest at Other School" in the CONSORT flow diagram.

Impact for Students (Groups 1 and 2 Combined)

To analyze outcome data, we used fixed effects, intent-to-treat (ITT) models. Using a "fixed effects" model means that we estimated the average impact across all the schools in the study rather than estimating a separate impact for each school. Using an ITT model simply means that the scores of any student who was assigned to the treatment condition are analyzed with the treatment condition, even if the student did not try hard in class or was often absent or if the teacher did not implement the program. The benefit of an ITT model is that it gives schools a good sense of how the program works in a real-life setting, not only when used at the highest level with the most eager students. We described our model and other methodological issues in detail earlier, in Chapter 4.

For the *Gates-MacGinitie* and the *Woodcock Reading Mastery*, our final regression models did not find any statistically significant impacts for the treatment. Students in the treatment condition outperformed those in the control condition on the *Gates-MacGinitie*, scoring an average of 0.72 points higher (p = .713) and on the *Woodcock Reading Mastery* word attack, scoring an average of 0.68 points higher (p = .333). The reverse occurred on the *Woodcock Reading Mastery* word identification, with students in the control scoring an average of 0.30 points higher than students in the treatment (p = .645). In all three cases, these observed differences were not statistically significant.

For the *MSP*, we found a statistically significant positive impact of the treatment on reading achievement, with students in the treatment outperforming those in the control group by an average of 3.08 points (p = .048). These results are summarized in Table 5.7.

Table 5.7
Overall Impact of the Intervention on Student Reading Achievement, Total Sample

	Control condition			Treatment condition				
	N	**Mean**	**(SD)**	**Model-adjusted mean**	**(SD)**	**Estimated impact**	**Effect size**	***P*-value**
Gates-MacGinitie	358	502.28	(25.9)	503.00	(24.2)	0.72	0.03	.713
Woodcock Reading Mastery word identification	357	95.66	(8.2)	95.36	(8.5)	-0.30	-0.04	.645
Woodcock Reading Mastery word attack	357	99.13	(8.9)	99.81	(9.4)	0.68	0.08	.333
MSP	401	383.07	(19.3)	386.15	(19.2)	3.08	0.16	.048

Table 5.7 also reports the standardized effect size (Glass's *delta*) for each of the four analyses. Effect size is a measure of the magnitude of the effect, which matters because an effect can be statistically significant but still not represent a large difference in terms of how much students learned. It is calculated as the difference between the means of the treatment and the means of the control group, divided by the standard deviation of the control group. The effect size for the *MSP* was 0.16.

These final regression models for the analyses included covariates representing the pretest, missing pretests, student demographic variables, and variables representing the school-group. Appendix C provides more detail about these results.

Sample Size and Attrition for Group 1

Group 1, made up of students who needed support in phonics and decoding, was the smaller of the two groups. The total analytic sample for Group 1 consisted of just 63 students for the *Gates-MacGinitie* and *Woodcock Reading Mastery* and 76 for the *MSP*.

As seen in Table 5.8, attrition from the ITT analytic sample was 35 percent for the *Gates-MacGinitie* and *Woodcock Reading Mastery* and 22 percent for the *MSP*. This is higher than attrition rates for the total sample. We calculated these percentages in the same way we calculated attrition from the total sample: the number of students we counted as assigned to the study included the students found ineligible, as well as those who initially participated in the study. Also, the number of students lost to the study excluded students who were lost from the sample exogenously (i.e., those who left the school before learning of their assignment).

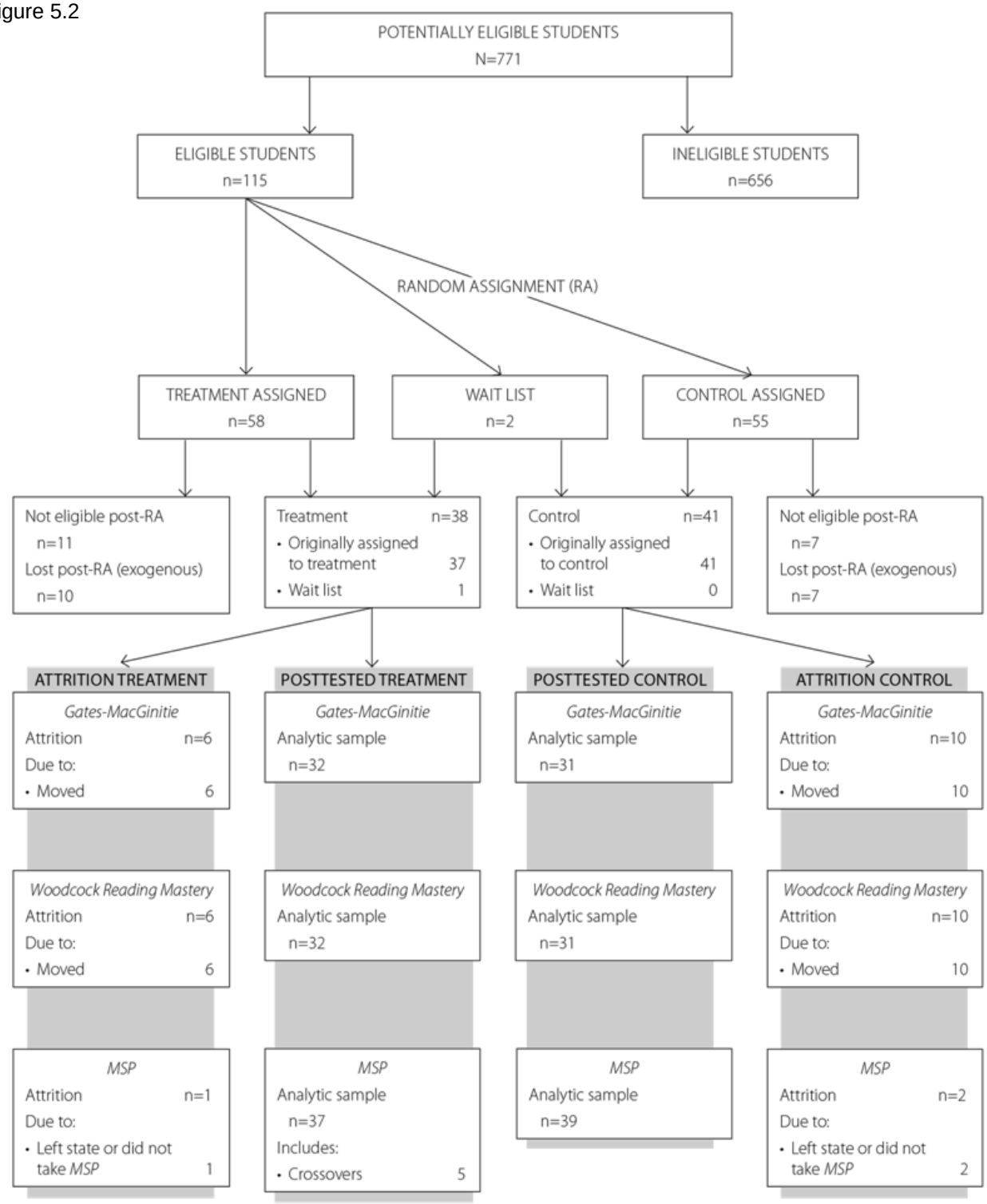
Table 5.8
Attrition Rates From Randomization to Posttest for Group 1

	Overall attrition rate	Treatment condition attrition rate	Control condition attrition rate	Differential attrition rate
Gates-MacGinitie	35.1%	34.7%	35.4%	0.7%
Woodcock Reading Mastery	35.1%	34.7%	35.4%	0.7%
MSP	21.6%	24.5%	18.8%	5.7%

Differential attrition (the difference between attrition in the treatment and control groups) was less than 1 percent for the *Gates-MacGinitie* and *Woodcock Reading Mastery* and about 6 percent for the *MSP*. As with the attrition in the overall sample, these rates are well within the range of what is acceptable attrition for an experimental study, according to What Works Clearinghouse standards (U.S. Department of Education, 2010).

Crossovers or Posttested at Other Schools. As noted in the earlier discussion about the overall sample, there were no crossovers for the *Gates-MacGinitie* and *Woodcock Reading Mastery*. In the analytic sample for the *MSP*, however, there were five students tested at other schools. These students had been in the treatment sample but transferred out of the district and were posttested at other schools in the state. This information and the entire path from potential eligibility to posttest are represented in Figure 5.2, the CONSORT flow diagram for Group 1.

Figure 5.2



CONSORT Flow Diagram, Group 1

Impact for Students in Group 1

For all four outcome measures, our final regression models did not find any statistically significant impacts for the treatment. Students in the treatment outperformed those in the control group on all four measures: scoring an average of 2.57 points higher on the *Gates-MacGinitie* (p = .581), an average of 0.77 points higher on *Woodcock Reading Mastery* word identification (p = .540), an average of 2.16 points higher on *Woodcock Reading Mastery* word attack (p = .540), and an average of 1.68 points higher on the *MSP* (p = .555). Table 5.9 summarizes these findings and reports the standardized effect sizes (Glass's *delta*).

Table 5.9
Overall Impact of the Intervention on Student Reading Achievement, Group 1

	Control condition			Treatment condition				
	N	**Mean**	**(SD)**	**Model-adjusted mean**	**(SD)**	**Estimated impact**	**Effect Size**	***P*-value**
Gates-MacGinitie	63	486.52	(19.3)	489.09	(22.7)	2.57	0.13	.581
Woodcock Reading Mastery word identification	63	87.74	(5.6)	88.51	(6.7)	0.77	0.14	.540
Woodcock Reading Mastery word attack	63	92.55	(6.6)	94.71	(9.3)	2.16	0.33	.540
MSP	76	376.33	(15.4)	378.01	(18.9)	1.68	0.11	.555

The final regression models for the analyses included covariates representing the pretest, missing pretests, student demographic variables, and variables representing the school. We provide further details about these analyses in Appendix C.

Sample Size and Attrition for Group 2

Most of the students in Washington Striving Readers (82%) were in Group 2, the group that spent the entire year working in *Read to Achieve.* The total analytic sample for Group 2 consisted of 295 students for the *Gates-MacGinitie*, 294 for the *Woodcock Reading Mastery*, and 325 for the *MSP*.

Attrition from the Group 2 analytic sample was 16 percent for the *Gates-MacGintie* and *Woodcock Reading Mastery* and 7 percent for the *MSP*, as shown in Table 5.10 (again including students found ineligible among those originally assigned to the study but excluding students lost for exogenous reasons).

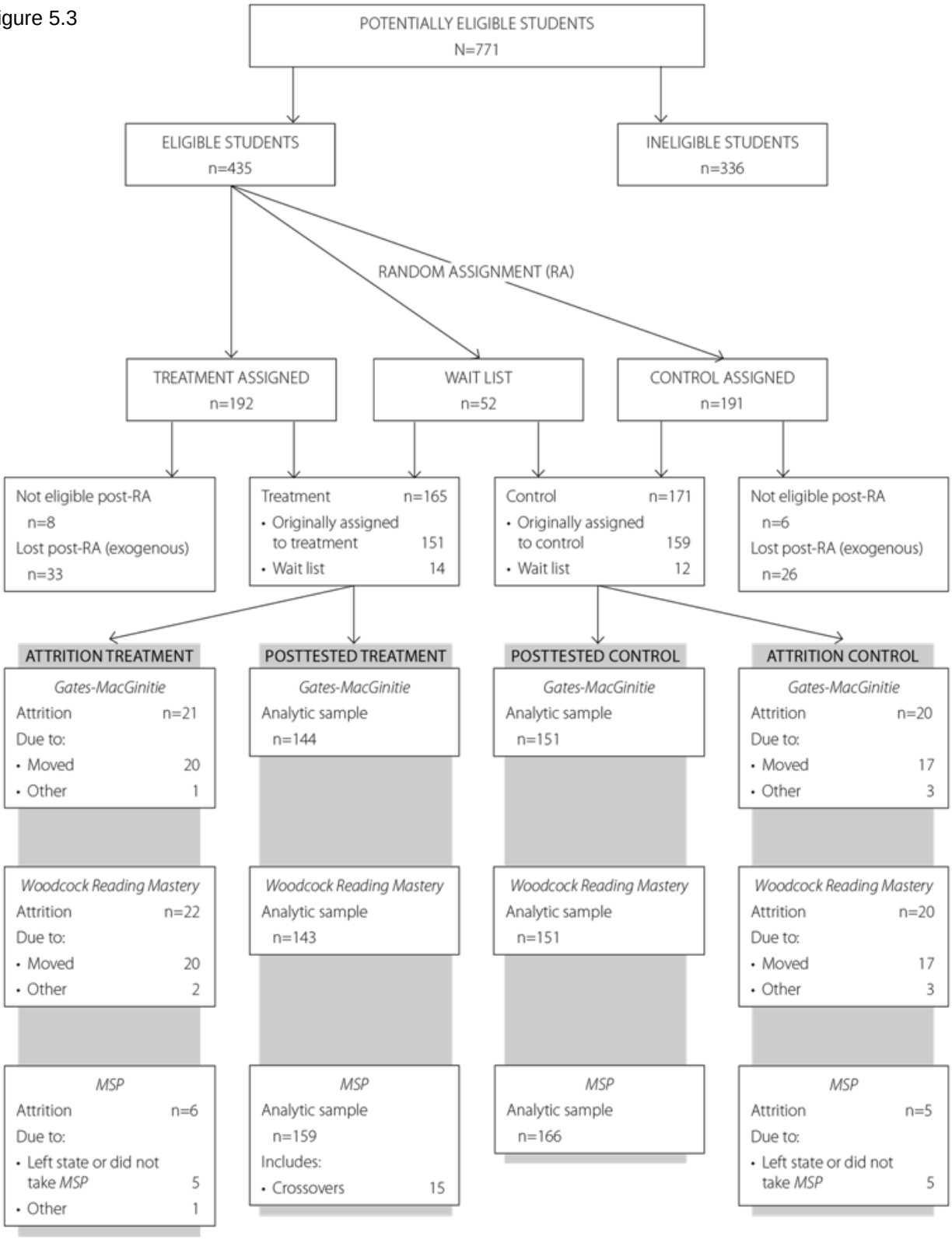
Table 5.10
Attrition Rates From Randomization to Posttest for Group 2

	Overall attrition rate	Treatment condition attrition rate	Control condition attrition rate	Differential attrition rate
Gates-MacGinitie	15.7%	16.8%	14.7%	2.1%
Woodcock Reading Mastery	16.0%	17.3%	14.7%	2.7%
MSP	7.1%	8.1%	6.2%	1.9%

Differential attrition (the difference between attrition in the treatment and control conditions) was less than 3 percent for all three ITT samples, also well within the acceptable range, according to What Works Clearinghouse standards (U.S. Department of Education, 2010).

Crossovers or Posttested at Other Schools. As noted earlier, there were also no crossovers for the *Gates-MacGinitie* and *Woodcock Reading Mastery*. The Group 2 analytic sample for the *MSP*, however, did include 15 crossovers from the treatment sample. These students had been in the treatment sample but transferred out of the district and were administered the *MSP* at other schools in Washington. This information and the entire path from potential eligibility to posttest are represented in Figure 5.3, the CONSORT flow diagram for Group 2.

Figure 5.3



CONSORT Flow Diagram, Group 2

Impact for Students in Group 2

Findings for Group 2 were similar to those for the total sample in terms of which group (treatment or control) performed better and in terms of effect sizes. Unlike our findings for the total sample, however, there were no statistically significant impacts for the treatment for Group 2 alone.

Students in the treatment condition outperformed those in the control condition on three of the four measures. As shown in Table 5.11, students in the treatment condition scored an average of 0.55 points higher on *Gates-MacGinitie* (p = .798), an average of 0.57 points higher on *Woodcock Reading Mastery* word attack (p = .472), and an average of 3.27 points higher on the *MSP* (p = .067). In contrast, on *Woodcock Reading Mastery* word identification, students in the control condition outperformed those in the treatment by an average of 0.25 points (p = .737).

Table 5.11
Overall Impact of the Intervention on Student Reading Achievement, Group 2

	Control condition			Treatment condition				
	N	**Mean**	**(SD)**	**Model-adjusted mean**	**(SD)**	**Estimated impact**	**Effect size**	***P*-value**
Gates-MacGinitie	295	505.52	(26.0)	506.07	(23.5)	0.55	0.02	.798
Woodcock Reading Mastery word identification	294	97.29	(7.7)	97.04	(8.0)	-0.25	-0.03	.737
Woodcock Reading Mastery word attack	294	101.48	(8.3)	102.05	(8.9)	0.57	0.07	.472
MSP	325	384.66	(19.9)	387.93	(18.5)	3.27	0.16	.067

The Glass's *delta* standardized effect size for Group 2 on the *MSP* was the same as the effect size for the total sample (ES = 0.16). The effect size for the other assessments were smaller (see Table 5.7). Appendix C provides more detail about these results, including covariates contained in the final models.

Chapter 6
Conclusions

The Washington Striving Readers program provided intensive in-school reading intervention to 176 middle school students who read significantly below grade level. The program intentionally differentiated the content of the intervention in order to address the specific student needs, as identified by a screening assessment. Most students (Group 2) spent the year in a class that used the *Read to Achieve* program. A much smaller group of students (Group 1) spent the first part of the year working with the *Phonics Blitz* program before moving into *Read to Achieve.*

We examined four aspects of program implementation: teachers' receipt of the intended professional development, their receipt of in-class coaching, their delivery of the programs as intended, and the completion of all the lessons that were supposed to be covered. For the first three aspects, implementation was high. The teachers who provided the intervention received the intended professional development and in-class coaching, and they delivered the intervention the way it was intended.

Lesson completion was the one aspect of implementation that was not high. The pacing of instruction was much slower than expected, meaning that teachers did not deliver all of the intended content. This was particularly true for students in Group 1, because *Phonics Blitz* was delivered much more slowly than planned. There are several possible reasons why this might have occurred. *Phonics Blitz* is designed to move briskly from one instructional activity to the next, each with its own specific routines. It may be that it takes longer than expected for teachers to become comfortable enough with the routines to complete them in the allotted time. Also, some teachers originally misunderstood some components of the program. For example, some teachers complained that it took an entire class period to have students read through all of the "Sentences to Read," one of the program components, when in fact the program intention was for teachers to select only a few of the sentences and then move on to the next task. If the program had continued a second year, one question to explore would have been whether the pacing of *Phonics Blitz* might have been faster once teachers were more experienced with the program. It is also possible that the expectations for the pacing of lesson delivery were unrealistically high.

The early end to the study also had consequences for the evaluation of program impact. The randomized controlled trial was originally designed to combine results from three years in order to be able to detect effects of the intervention. Because the program only ran a single year, our sample size was smaller than planned, making it less likely that we would find significant effects.

Results of the *Gates-MacGinitie* and the *Woodcock Reading Mastery* word identification and word attack subtests revealed no significant differences between the treatment and control groups. There was, however, a significant positive impact on the *MSP,* the Washington state reading assessment. The size of this effect (Glass's *delta*) was 0.16—not a large impact but comparable in size to that found in other Striving Readers programs (Faddis et al., 2010; Hamilton et al., 2011).

Another way to think about the size of the impact on the *MSP* is to consider the effect size of the average annual gain of middle school students in reading. Between fifth and sixth grades, the average annual gain in effect size is 0.32, between sixth and seventh 0.23, and between seventh and eighth 0.27 (Hill, Bloom, Black, & Lipsey, 2007). If we consider 0.25 to be a conservative estimate for middle-schoolers overall, an effect size of 0.16 represents 59 percent of that gain or about five months of a nine-month school year.

This improvement reduced the gap between low-performing readers and their peers who read at grade level, but did not close that gap. Students in the treatment condition still had average *MSP* scores that put them below the cut point to be considered proficient readers.

Cutting the study short inevitably meant that some of what we might have learned was lost. We do not know, for example, whether lesson completion might have increased with more experience, and whether covering more lessons might have affected student learning. Nevertheless, there are meaningful lessons from this one-year study that can have important implications for those implementing similar interventions in the future.

The first is that it is important to attend not only to the fidelity of program implementation but to the amount of material taught during the year (lesson completion). If we had measured through observations only whether the program was delivered as intended, we might have missed the fact that teachers were not able to finish teaching all the material in the *Read to Achieve* program. It may be that when schools implement new intervention programs, teachers need additional support to ensure appropriate pacing—or simply time to thoroughly learn the program.

Another lesson from this study is that it is possible to make a statistically significant difference in students' overall literacy achievement in the course of one school year. Students in the treatment condition performed better on the state reading assessment than did students in the control condition. But, this positive finding is tempered by the fact that the gains made were not sufficient to bring students up to a proficient level. In light of these and other recent findings (e.g., Vaughn et al., 2011), it seems that middle school students who read substantially below grade level may need more than one extra reading class for one year. A summer program, an additional intervention class, and/or a second year in intervention might help students make additional progress.

Limitations

There are several limitations to our evaluation. As noted previously, the early end to the study and resulting smaller sample size made it less likely we would find significant effects. It also had implications for our study of program implementation. We had only two sets of observation data and implementation ratings from the first year, instead of three sets as originally planned. Our number of observations for *Phonics Blitz,* in particular, was lower than ideal. Also, we lost the opportunity to recheck interrater reliability on the *Read to Achieve* protocol following our recalibration of ratings in the winter. Finally, we only saw the program in its first year, when the teachers were still learning how it all worked. It is possible that implementation and lesson completion might have looked different in subsequent years, when teachers had more experience.

References

Agodini, R., Harris, B., Atkins-Burnett, S., Heaviside, S., Novak, T., & Murphy, R. (2009). *Achievement effects of four early elementary school math curricula: Findings from first graders in 39 schools* (NCEE 2009-4052). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Creswell, J.W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, CA: SAGE.

Dynarski, M., Agodini, R., Heaviside, S., Novick, T., Carey, N., Campuzano, L., & Sussex, W. (2007). *Effectiveness of reading and mathematics software products: Findings from the first student cohort. Report to Congress* (NCEE 2007-4005). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Faddis, B.J., Beam, M., Maxim, L., Vale, E., Hahn, K., & Hale, R. (2010). *Portland Public Schools' Striving Readers program: Year 3 evaluation report.* Portland, OR: RMC Research.

Feldman, J., Schenck, A., Coffey, D., & Feighan, K. (2010). *Memphis Striving Readers Project: Year 3 evaluation report.* Memphis, TN: Research for Better Schools.

Hamilton, J., Meisch, A., Chen, E., Quintanilla, P., Fong, P., Gray-Adams, K., & Thornton, N. (2011). *Striving Readers study: Targeted & whole-school interventions-year 4.* Rockville, MD: Westat.

Hill, C.J., Bloom, H.S., Black, A.R., & Lipsey, M.W. (2007). *Empirical benchmarks for interpreting effect sizes in research.* New York: MDRC.

MacGinitie, W.H., MacGinitie, R.K., Maria, K., and Dreyer, L.G. (2002). *Gates-MacGinitie Reading Tests: technical report forms S and T (4th ed.). Itasca, IL: Riverside.*

Mahoney, J.L., & Zigler, E.F. (2006). Translating science to policy under the No Child Left Behind Act of 2001: Lessons from the national evaluation of the 21st-Century Community Learning Centers. *Journal of Applied Developmental Psychology, 27*(4), 282–294.

Marchand-Martella, N. & Martella, R. (2010). *SRA Read to Achieve: Comprehending Content-Area Text and Comprehending Narrative Text,* McGraw-Hill SRA

Penuel, W.R., Frank, K.A., Fishman, B.J., Sabelli, N., & Cheng, B. (2009). *Expanding the scope of implementation research in education to inform design*. Retrieved from SRI International, Center for Technology in Learning website: http://ctl.sri.com/publications/downloads/ExpandingScopeImplementationResearchEducationInformDesign.pdf

Pianta, R.C., La Paro, K.M., & Hamre, B.K. (2008). *Classroom Assessment Scoring System (CLASS).* Baltimore, MD: Paul H. Brookes.

Puma, M.J., Olsen, R.B., Bell, S.H., & Price, C. (2009). *What to do when data are missing in group randomized controlled trials* (NCEE 2009-0049). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Raudenbush, S.W. & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods.* Thousand Oaks, CA: Sage Publications.

Really Great Reading. (2010). *Phonics Blitz* (2nd ed.). Cabin John, MD: Author. (APA style)Schulz, K.F., Altman, D.G., & Moher, D. (2010).

Schulz, K.F., Altman, D.G., & Moher, D. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomized trials. *Annals of Internal Medicine, 152*(11), 726–732.

Smith M.W., Brady, J.P., & Anastasopoulos, L. (2008). *Early Language & Literacy Classroom Observation: Pre-K tool.* Baltimore, MD: Paul H. Brookes.

State of Washington, Office of Superintendent of Public Instruction. (2012a). *State testing 2012* [Parent/guardian handout]. Retrieved from http://www.k12.wa.us/resources/pubdocs/StateTesting.pdf

State of Washington, Office of Superintendent of Public Instruction. (2012b). *Reading assessment: Updates for 2012. Measurements of student progress: Reading, grades 6-8.* Retrieved from http://www.k12.wa.us/Reading/Assessment/pubdocs/ReadingAssessmentUpdates2012_Grades6_8.pdf

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2010). *What Works Clearinghouse: Procedures and standards handbook (version 2.1)*. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v2_1_standards_handbook.pdf

Vaughn, S., Wexler, J., Roberts, G., Barth, A, Cirino, P., Romain, M., Francis, D., Fletcher, J., Denton, C. (2011). Effects of individualized and standardized interventions on middle school students with reading disabilities. *Exceptional Children, 77* (4:391-407).

Woodcock, R.W. (1998). *Woodcock Reading Mastery Tests-Revised Forms G and H Examiners Manual*. American Guidance Service, Inc.

Appendix A
Washington Striving Readers Implementation Measures

Phonics Blitz Observation Protocol

Read to Achieve Observation Protocol

Teacher Interview Protocols

Striving Readers
Phonics Blitz Observation Protocol

I. INFORMATION

Observer	Date	Period
School	Observation Start Time	
District	Observation End Time	
Teacher	❑ Regular ❑ Substitute	
Highest number of students at any time	Lesson number(s)	

Each item is scored not very true, somewhat true, very true which will be given a 1, 2, and 3– point value.

Not very true of this lesson is the equivalent of low (or absent) implementation. This may mean:

- I never or rarely observed this descriptor
- I observed an inappropriate modification of this descriptor
- There was some evidence of this descriptor, but with many caveats.

Somewhat true of this lesson is the equivalent of medium or mixed implementation. This may mean:

- This descriptor is what I observed some of the time or
- This descriptor is somewhat observed all of the time or
- This descriptor is somewhat true, but with some caveats

Very true of this lesson is the equivalent of high implementation. Note that this does not have to mean perfect implementation. Rather, it is good and appropriate implementation which means:

- This descriptor is what I observed all the time or
- This descriptor is what I observed most of the time or
- I observed an appropriate modification of this descriptor

"Not relevant" should be used when this isn't an issue that can be scored. For example, if there are no errors made by students, then you could not rate positive error correction. Or, for example, a few lessons don't have words to preview.

II. FIDELITY

Oral Reading

☐ Not part of the observation

	Not very true of this lesson	Somewhat true of this lesson	Very true of this lesson	Not Relevant
1. T reviews all words to preview				
2. Ss skim passage (lessons 26-50 only)				
TIMED READING				
3. Three students each take individual turns doing timed reading				
4. All students who aren't readers are actively checking the reader				
5. Teacher reviews a maximum of three errors per student.				
COMPREHENSION QUESTIONS (lessons 26-50 only)				
6. Ss read and answer questions silently				
7. Work is reviewed: S reads answer and other Ss check. Teacher explains if there is disagreement.				

Phonological Awareness/ Phonemic Awareness (lessons 1-25)

☐ Not part of the observation

	Not very true of this lesson	Somewhat true of this lesson	Very true of this lesson	Not Relevant
1. T states objectives				
2. T introduces/reviews concept as indicated in the lesson				
Modeling and practice -1st time				
3. T models (I do) as indicated				
4. T & Ss do activity together (we do) as indicated				
5. Ss complete activity without T (you do) as indicated				
Modeling and practice – 2nd time				
6. T models (I do) as indicated				
7. T & Ss do activity together (we do) as indicated				
8. Ss complete activity without T (you do) as indicated				
Modeling and practice – 3rd time				
9. T models (I do) as indicated				
10. T & Ss do activity together (we do) as indicated				
11. Ss complete activity without T (you do) as indicated				
Overall				
12. T never writes words, refers to letter names, or uses letter tiles (posters are okay)				
13. T always uses correct phoneme pronunciation, motions, and sound/phoneme names				
14. Ss always use fingers when stretching sounds				
15. Positive error correction				

Phonics (all lessons)

☐ Not part of the observation

	Not very true of this lesson	Somewhat true of this lesson	Very true of this lesson	Not Relevant
1. T states objectives				
2. T introduces/reviews concept as indicated in the lesson plan.				
Modeling and practice -1st time				
3. T models (I do) as indicated				
4. T & Ss do activity together (we do) as indicated				
5. Ss complete activity without T (you do) as indicated				
Modeling and practice – 2nd time				
6. T models (I do) as indicated				
7. T & Ss do activity together (we do) as indicated				
8. Ss complete activity without T (you do) as indicated				
Modeling and practice – 3rd time				
9. T models (I do) as indicated				
10. T & Ss do activity together (we do) as indicated				
11. Ss complete activity without T (you do) as indicated				
Overall				
12. T uses tiles/white boards when called for				
13. Ss use tiles/white boards when called for				
14. Positive error correction used				

Word Sort (most lessons)

☐ Not part of the observation

	Not very true of this lesson	Somewhat true of this lesson	Very true of this lesson	Not Relevant
1. T explains purpose of word sort				
2. T and Ss practice the full first line together.				
3. All Ss complete Word Sort independently				
4. Ss check work and make error corrections as needed				

Detective Work (most lessons)

☐ Not part of the observation

	Not very true of this lesson	Somewhat true of this lesson	Very true of this lesson	Not Relevant
1. T models and students imitate the teacher for all of column 1				
2. In pairs, students complete column 2 and 3, including writing the number correct and using proper error correction				

Words to Read (all lessons)

☐ Not part of the observation

	Not very true of this lesson	Somewhat true of this lesson	Very true of this lesson	Not Relevant
1. Every student individually reads at least one line with three words				
2. Ss participate as checkers (marking errors and identifying incorrect words)				
3. Teacher uses appropriate error correction				

Sentences to Read (all lessons)

☐ Not part of the observation

	Not very true of this lesson	Somewhat true of this lesson	Very true of this lesson	Not Relevant
1. Every student in the class reads at least one sentence individually				
2. Students track errors on paper				
3. Teacher uses appropriate error correction (thumbs up, helping hand)				

Overall

	Not very true of this lesson	Somewhat true of this lesson	Very true of this lesson	Not Relevant
1. Teacher does not extend concepts (e.g., talk all about blending) or bring in outside concepts				
2. Classroom is set up appropriately (in a "U" shape or desks parallel so ss can see each other).				

III. Classroom Characteristics. *Complete this section during the last 5 minutes. If an issue cannot be rated (e.g., there were no behavior problems), leave that item blank.*

	1	2	3	4
Classroom Climate	Classroom is rarely or never characterized by positive affect (smiling, enthusiasm), positive communications and respect among	Classroom is sometimes characterized by positive affect (smiling, enthusiasm), positive communications and respect among	Classroom is often characterized by positive affect (smiling, enthusiasm), positive communications and respect among teachers and students.	Classroom is always characterized by positive affect (smiling, enthusiasm), positive communications and respect among teachers and
Organization of Materials	Very disorganized materials (students or teacher can't find things).	Somewhat disorganized materials (students and/or teacher sometimes cannot find things easily)	Mostly organized materials (students and teacher usually find things easily).	Very well organized materials (students and teacher almost always find things easily).
Classroom Routines	Routines are unclear to students and/or no routines are established. Many students are confused about what to do.	Some unclear or unestablished routines. Some students are confused about what to do.	Mostly clear and established routines. A few students may be confused or there may be 1 or 2 instances of lack of clarity.	Very clear and established routines. All students know what to do with rare exceptions.
Student Engagement	Few students actively participate throughout the lesson. Few students are on task.	Some students actively participate throughout the lesson. Some students are on task for most/all of the lesson.	Most students are actively participating throughout the entire lesson. Most students are on task for most of the lesson.	All students are actively participating throughout the entire lesson. All students are on task for the entire lesson with rare exceptions.
Addressing Behavior Problems	The teacher is always or almost always ineffective at addressing students' behavior problems.	The teacher is often ineffective at addressing students' behavior problems.	The teacher is usually effective at addressing students' behavior problems.	The teacher is always effective at addressing students' behavior problems. Or, no student behavior problems.
Lesson Pacing	The pace of the lesson is never appropriate; it is always too fast or too slow for students.	The pace of the lesson is often not appropriate; it is usually too fast or too slow for students.	The pace of the lesson is usually appropriate, with a few instances of being too fast or slow.	The pace of the lesson is always appropriate for students (with rare exceptions).
Teacher Monitoring	Teacher rarely or never monitors independent/ partner/group work (e.g., teacher sits at desk or does other things).	Teacher sometimes monitors independent/ partner/group work but sometimes does not.	Teacher usually monitors independent/partner/group work although her attention may turn away for a few minutes.	Teacher always monitors student independent/ partner/group work.

Striving Readers
Read to Achieve Observation Protocol

I: INFORMATION

Observer Name	Date	Period
School Name	Observation Start Time	
District Name	Observation End Time	
Teacher Name	❑ Regular ❑ Substitute	
Highest Number of Students	❑ Content ❑ Narrative	

II: READ TO ACHIEVE FIDELITY

In the table below, transfer the total points you gave for each part observed. Leave any areas not observed blank.

Unit	Lesson	Part	Routines/ Activities (1-5)	Levels of Support (1-3)	Error Correction (1-3 each)			
					A	B	C	D
		Comprehension						
		Vocabulary						
		Comprehension/Vocabulary						
		Fluency (other than timings)						
		Higher Order Thinking						
		Beyond the Book						
		Other						
		Fluency Hot/Cold Timings	**Total Potential Points** (circle one) 9 12		**Total Points Earned**			

III: CLASSROOM CHARACTERISTICS

	POINTS (1-4)
Classroom Climate	
Organization of Materials	
Classroom Routines	
Student Engagement	
Addressing Behavior Problems	
Lesson Pacing	

Teacher Monitoring	

IIa: Fluency Hot or Cold Timings

Hot or cold timing	**Yes (3 points)**	**No (0 points)**
Students work in partners or appropriate alternative.		
All students have a turn.		
All students record and graph WCPM.		
Self-reflection activity is completed as outlined after hot timing. (Leave blank if not applicable)		
Total possible points	(circle one) 9 12	
Total points earned		

IIb: Lesson fidelity

Unit: ____ **Lesson**: _____ **Teacher**: _____ **Period**: _____

Part (circle 1): Comp Vocab Comp/Vocab Fluency* HOT BTB Other

Read to Achieve Activities and Routines				
1 Teacher follows no or almost no activities and routines and/or modifications are rarely appropriate.	**2** Teacher follows a few activities and routines. And/or modifications are seldom appropriate.	**3** Teacher follows some activities and routines. And/or modifications are sometimes appropriate.	**4** Teacher follows most activities and routines. And modifications are usually appropriate.	**5** Teacher follows all or almost all activities and routines. And modifications are almost always appropriate.

Teacher Support		
1 Level of support is always or almost always too high or too low. That is, teacher support is usually too much or too little for what students appear to need.	**2** Level of support is sometimes too high or too low. That is, teacher support is sometimes too much or too little support for what students appear to need.	**3** Level of support is always or almost always appropriate. That is, support a) matches what the program calls for, b) is increased due to obvious student need for firming, or c) is thinned down because students are obviously "getting it."

Error Correction *(if no student errors, leave blank)*

	1 **Rarely…**	**2** **Sometimes…**	**3** **Always or almost always…**
A	• Errors are immediately addressed	• Errors are immediately addressed	• Errors are immediately addressed
B	• Errors are addressed accurately	• Errors are addressed accurately	• Errors are addressed accurately
C	• Students practice the correct answer or have another chance to get it right (e.g., they are not "let off the hook")	• Students practice the correct answer or have another chance to get it right (e.g., they are not "let off the hook")	• Students practice the correct answer or have another chance to get it right (e.g., they are not "let off the hook")
D	• Teacher tone is positive	• Teacher tone is positive	• Teacher tone is positive

*This rubric is not used for hot or cold timings.

Repeat this page for each "part" of the lesson.

III. Classroom Characteristics. *Complete this section during the last 5 minutes. If an issue cannot be rated (e.g., there were no behavior problems), leave that item blank.*

	1	2	3	4
Classroom Climate	Classroom is rarely or never characterized by positive affect (smiling, enthusiasm), positive communications and respect among teachers and students.	Classroom is sometimes characterized by positive affect (smiling, enthusiasm), positive communications and respect among teachers and students.	Classroom is often characterized by positive affect (smiling, enthusiasm), positive communications and respect among teachers and students.	Classroom is always characterized by positive affect (smiling, enthusiasm), positive communications and respect among teachers and students.
Organization of Materials	Very disorganized materials (students or teacher can't find things).	Somewhat disorganized materials (students and/or teacher sometimes cannot find things easily).	Mostly organized materials (students and teacher usually find things easily).	Very well organized materials (students and teacher almost always find things easily).
Classroom Routines	Routines are unclear to students and/or no routines are established. Many students are confused about what to do.	Some unclear or unestablished routines. Some students are confused about what to do.	Mostly clear and established routines. A few students may be confused or there may be 1 or 2 instances of lack of clarity.	Very clear and established routines. All students know what to do with rare exceptions.
Student Engagement	Few students actively participate throughout the lesson. Few students are on task.	Some students actively participate throughout the lesson. Some students are on task for most/all of the lesson.	Most students are actively participating throughout the entire lesson. Most students are on task for most of the lesson.	All students are actively participating throughout the entire lesson. All students are on task for the entire lesson with rare exceptions.
Addressing Behavior Problems	The teacher is always or almost always ineffective at addressing students' behavior problems.	The teacher is often ineffective at addressing students' behavior problems.	The teacher is usually effective at addressing students' behavior problems.	The teacher is always effective at addressing students' behavior problems. Or, no student behavior problems.
Lesson Pacing	The pace of the lesson is never appropriate; it is always too fast or too slow for students.	The pace of the lesson is often not appropriate; it is usually too fast or too slow for students.	The pace of the lesson is usually appropriate, with a few instances of being too fast or slow.	The pace of the lesson is always appropriate for students (or there is only a rare exception).
Teacher Monitoring	Teacher rarely or never monitors independent/ partner/group work (e.g., teacher sits at desk or does other things).	Teacher sometimes monitors independent/ partner/group work but sometimes does not.	Teacher usually monitors independent/partner/group work although her attention may turn away for a few minutes.	Teacher always monitors student work in independent/partner/group work.

Washington Striving Readers

Teacher Interview - Fall

Note: Site visitors should bring the pacing schedules to interviews.

Each Striving Readers teacher will be asked the questions in this protocol. Topics include teacher feedback on training, coaching, and both the *Phonics Blitz* and *Read to Achieve* programs. The information will be used only for evaluation purposes. Please be candid in your responses. Your individual responses will not be shared with anyone from your school or the Striving Readers program. Teacher responses are important as they help to describe both the successes and challenges of implementing these new programs.

Phonics Blitz

I have a few questions about your experience teaching *Phonics Blitz.*

How easy or challenging is teaching *Phonics Blitz*?
Why?
What, specifically, is challenging?

What *Phonics Blitz* lesson(s) are you teaching today?
(Consult pacing calendar and tell them how many days they are behind/ahead of schedule)

If behind schedule: What has slowed you down?
If ahead of schedule: What has allowed you to speed up?
If on schedule: Have you had any challenges following the pacing schedule?

What, modifications (or "tweaks"), if any, have you made when teaching *Phonics Blitz*?

Why?
Have you had to supplement *Phonics Blitz* with any other materials? Yes/No
If yes, please describe.

Think about the summer training you had for *Phonics Blitz.*

How well did that training prepare you to implement *Phonics Blitz*?
What if anything would have improved the training?

Have you received a visit and follow-up training from Linda Farrell (*Phonics Blitz*)? Yes/No

On a scale of 1-10 where 10 is the highest, how useful was this assistance?
Why?

Read to Achieve

I'd like to ask you about your experience teaching *Read to Achieve.*

How easy or challenging is teaching *Read to Achieve*?
Why?
What, specifically, is challenging?

What *Read to Achieve* lesson(s) are you teaching today? *(Separate answers for Group 1 and Group 2. Then, consult pacing calendar and calculate how many days they are behind/ahead of schedule)*

If behind schedule: What has slowed you down?

If ahead of schedule: What has allowed you to speed up?
If on schedule: Have you had any challenges following the pacing schedule?

What, modifications (a.k.a., "tweaks", if any, have you had to make when teaching *Read to Achieve*?
Why?
Have you had to supplement *Read to Achieve* with any other materials? Yes/No
If yes, describe.

Think about the summer training you had for *Read to Achieve*.
How well did that training prepare you to implement *Read to Achieve*?
What if anything would have improved the training?

Coaching

What kinds of assistance and support have you received from your coach so far this year?
Please provide specific example(s) of how the coach has helped you improve how you teach *Phonics Blitz*.
Please provide specific example(s) of how the coach has helped you improve how you teach *Read to Achieve*.

On a scale of 1-10, with 10 the highest, how useful has the coaching aspect of Striving Readers been for you? (1-10)
Why?

What, if anything, would you change about the coaching offered through Striving Readers?

Other questions

Have you shared anything that you've learned at these trainings with other teachers at your school?
Yes/No
If yes, explain specifically what have you shared and with whom.

Do you think students were accurately placed in these two programs? (Yes/Maybe/No)
Why or why not?

How many students are currently enrolled in each of your classes?

	Group 1 (PB + RtA) Or Group 2 (only RtA)?	Number of students enrolled
Period 1		
Period 2		
Period 3		
Period 4		
Period 5		
Period 6		

Washington Striving Readers

Teacher Interview - Winter

Note: Site visitors should remember to bring the pacing schedules to the interviews.

Phonics Blitz

When did you – or when will you – finish *Phonics Blitz*? (date)

a) How easy or challenging is teaching *Phonics Blitz*?
b) Why?
c) What, specifically, is challenging?

What, if any, modifications (a.k.a., "tweaks") have you made when teaching *Phonics Blitz*?

a) Why?
b) Have you had to supplement *Phonics Blitz* with any other materials? Yes/No
c) If yes, please describe.

What training and support do you need to implement *Phonics Blitz* well next year?

Read to Achieve

How easy or challenging is teaching *Read to Achieve*?
Why?
What, specifically, is challenging?

What *Read to Achieve* lesson(s) are you teaching today? *(Separate answers for Group 1 and Group 2. Then, consult pacing calendar and calculate how many days they are behind/ahead of schedule)*
Are you using the fast "skipping schedule"? If yes, describe how you are using it.
If behind schedule: What has slowed you down?
If ahead of schedule: What has allowed you to speed up?
If on schedule: Have you had any challenges following the pacing schedule?

What, if any, modifications ("tweaks") have you had to make when teaching *Read to Achieve*?
Why?
Have you had to supplement *Read to Achieve* with any other materials? Yes/No
If yes, describe.

Did you attend the December training about *Read to Achieve*? Yes/No
On a scale of 1-10 with ten being the highest, how useful was the training?
Why?

In the future, what additional training and support do you need to implement *Read to Achieve*?

Coaching

What kinds of assistance and support have you received from your coach since (date of last interview)?

Please provide specific example(s) of how the coach has helped you improve how you teach *Phonics Blitz*.

Please provide specific example(s) of how the coach has helped you improve how you teach *Read to Achieve*.

On a scale of 1-10 with 10 the highest, how useful has the coaching aspect of Striving Readers been? (1-10)

Why?

What would you change about the coaching offered through Striving Readers?

Other questions

Did you attend the AIMSweb data training in November? Yes/No

On a scale of 1 to 10 with ten being high, how useful was the AIMSweb training? (1-10)

Why?

How are you using the AIMSweb data from Striving Readers?

Have you shared anything that you've learned at these trainings with other teachers at your school?

Yes/No

If yes, explain specifically what have you shared and with whom.

Do you think students were accurately placed in these two programs? (Yes/Maybe/No)

Why or why not?

How many students are currently enrolled in each of your classes?

	Group 1 (PB + RtA) Or Group 2 (only RtA)?	Number of students enrolled
Period 1		
Period 2		
Period 3		
Period 4		
Period 5		
Period 6		

Note: Interviewers may also need to ask the teacher follow-up questions about the observations (e.g., for more context about teacher level of support in Read to Achieve).

Appendix B
Baseline Equivalence of Treatment and Control Groups

In a randomized controlled trial, randomization should ensure that the treatment and control groups are equivalent, or similar, on important student characteristics known to be related to student achievement. These characteristics include gender, student ethnicity, student receipt of free/reduced-price lunch (FRL), student receipt of special education services, English language learner (ELL) status and pretest scores. Because it is possible that randomization does not always yield equivalent groups, we conducted some basic analyses to see whether the treatment and control groups were, in fact, equivalent on these key demographic features and on pretest measures.

Tables B1 through B3 provide the results from our tests. The chi square (X^2) statistic indicates the magnitude of any difference between the two groups, and the *p*-value indicates whether the difference is statistically significant. No differences were statistically significant, meaning that treatment and control groups were similar on these characteristics.

These variables we included here are the same as used in our analysis. However, to take a closer look at any ethnic differences, here we unpacked the "other" category into its four parts. We did not test the statistical significance of any difference among these groups, because they were small and not used in our final analyses.

Table B1

Demographic Characteristics of the Gates-MacGinitie Analytic Samples by Group

Group	Demographic Characteristic	Treatment	Control	X^2	*P*
Total Sample N=358	Male	55%	59%	0.653	.419
	African American	15%	18%	0.500	.480
	Latino	13%	18%	1.739	.187
	White	44%	42%	0.239	.625
	Other	27%	22%	0.134	.714
	• American Indian	4%	3%	NA	NA
	• Asian	9%	7%	NA	NA
	• Pacific Islander	6%	6%	NA	NA
	• Multiracial	9%	7%	NA	NA
	FRL	59%	61%	0.134	.714
	Special Education	6%	6%	0.092	.761
	ELL	13%	11%	0.366	.545
Group 1 N=63	Male	47%	52%	0.141	.707
	African American	13%	23%	1.110	.292
	Latino	13%	19%	0.554	.457
	White	47%	39%	0.429	.613
	Other	28&	19%	0.668	.414
	• American Indian	13%	3%	NA	NA
	• Asian	3%	13%	NA	NA
	• Pacific Islander	13%	3%	NA	NA
	• Multiracial	0%	0%	NA	NA
	FRL	59%	62%	0.011	.916
	Special Education	13%	7%	0.669	.414
	ELL	28%	19%	0.668	.414
Group 2 N=295	Male	57%	61%	0.483	.487
	African American	16%	17%	0.830	.774
	Latino	13%	18%	1.230	.267
	White	44%	42%	0.056	.813
	Other	27%	23%	0.825	.364
	• American Indian	2%	3%	NA	NA
	• Asian	10%	5%	NA	NA
	• Pacific Islander	4%	6%	NA	NA
	• Multiracial	10%	8%	NA	NA
	FRL	59%	62%	0.202	.653
	Special Education	5%	5%	0.029	.864
	ELL	10%	9%	0.017	.895

Table B4 shows results of our test of baseline equivalence of the treatment and control groups for student achievement at pretest. We used linear regression to test these differences using the first model described in Chapter 4. This model included all student level covariates and included the school-group variable (which school students attended whether they were eligible to be in Group 1 or Group 2) as the blocking variable to account for the nesting of students within schools and groups. The *beta* (β) in the table indicates the magnitude of the difference between treatment and control groups, and the *p*-value indicates the statistical significance of this difference. No differences were statistically significant, indicating that treatment and control groups were similar on student achievement at pretest.

Table B4
Pretest Equivalence of the Analytic Sample

		Control Group	**Treatment Group**		
Total Sample	**N**	**Mean (SD)**	**Model-adjusted Mean (SD)**	***β***	***P*-value**
Gates-MacGinitie	358	496.50 (32.1)	495.69 (27.9)	-0.81	.792
Woodcock Reading Mastery word attack	357	98.76 (9.0)	98.4 (8.3)	-0.36	.681
MSP	401	377.79 (49.6)	374.41 (62.6)	-3.38	.516
		Control Group	**Treatment Group**		
Group 1	**N**	**Mean (SD)**	**Model-adjusted Mean (SD)**	***β***	***P*-value**
Gates-MacGinitie	63	478.03 (31.1)	480.3 (24.7)	2.27	.765
Woodcock Reading Mastery word attack	63	92.16 (6.5)	91.93 (6.8)	-0.23	.893
MSP	76	369.95 (63.4)	355.21 (104.7)	-14.74	.397
		Control Group	**Treatment Group**		
Group 2	**N**	**Mean (SD)**	**Model-adjusted Mean (SD)**	***β***	***P*-value**
Gates-MacGinitie	295	500.42 (31.1)	498.85 (27.1)	-1.57	.647
Woodcock Reading Mastery word attack	294	100.16 (8.9)	99.79 (7.9)	-0.37	.712
MSP	325	379.87 (45.8)	379.26 (45.9)	-0.61	.899

Appendix C
Detailed Regression Analysis Results

Tables C1 – C4 report in detail the results from the overall impact of the intervention on student reading achievement for the total sample: 358 students for *Gates-MacGinitie,* 357 students for the *Woodcock Reading Mastery* (word identification and word attack subtests), and 401 students for the *MSP*.

Table C1
Summary of Regression Analyses for Total Sample, Gates-MacGinitie

Variable	*B*	SE	DF	*t*	*p*
Constant	512.110	2.725	14	187.935	.000
Treatment	.716	1.946	343	.368	.713
Pretest	.462	.035	343	13.329	.000
Missing Pretest	-3.449	6.575	343	-.525	.600
White	9.746	2.035	343	4.789	.000
Free or Reduced Price Lunch	5.354	2.032	343	2.635	.009
School 1 Group 1	-13.208	5.152	343	-2.563	.011
School 2 Group 1	-19.572	7.969	343	-2.456	.015
School 3 Group 1	-14.345	5.216	343	-2.750	.006
School 4 Group 1	-17.567	5.132	343	-3.423	.001
School 5 Group 1	-23.892	8.597	343	-2.779	.006
School 2 Group 2	-6.371	3.977	343	-1.602	.110
School 3 Group 2	-15.507	3.349	343	-4.630	.000
School 4 Group 2	-9.725	3.382	343	-2.876	.004
School 5 Group 2	-6.891	3.418	343	-2.016	.045

Table C2
Summary of Regression Analyses for Total Sample, Woodcock Reading Mastery Word Identification

Variable	*B*	SE	DF	*t*	*p*
Constant	95.178	.925	13	102.847	.000
Treatment	-.299	.650	343	-.461	.645
Pretest	.573	.042	343	13.754	.000
Missing Pretest	.930	1.809	343	.514	.607
White	1.455	.669	343	2.175	.030
School 1 Group 1	-3.741	1.729	343	-2.164	.031
School 2 Group 1	-3.133	2.665	343	-1.176	.241
School 3 Group 1	-5.815	1.778	343	-3.272	.001
School 4 Group 1	-2.345	1.784	343	-1.314	.190
School 5 Group 1	-1.362	2.904	343	-.469	.639
School 2 Group 2	2.240	1.343	343	1.668	.096
School 3 Group 2	.193	1.127	343	.172	.864
School 4 Group 2	2.249	1.139	343	1.973	.049
School 5 Group 2	2.031	1.151	343	1.764	.079

Table C3
Summary of Regression Analyses for Total Sample, Woodcock Reading Mastery Word Attack

Variable	*B*	SE	DF	*t*	*p*
Constant	97.533	1.000	13	97.576	.000
Treatment	.680	.702	343	.969	.333
Pretest	.670	.045	343	14.894	.000
Missing Pretest	.123	1.954	343	.063	.950
White	2.605	.723	343	3.605	.000
School 1 Group 1	-2.079	1.867	343	-1.113	.266
School 2 Group 1	1.124	2.878	343	.390	.696
School 3 Group 1	.010	1.920	343	.005	.996
School 4 Group 1	-.622	1.927	343	-.323	.747
School 5 Group 1	3.759	3.137	343	1.199	.232
School 2 Group 2	4.284	1.451	343	2.953	.003
School 3 Group 2	.686	1.218	343	.563	.573
School 4 Group 2	1.606	1.231	343	1.305	.193
School 5 Group 2	4.218	1.244	343	3.392	.001

Table C4
Summary of Regression Analyses for Total Sample, MSP

Variable	*B*	SE	DF	*t*	*p*
Constant	173.654	17.815	16	9.748	.000
Treatment	3.075	1.550	384	1.984	.048
Pretest	.550	.046	384	12.014	.000
Missing Pretest	204.897	18.058	384	11.347	.000
Male	-4.759	1.611	384	-2.954	.003
African American	4.697	2.334	384	2.013	.045
White	3.765	1.776	384	2.120	.035
ELL	-7.481	2.640	384	-2.833	.005
School 1 Group 1	-8.317	3.882	384	-2.143	.033
School 2 Group 1	-23.144	6.708	384	-3.450	.001
School 3 Group 1	-5.149	4.092	384	-1.258	.209
School 4 Group 1	-1.363	4.106	384	-.332	.740
School 5 Group 1	-.494	5.932	384	-.083	.934
School 2 Group 2	-1.133	3.184	384	-.356	.722
School 3 Group 2	1.009	2.690	384	.375	.708
School 4 Group 2	-1.792	2.747	384	-.653	.514
School 5 Group 2	.431	2.752	384	.156	.876

Tables C5-C8 show more detail about the impact of the intervention on student reading achievement for Group 1: 63 students for *Gates-MacGinitie* and *Woodcock Reading Mastery* subtests and 76 students for the *MSP*.

Table C5

Summary of Regression Analyses for Group 1, Gates-MacGinitie

Variable	***B***	**SE**	**DF**	***t***	***p***
Constant	496.646	5.267	9	94.297	.000
Treatment	2.574	4.633	53	0.555	.581
Pretest	.363	.089	53	4.106	.000
Missing Pretest	-11.042	18.849	53	-0.586	.560
White	9.373	4.948	53	1.894	.064
Special education	-7.998	8.244	53	-0.970	.336
School 2 Group 1	-6.337	9.074	53	-0.698	.488
School 3 Group 1	-2.336	6.345	53	-0.368	.714
School 4 Group 1	-4.252	6.362	53	-0.668	.507
School 5 Group 1	-8.364	9.609	53	-0.870	.388

Table C6

Summary of Regression Analyses for Group 1, Woodcock Reading Mastery Word Identification

Variable	***B***	**SE**	**DF**	***t***	***p***
Constant	90.826	1.374	10	66.089	.000
Treatment	.768	1.246	52	0.616	.540
Pretest	.464	.107	52	4.321	.000
Missing Pretest	-6.495	3.178	52	-2.044	.046
Male	2.695	1.434	52	1.878	.066
Latino	4.816	1.942	52	2.480	.016
White	3.438	1.347	52	2.552	.014
School 2 Group 1	1.507	2.312	52	0.652	.517
School 3 Group 1	-1.323	1.754	52	-0.754	.454
School 4 Group 1	.301	1.787	52	0.168	.867
School 5 Group 1	2.066	2.481	52	0.833	.409

Table C7

Summary of Regression Analyses for Group 1, Woodcock Reading Mastery Word Attack

Variable	***B***	**SE**	**DF**	***t***	***p***
Constant	95.034	1.576	9	60.285	.000
Treatment	2.209	1.426	53	1.549	.127
Pretest	.768	.123	53	6.233	.000
Missing Pretest	-8.629	3.629	53	-2.378	.021
Male	3.293	1.540	53	2.138	.037
White	3.868	1.462	53	2.646	.011
School 2 Group 1	4.110	2.653	53	1.549	.127
School 3 Group 1	3.933	1.997	53	1.969	.054
School 4 Group 1	2.528	2.049	53	1.234	.223
School 5 Group 1	6.448	2.850	53	2.263	.028

Table C8
Summary of Regression Analyses for Group 1, MSP

Variable	*B*	SE	DF	*t*	*p*
Constant	141.766	32.594	8	4.349	.000
Treatment	1.678	2.829	67	0.593	.555
Pretest	.613	.085	67	7.243	.000
Missing Pretest	236.047	31.876	67	7.405	.000
ELL	-6.604	3.586	67	-1.842	.070
School 2 Group 1	-14.865	5.796	67	-2.565	.013
School 3 Group 1	4.645	3.788	67	1.226	.224
School 4 Group 1	6.821	3.749	67	1.819	.073
School 5 Group 1	7.327	5.112	67	1.433	.156

Tables C9-C12 show more detail about the overall impact of the intervention on student reading achievement for Group 2: 295 students for *Gates-MacGinitie*, 294 students for the *Woodcock Reading Mastery* subtests, and 325 students for the *MSP*.

Table C9
Summary of Regression Analyses for Group 2, Gates-MacGinitie

Variable	*B*	SE	DF	*t*	*p*
Constant	511.990	2.790	9	183.539	.000
Treatment	.554	2.161	285	0.256	.798
Pretest	.480	.038	285	12.676	.000
Missing Pretest	-2.918	7.080	285	-.412	.681
White	9.822	2.235	285	4.395	.000
Free or Reduced Price Lunch	6.043	2.231	285	2.708	.007
School 2 Group 2	-6.230	4.013	285	-1.552	.122
School 3 Group 2	-15.267	3.381	285	-4.515	.000
School 4 Group 2	-9.637	3.407	285	-2.828	.005
School 5 Group 2	-6.807	3.443	285	-1.977	.049

Table C10
Summary of Regression Analyses for Group 2, Woodcock Reading Mastery Word Identification

Variable	*B*	SE	DF	*t*	*p*
Constant	95.324	.957	14	99.588	.000
Treatment	-.247	.734	279	-0.336	.737
Pretest	.579	.045	279	12.869	.000
Missing Pretest	3.842	2.126	279	1.807	.072
School 2 Group 2	1.710	1.361	279	1.256	.210
School 3 Group 2	-.148	1.142	279	-0.129	.897
School 4 Group 2	2.013	1.164	279	1.729	.085
School 5 Group 2	1.730	1.176	279	1.471	.142

Table C11

Summary of Regression Analyses for Group 2, Woodcock Reading Mastery Word Attack

Variable	***B***	**SE**	**DF**	***t***	***p***
Constant	97.665	1.034	8	94.426	.000
Treatment	.565	.784	285	0.720	.472
Pretest	.657	.048	285	13.585	.000
Missing Pretest	2.989	2.270	285	1.317	.189
White	2.234	.808	285	2.764	.006
School 2 Group 2	4.020	1.479	285	2.719	.007
School 3 Group 2	.545	1.241	285	0.439	.661
School 4 Group 2	1.459	1.251	285	1.166	.245
School 5 Group 2	4.022	1.265	285	3.180	.002

Table C12

Summary of Regression Analyses for Group 2, MSP

Variable	***B***	**SE**	**DF**	***t***	***p***
Constant	178.357	20.297	11	8.787	.000
Treatment	3.273	1.778	313	1.840	.067
Pretest	.538	.052	313	10.311	.000
Missing Pretest	189.047	21.495	313	8.795	.000
Male	-6.005	1.871	313	-3.210	.001
African American	4.659	2.685	313	1.735	.084
White	4.520	2.014	313	2.244	.026
ELL	-7.252	3.345	313	-2.168	.031
School 2 Group 2	-1.500	3.319	313	-0.452	.651
School 3 Group 2	.875	2.800	313	0.313	.755
School 4 Group 2	-2.314	2.865	313	-0.808	.420
School 5 Group 2	.174	2.861	313	0.061	.952