Exploring the Relative Effectiveness of Reading Interventions for High School Students

Laura Lang and Joseph Torgesen
Florida State University, Florida, USA

William Vogel
Seminole County Public Schools, Sanford, Florida, USA

Carol Chanter
Scholastic Education, Winter Springs, Florida, USA

Evan Lefsky
Florida Department of Education, Tallahassee, Florida, USA

Yaacov Petscher
Florida State University, Florida, USA

Abstract: The purpose of this research was to explore the relative effectiveness of intensive reading interventions for struggling high school readers. A yearlong randomized control study was conducted to estimate causal effects, as measured by the criterion-referenced state assessment test, for 1,265 ninth-grade students in 89 classes across 7 high schools in a large school district. Students in the high risk group and the moderate risk group were randomly assigned to one of four intensive reading interventions (three new interventions and a “business as usual” control condition.) Results indicated that for all four interventions, gains made by students in the high risk group exceeded the benchmark for expected annual growth. For the moderate risk group, random effects mixed modeling showed that reliable differences were observed in the state outcome gain scores between two of the intensive interventions and the “business as usual” control condition (Glass’s adjusted $\Delta = .27, .30$).

Keywords: High school reading interventions, randomized experiment, adolescent literacy

Address correspondence to Laura Lang, University Center 4605, PO Box 30625-40, Tallahassee, FL 32306-2540, USA. E-mail: llang@fsu.edu
Over the past several years, adolescent literacy has become a focal point of concern for educators, researchers, and policymakers (Biancarosa & Snow, 2006). This is due, at least in part, to the adoption of standards-based state accountability systems and the 2001 passage of the federal No Child Left Behind Act (2002). Most state policies and the federal legislation mandate reforms in PK–12 education are designed to result in (a) higher student achievement, (b) elimination of the performance gap between higher and lower performing students (see http://www.ed.gov/policy/elsec/leg/esea02/index.html), and (c) more rigorous academic curriculum and coursework (Achieve, 2006). Despite the mandate to improve, the most recent national data indicate little, if any, progress in improving high school students’ reading performance or in closing the achievement gap (Grigg, Donahue, & Dion, 2007). National Assessment of Educational Progress data reported by the National Center for Educational Statistics suggest that the percentage of high school seniors performing below Basic in reading has increased rather than decreased, with 20% in this lowest category in 1992 compared to 27% in 2005, and that for both White and Black high school students, average reading performance has actually declined (Perie, Grigg, & Donahue, 2005). In other words, despite state and national mandates, most high schools have not been successful in meeting the instructional needs of students who enter ninth grade with less than proficient reading skills.

There is evidence, however, that high schools have been relatively successful in increasing the rigor of coursework and rates of participation in advanced courses (Shettle et al., 2007). This is good news for many students because it appears to increase the likelihood that they will enroll and persist in postsecondary education (Horn & Nunez, 2000). Unfortunately, when the demands of more rigorous coursework are not coupled with appropriate instructional supports and improvement efforts, the likelihood that struggling students will remain in high school until graduation is actually diminished (Heck & Mahoe, 2006).

THE INSTRUCTIONAL NEEDS OF STRUGGLING HIGH SCHOOL READERS

The scope and complexity of the challenges involved in addressing the instructional needs of struggling high school readers are detailed in a number of recent publications (Biancarosa & Snow, 2006; Hasselbring & Goin, 2004; National Association of Secondary School Principals, 2005; National Governors Association, 2005; Torgesen et al., 2007). These students are typically behind their peers in the development of a broad range of skills and knowledge required for the fluent comprehension of grade-level text. These students often struggle with accurate and fluent word reading, they frequently have limited vocabularies and content knowledge, and their flexible use of active reading comprehension strategies is almost always poorly developed (Hasselbring & Goin, 2004; Hock et al., in press).
Providing struggling high school readers with the sheer amount of instructional support and guided practice necessary to close the gap in all these essential areas is a significant challenge made even more difficult by the fact that many struggling readers have become disengaged from academics and reading (National Association of Secondary School Principals, 2005). McKenna, Kear, and Ellsworth (1995) discovered a steady decline in students’ attitudes toward academic, and even pleasure, reading as they moved up in grade level, which in turn resulted in decreased levels of reading frequency and an avoidance of difficult reading tasks (Bean, 2000). Despite the myriad issues, promising intensive interventions for struggling adolescent readers have been developed over the past few years. The effectiveness of these programs, however, has not been systematically demonstrated (Balfanz, McPartland, & Shaw, 2002; Biancarosa & Snow, 2006).

In addition to identifying effective intensive interventions, a second major challenge for high schools is to include reading instruction across content areas so that students are stimulated and explicitly taught the increasingly sophisticated comprehension strategies and reasoning skills required to meet the demands of coursework and texts that become more complex at each grade level (RAND Reading Study Group, 2002). In addition to helping students acquire active and flexible comprehension/reasoning strategies, instruction in the content areas must also provide stronger support for the development of academic vocabulary and content knowledge (Bulgren, 2004). Not only do students think more effectively within domains where their knowledge is deep (Shaffer, 2002), they are able to draw upon widening circles of content knowledge to construct meaning from text that assumes the reader will have related prior knowledge. Although research indicates integrating reading instruction in content area classes is necessary to meet the instructional needs of struggling high school readers, high school teachers generally do not see reading instruction as their responsibility (Roe, Stoodt, & Burns, 1998).

Given the depth and breadth of these challenges, researchers suggest that improving the capacity of high schools to meet the instructional needs of all students requires significant changes on many dimensions (Biancarosa & Snow, 2006). In this 3-year study, we focused on two of the most important aspects of improving instruction for struggling readers: (a) providing intense and focused interventions (Torgesen, 2005a), and (2) incorporating reading strategy instruction in two content area classes (RAND Reading Study Group, 2002). Professional development was provided to content area teachers to help them with reading strategy instruction because this study was done as part of a general effort to improve literacy instruction for adolescents in a school district that partnered with us in the research. The professional development and support provided to content area teachers was not extensive, and it acted as a general background condition that was common to all students who received reading interventions. The focus of this article is on the differential impact of four different interventions to which students of two broad reading levels were randomly assigned.
OVERVIEW OF A 3-YEAR STUDY

The foci of Year 1 of the study, 2004–2005, included building a partnership between district and school leaders and researchers who jointly planned and internally funded a study with three distal outcomes (see the appendix): (a) to provide evidence for practitioners to use in making decisions; (b) to increase reading achievement for all students, but especially those who enter high school with less than proficient reading skills; and (c) to contribute to the body of systematic knowledge regarding the impact of interventions for at-risk high school students. Four intensive reading intervention programs were selected for struggling readers. School leaders identified teachers to be trained in one the four intensive interventions. Publishers of the two commercially available interventions were asked to participate in the provision of materials, in conducting professional development for teachers and school leaders, and in developing fidelity of implementation checklists. A third treatment, developed by the state department of education, was selected and the author was asked to provide the same types of support to the project. During the first year, all of the participating high schools implemented an intensive reading intervention that was a continuation and refinement of what they were already doing, and that served as a “business as usual” control for the year in which impacts were examined. It is important to note that the control condition was considered a fourth approach to reading intervention rather than the absence of treatment.

During Year 2 of the study, 2005–2006, teachers of 9th- and 10th-grade struggling readers implemented one of the four intensive reading interventions during a 90-min block that replaced students’ elective classes. Further, science and social teachers began integrating content enhancement and student-focused reading strategies for all 9th- and 10th-grade students (Bulgren, Schumaker, & Deshler, 1988; Ellis, 1999; Lenz & Deshler, 2000; Schumaker & Deshler, 1994). Intensive reading teachers received coaching and feedback, related to both fidelity and quality of implementation, provided by the school-level reading coach assigned to each school, the project coordinator, the publishers of the two commercially developed interventions, and the author of the third. Professional development continued throughout the year for teachers, and intervention-specific monthly support meetings were held to address concerns. Issues related to unforeseen scheduling requirements, random assignment at the student level and obtaining consent for participation were also addressed. The study procedures were fine-tuned at the end of Year 2, including an exclusive focus on 9th-grade students. The school district also agreed to provide small incentives for students to increase the return of consent forms so that a more representative sample could be obtained.

In Year 3, 2006–2007, a randomized field trial was undertaken to investigate the following questions:

1. For high risk struggling readers:
a. Are there differences in the impact of one or more of the three interventions when compared to that of the business as usual intervention control group?

b. Are reading gains greater for students in the intervention groups compared to Florida students statewide?

2. For moderate risk struggling readers:

a. Are there differences in the impact of one or more of the three interventions when compared to that of the business as usual intervention control group?

b. Are reading gains greater for students in the intervention groups compared to Florida students statewide?

**METHOD**

**Participants**

Seven comprehensive high schools in a large Florida district were included in the study. Within these schools, 1,265 ninth-grade students were identified as struggling readers based on prior year (2005–2006) reading performance on the Florida Comprehensive Assessment Test (FCAT). Preliminary review of the sample data for the 1,265 students indicated that approximately 5% of the pretest data (i.e., 63 cases) and 8% of the posttest data (i.e., 101 cases) were missing, thus, we used multiple imputation (MI) to estimate missing data. The presence of missing data is a typical phenomenon observed in most research, and a variety of methods exist to handle the problems that arise when data for participants are missing. Although methods such as casewise deletion, mean replacement, and single imputation are often used to correct for missingness, these procedures characteristically result in biased estimates of treatment effects, liberal observed p-values, and poor statements about statistical significance (Little & Rubin, 2002). MI is a technique for handling missing data that is regarded as superior to other methods. An MI procedure involves the simulation of a particular number of datasets, specified by the user, and estimates the between- and within-imputation variances. The results from the all imputed data files are combined and will generally produce a valid estimate of the true treatment effect (Rubin, 1996).

To apply MI, it is important to assume that the data are missing at random. Thus, the data were examined to determine if the probability of missing data on either the pretest or posttest of the FCAT was related to its particular value. In addition, within intervention groups, data appeared to be missing at random. Little’s (1988) test for missingness was used to test the extent to which data were truly missing at random. Findings indicated that the data were missing at random, \( \chi^2(11) = 807.98, p < .001 \), and that multiple imputation was a viable option for handling the missing cases. Further, the correlation
between missingness and treatment was .02, providing evidence that cases were specifically unrelated to treatment conditions.

Studentized residuals in the multiply imputed data greater than an absolute value of 2.5 (n = 68) were examined with respect to influencing the results. Each outlier was studied within each group observed in order to understand the departure from the rest of the data. All 68 cases were scores that did not properly appear to be a part of the population of interest and were removed from the sample, leaving a reduced sample of 1,197 participants. Removal of the outliers did not significantly adjust the demographics of the sample, with the proportions of students’ characteristics from the across risk and treatment groups approximating the full sample mean proportions (Table 1).

Of the 1,197 students in the reduced sample, 385 were identified as Level 1, which indicated that students were reading below a fourth-grade reading level, and 812 as Level 2, which suggests that these students were reading between a fourth- and sixth-grade level (Florida Department of Education, 2007b). Slightly more were female (n = 642, 50.8%), and in terms of ethnicity, 50.6% of the sample was identified as Caucasian, followed by 20.4% as Latino, 20.4% as Black, and 8.6% as other. Forty-three percent of the participants were identified as eligible for free and/or reduced lunch and 12.2% as English language learners. Students with disabilities, unless prohibited by their Individualized Educational Plan, and students with limited English proficiency were included in this sample.

**Measures**

The FCATest–Sunshine State Standards (FCAT–SSS) is a component of Florida’s testing program designed to assess student achievement in reading, writing, mathematics, and science represented in Florida’s SSS (Florida Department of Education, 2001a). The SSS Reading portion of the FCAT is a group administered, criterion-referenced test consisting of six to eight informational and literary reading passages (Florida Department of Education, 2005). In each grade (3–10), students respond to between 6 and 11 multiple-choice items for each passage and are assessed across four content clusters: reading comprehension in the areas of words and phrases in context, main idea, comparison/cause and effect, and reference and research. In Grades 4, 8, and 10, constructed-response items are included. Reliability for the FCAT–SSS has been shown to be high at .90; moreover, test score content and concurrent validity have been established through a series of expert panel reviews and data analysis (Florida Department of Education, 2001a). The construct validity of the FCAT–SSS as a comprehensive assessment of reading outcomes recently received strong support in an empirical analysis of its relationships with a variety of other reading comprehension, language, and basic reading measures (Schatschneider et al., 2004). It
Table 1. Full and reduced sample demographic characteristics

<table>
<thead>
<tr>
<th></th>
<th>Full Sample&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Reduced Sample&lt;sup&gt;d&lt;/sup&gt;</th>
<th>Reduced Sample–High Risk&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Reduced Sample–Moderate Risk&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reach&lt;sup&gt;e&lt;/sup&gt;</td>
<td>READ 180&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>% Male</td>
<td>49</td>
<td>49</td>
<td>49</td>
<td>50</td>
</tr>
<tr>
<td>% White</td>
<td>51</td>
<td>51</td>
<td>44</td>
<td>48</td>
</tr>
<tr>
<td>% Black</td>
<td>20</td>
<td>20</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td>% Latino</td>
<td>20</td>
<td>20</td>
<td>28</td>
<td>20</td>
</tr>
<tr>
<td>% FRL</td>
<td>58</td>
<td>58</td>
<td>48</td>
<td>52</td>
</tr>
<tr>
<td>% ELL</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Note. RISE = Reading Intervention through Strategy Enhancement; SOAR = School Offered Accelerated Reading; FRL = free or reduced lunch; ELL = English language learner.

<sup>a</sup>n = 385. <sup>b</sup>n = 812. <sup>c</sup>n = 1,265. <sup>d</sup>n = 1,197. <sup>e</sup>n = 91. <sup>f</sup>n = 100. <sup>g</sup>n = 104. <sup>h</sup>n = 90. <sup>i</sup>n = 199. <sup>j</sup>n = 207. <sup>k</sup>n = 204. <sup>l</sup>n = 202.
is important to note that the FCAT–SSS is the primary measure in Florida’s accountability system and its results are important to school leaders and teachers.

In addition to the scale score, a developmental scale score (FCAT–DSS; range = 0–3,000), established through linking studies (Florida Department of Education, 2001b), allows one to monitor academic growth each year. The FCAT–DSS was created through the use of external (not contributing to a student’s test scores), anchor tests composed of adjacent grade items embedded within each grade level’s test. Using item response theory scaling methodologies, a linear adjustment was made to each grade level test’s scale. Expected learning gains in the form of scale score point gains on this vertical scale were established for each grade level as an element in school accountability measures. The DSS score from the FCAT was used in evaluating the impact of the interventions in this study because it allowed us to directly estimate student growth from one year to the next in the reading construct (reading comprehension) assessed by the FCAT.

Students completing the reading portion of the FCAT–SSS are placed in one of five performance levels based on a within-grade scale score that ranges from 100 to 500. Achievement Level 1 and Achievement Level 2 reflect below-grade-level performance in reading, with Level 1 being the lowest level of reading performance and Level 2 indicating below-grade-level performance. Levels 3, 4, and 5 represent proficiency in reading comprehension at or above grade-level standards. In an earlier study examining the relationship of FCAT–SSS Reading to other reading measures for a statewide representative sample of 7th- and 10th-grade students (Schatschneider et al., 2004), the average score on the Stanford Achievement Test, Ninth Edition (Stanford 9) of students at Achievement Level 1 in 10th grade was the 25th percentile. The corresponding average score for students at Achievement Level 2 was the 44th percentile. The overall correlation between the Stanford 9 and the FCAT–SSS scores was .78 at 7th grade and .74 at 10th grade. The overall conclusion from the independent evaluation of the FCAT (Schatschneider et al., 2004) is that it has strong concurrent validity when compared to other measures of reading comprehension, and that there is stability in the core reading constructs it assesses between 7th and 10th grades.

In our study Achievement Level 1, hereafter referred to as High Risk, \(n = 385\), and Achievement Level 2, hereafter referred to as Moderate Risk \(n = 812\), students were analyzed as distinct groups, with risk treated as random coefficient for the purpose of statistical modeling.

**Interventions**

Students were randomly assigned to one of the three treatments or a “business as usual” school-designed intervention control group for the entire 2006–2007 school year. Each of the seven schools offered varying numbers of 90-min
intervention classes, based on their capacity, to groups of no more than 21 students, with 87 individual classes taught by 31 teachers. Teachers were assigned to teach an intensive intervention by school administrators according to routine processes and in light of state certification and prior experience. No teacher taught more than one of the intensive intervention programs, and 28 of the teachers were implementing the same program taught during the 2nd year of the study. Of the teachers who did not teach the same intervention for 2 years, 1 teacher taught one intervention during the 2nd year and a different intervention in the 3rd year, 1 taught the control condition in the 2nd year and an intervention in the 3rd year of the study, and 1 teacher joined the study as an intervention teacher during the 3rd year of the study. In other words, all but 3 teachers had a full-year professional development and experience with the specific curriculum implemented during the randomized field trial conducted during the 2006–2007 school year. In addition, all students in ninth grade, including study participants, were enrolled in science and social studies classes taught by teachers who employed content enhancement routines and student-focused reading strategies (Bulgren et al., 1988; Deshler & Roth, 2000; Ellis, 1999; Lenz & Deshler, 2000; McFarland, 1998; Schumaker & Deshler, 1994).

The four approaches to intensive intervention, or conditions, included READ 180 (http://teacher.scholastic.com/products/read180), the REACH System 2002 (REACH; https://sraonline.com/products.html?tid=1&sid=82), Reading Intervention through Strategy Enhancement (RISE; Lefsky, 2004), and School Offered Accelerated Reading (SOAR). Although each of these approaches contained elements that are consistent with the findings of scientifically based research in reading, only READ 180 (Smith, Rissman, & Grek, 2004) and elements of the REACH System (Rissman, 2004) were the subject of prior systematic research. However, few studies included random assignment methodologies. All programs were monitored for implementation fidelity at least four times during the 2006–2007 school year. These observations were conducted by program experts, district literacy coaches, and school-based and district administrator.

As described next, each approach contains unique instructional features and strategies.

READ 180

READ 180 is a 90-min-a-day intensive reading program designed to meet the needs of students whose reading achievement is below the proficient level. The 90-min instructional model is designed to provide students with the opportunity to achieve reading success through a combination of instructional, modeled, and independent reading components. The program begins with 20 min of teacher-led, whole-group instruction followed by three 20-min rotations. The rotations last for a total of 60 min and include (a) small-group direct instruction,
(b) READ 180 software, and (c) independent and modeled reading. Once all rotations are complete, the class convenes for 10 min of whole group wrap-up.

During the first 20 min of READ 180, teachers use the rBook teaching system to teach reading skills and strategies, vocabulary and word study, and writing and grammar. The nine rBook workshops cover content-area nonfiction, with a high percentage of science and social studies passages as well as grade-appropriate, high-interest literature. Students begin each workshop viewing videos intended to anchor instruction by helping students build background and activate prior knowledge to increase their ability to create mental models from text. They then explore the workshop topic with the goal of building high-utility academic vocabulary that students will encounter across content areas. Comprehension instruction is designed to scaffold students’ access to text while actively engaging them. Students review vocabulary and are instructed to use high-utility academic words expressively—both orally and in writing. Writing instruction in the rBook is intended to guide struggling writers to master key types of writing and help to prepare them for standardized tests.

READ 180 topic software is designed to engage students in individualized and adjusted instruction intended to develop skills in decoding, fluency, vocabulary, comprehension, and spelling. Adaptive technology measures students’ automaticity in reading and provides appropriate practice activities based on an individual student’s needs. In Modeled and Independent Reading, students practice fluency and reading comprehension skills through reading READ 180 Audiobooks and leveled Paperbacks. Students use Audiobooks to hear modeled fluent reading as they track the text. The Audiobooks include a virtual reading coach who steps in to model comprehension, vocabulary, and self-monitoring strategies. Students’ Lexile scores (determined by the Scholastic Reading Inventory) and areas of interest are used to identify titles in the READ 180 Paperbacks that they can successfully read independently. Each READ 180 session ends with a 10-min summary activity. During this time, teachers are instructed to summarize and review skills with the whole class.

REACH

REACH is composed of three programs that have been integrated into one comprehensive reading/language arts program. REACH—for Grades 6 through 12—is designed to accelerate learning for students whose reading achievements are significantly below grade level. The programs included in REACH are Corrective Reading, Reasoning and Writing, and Spelling Through Morphographs. REACH also includes placement and progress assessments, and collections of high-interest chapter books.

REACH is designed to provide explicit, systematic, intensive instruction to enable students to progress to grade-level standards and to participate in a comprehensive reading/language-arts program at their grade level. The direct instruction approach found in REACH is designed to assist students in acquiring
critical skills and strategies by using certain principles. Each skill is designed
to be taught through teacher modeling or demonstration and practiced using
numerous and varied examples until it is mastered. Frequent student–teacher
interactions are intended to provide ongoing monitoring of student progress.

Students are placed in the REACH System according to their skill level
as determined by a diagnostic test that indicates reading and language arts
proficiency. Tests are also administered every 5 to 10 lessons, depending on the
program and level, to provide teachers student data needed to judge the progress
of individual students, identify concepts needing extra work, and target specific
exercises for students who need extra help.

Corrective Reading Decoding A teaches basic print and phonemic aware-
ness, whereas Corrective Reading Decoding B1 reviews the critical phonemic
awareness skills for reading the basic long and short vowel letter–sound re-
lationships, and the most common consonant digraphs. Decoding B1 is de-
signed to counter the common guessing strategy that many struggling read-
ers use—a strategy that too often results in reading errors and, consequently,
misunderstandings. Corrective Reading Decoding B2 is designed to build flu-
ency and accuracy as students become engaged in more interesting stories
similar to trade books. Comprehension begins to be integrated into decoding
as students are held more accountable for recalling details and major ideas. By
the end of Decoding B2, students are reading more complex text at a minimum
rate of 110 words per minute with 98% accuracy. Corrective Reading Decoding
C finishes the decoding sequence and provides a transition to outside reading.
The emphasis of Decoding C is on building word knowledge, word analysis
techniques, and the ability to comprehend more difficult expository passages.
By the end of Level C, students should be reading at a rate of 130 words per
minute with 98% accuracy for a sustained period of 5 min.

Corrective Reading Comprehension A is an oral language program that
requires no reading ability. Comprehension A focuses on the world knowledge,
vocabulary, and thinking–reasoning skills that provide a foundation for com-
prehension. Corrective Reading Comprehension B uses content from science
and social studies to teach the skills necessary for students to comprehend
expository text and respond to written questions that involve deductions and
rule applications. Corrective Reading Comprehension C teaches students to
use higher order thinking skills independently. They are taught to infer def-
initions from context, read for basic information, recognize the main idea,
draw conclusions from basic evidence, and identify contradictions and faulty
arguments.

Reasoning and Writing B uses storytelling and other activities to help
students develop an awareness of narrative structure and logical connection.
Students are taught higher level comprehension skills, such as cause-and-effect
reasoning, predicting story events based on an understanding of character, and
how perspective or point of view can impact meaning. Reasoning and Writing
C provides opportunities for students to apply what they have learned about
communication and reporting. Writing is presented as an ongoing process that begins with drafting and leads through stages of revising, editing, and proofreading. Reasoning and Writing D emphasizes expository writing—the kind of writing most often done in everyday life. Reasoning and Writing E is designed to assist students in reaching a level of sophistication in writing commensurate with usual expectations for middle school students. Reasoning and Writing F extends the skills begun in Level E as it teaches students to use a variety of reference material to develop complex arguments, examine patterns and write critiques.

Spelling Through Morphographs is a spelling program designed for students in Grades 6 through 12 that teaches students a set of specific strategies that deal with the structure of words. This approach to spelling teaches student to spell units (bases and affixes) of words and to put them together to spell words. The strategies taught in the program, combined with repeated practice and application, teach students to spell unfamiliar words and to remember familiar words more successfully than they would by using other spelling methods.

High Interest Readers (published by High Noon Books) are designed to enhance the learning process, and Ravenscourt Books provide additional support for boosting reading fluency and comprehension skills. Corresponding Fluency Audiocassettes and CDs are designed to assist students in learning correct phrasing, intonation, and expression while improving their reading rate. Evaluation and tracking software include comprehension quizzes that are intended to assess student progress.

RISE

RISE (Lefsky, 2004) is an intervention guided by the philosophy that teachers—given time, resources, and strong professional development support—can create effective curriculum that is engaging and provides remediation for struggling adolescent readers. It is designed to address the instructional needs of adolescent readers with poor literacy skills through a planned yet flexible approach to intervention. RISE teachers are expected to differentiate and alter the instruction when classroom-based assessments indicate that individual students are not making progress in the intervention. According to Lefsky, the key to successful implementation of RISE is strong professional development and up-front thinking involved in planning the text set unit.

Five questions guided the creation of curriculum in the RISE intervention:

1. Are all students provided many books they can read and want to read?
2. When students read/write, are they encouraged to write about what they know and care about?
3. Are all students given adequate time to explore topics and themes through reading, writing, and discussion?
4. Are students provided with explicit instruction to assist them with reading, writing, and discussion of text?
5. Do all of students get opportunities to demonstrate and use their strengths in reading, writing, and discussion, or do these activities only accentuate their weaknesses?

The foundation for all phases of the RISE daily instruction is the text set unit plan. A text set is a collection of related texts of varying genres and reading levels (Hamman, 1995). By varying the length and level of reading difficulty of the texts students read, the RISE intervention is designed to ensure that students are challenged but not frustrated. The use of multiple texts is intended to motivate adolescents to take an interest in the topic of study; see that it is relevant to their lives (Luke & Freebody, 1997); and to build background knowledge, vocabulary, and strategies for monitoring and improving comprehension (Duke & Pearson, 2002).

As part of the RISE intervention, teachers are provided professional development on both the creation of text set units and the foundations of differentiated literacy instruction. Each RISE teacher creates themed units, based on the interests of their students, which include a variety of text, plans for differentiated instruction, as well as plans to assess students’ progress.

The daily components of the RISE Intervention include independent reading practice, whole group lesson and daily small group differentiated instruction.

**Independent Reading Practice.** Students in the RISE classroom are instructed to devote the first 15 to 20 min of each class to independent reading practice monitored by the teacher. This practice occurs in texts selected by students from the classroom library or other source. The focus of independent reading practice is independent-level pleasure reading or reading related to the topic being studied in the text set unit. The intent of this time is for students to receive successful practice; therefore, it is important that the teacher monitor student reading to ensure that the book selected is not too difficult and that the student is making progress toward completion with full comprehension. Teachers are provided professional development on how to select high-interest, level-appropriate texts for their classrooms, as well as how to monitor the progress of students during independent reading practice with fluency and comprehension checks. This time also provides teachers with the opportunity to individually conference with students on a daily basis.

**Whole Group Lesson.** After completing the independent reading practice, the teacher provides whole class instruction or facilitates a class discussion on a common piece of text. During this time, the teacher might model a strategy or skill for students through a think-aloud or interactive read-aloud from the text set while providing students the opportunity to practice the strategies in
engaging yet challenging text. The focus of this aspect of the lesson is building interest, vocabulary, background knowledge, and strategic thinking on a topic being studied in the text set unit.

Daily Small-Group Differentiated Instruction. As an extension of the whole group lesson or discussion, students participate in three to four different workstations. Each workstation consists of four to five students and lasts 15 to 20 min. One small-group workstation is led by the teacher, another by students, and a third group participates in either a technology program or independent work, such as extended research in an area of individual interest or a listening station. The teacher-led station provides students the opportunity to receive differentiated instruction in specific areas of need and allows the teacher to provide more focused assessment and intervention. During the student-led station, students participate in reciprocal teaching, literature circles, and text-based discussions from the text set unit.

SOAR

SOAR was the title of the school-based intervention considered as the “business as usual” control. Some of the materials utilized in the SOAR classrooms included, but were not limited to *The Reading & Writing Sourcebook* by Great Source, *The Reader’s Handbook* by Great Source, *Reading Nonfiction* by Jamestown, and the *Daybook of Critical Reading and Writing* by Great Source. In addition to general reading instruction based on these materials, the SOAR classes also included FCAT preparatory activities specifically aligned with the SSS and Benchmarks. These preparatory activities were provided through a software program called FCAT Explorer (Florida Department of Education, 2007a). This type of practice provided students opportunities to answer questions based on the types of text (70% informational and 30% literary) and length of passages (range of words = 300–1,400; average number of words = 800) they would encounter on the ninth-grade test (Florida Department of Education, 2005). However, in that 80% or more of the ninth-grade test items were moderate to high in terms of cognitive complexity (Florida Department of Education, 2005), “teaching the test” was not possible (Greene, Winters, & Forster, 2004; Hassler, Buck, & Torgesen, 2004).

Experimental Design

Students scoring high risk or moderate risk on the prior year (2005–2006) reading portion of the FCAT–SSS were randomly assigned, within school and within level, to classrooms where one of four intensive reading interventions were taught (REACH, READ 180, RISE, or SOAR), using a Latin-square design that began with the lowest performing student in each school and continued
until all available seats in each intervention classroom at a given school were filled. In addition to assigning students to intervention classes based on their risk level according to prior year FCAT scores, students who were randomly assigned to REACH were administered the pretest included in the curriculum and rescheduled to intervention classes based on their identified instructional levels. Separate classes for High Risk and Moderate Risk students were maintained during this process. Overall, there were 290 students assigned to REACH (24.2%), 307 students assigned to READ 180 (25.6%), 308 students assigned to RISE (25.7%), and 292 students assigned to the SOAR (24.4%). The distribution of random assignment based on prior-year FCAT achievement resulted in an equitable percentage of students, by risk level, assigned to REACH (high risk = 24.5%, moderate risk = 23.6%), READ 180 (high risk = 25.5%, moderate risk = 26%), RISE (high risk = 25.1%, moderate risk = 27%), and the SOAR groups (high risk = 24.8%, moderate risk = 23.4%).

RESULTS

The unconditional and conditional mixed models were fit to 10 multiply imputed data sets, created using the MI procedure in SAS. The fitted means, variance components, significance tests, and between-group contrasts were calculated by using SAS PROC MIANALYZE, which combined the results across the imputed datasets. The MI procedure was conducted using a Markov Chain Monte Carlo method with a single chain and created 10 imputations. To improve the precision of estimates, the imputation was run within each risk by treatment group. Imputed results did not differ substantially from analyses conducted with the original data set.

Nature and Variability in Reading Scores by Risk Level and Intervention Group

A summary of the mean reading performance for each group by risk level is reported in Table 2. The gains made across all instructional conditions by the high risk students ranged from 124.20 DSS points for the READ 180 group to 170.52 DSS points for students receiving SOAR, whereas Moderate Risk students’ gains ranged from 69.85 DSS points for SOAR to 104.53 DSS points for READ 180. To ascertain the extent to which these gains might be related to fidelity of implementation, data related to quarterly observation checks, conducted by two trained district reading specialists for each intensive intervention teacher, were reviewed. The results indicated high levels of fidelity to treatments, with 90% or higher on-model indicators met for all teachers and each of the interventions throughout the study.
Table 2. Descriptive statistics for interventions by risk level

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th></th>
<th>Posttest</th>
<th></th>
<th>Gain Score</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>High risk</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>REACH</td>
<td>1567.96</td>
<td>139.96</td>
<td>1706.00</td>
<td>160.52</td>
<td>138.04</td>
<td>172.96</td>
</tr>
<tr>
<td>READ 180</td>
<td>1562.25</td>
<td>136.48</td>
<td>1686.45</td>
<td>196.92</td>
<td>124.20</td>
<td>190.88</td>
</tr>
<tr>
<td>RISE</td>
<td>1549.71</td>
<td>174.69</td>
<td>1710.13</td>
<td>180.33</td>
<td>160.42</td>
<td>164.09</td>
</tr>
<tr>
<td>SOAR</td>
<td>1558.69</td>
<td>247.17</td>
<td>1729.21</td>
<td>236.27</td>
<td>170.52</td>
<td>172.58</td>
</tr>
<tr>
<td>Moderate risk</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>REACH</td>
<td>1794.62</td>
<td>51.84</td>
<td>1871.81</td>
<td>156.09</td>
<td>77.19</td>
<td>145.98</td>
</tr>
<tr>
<td>READ 180</td>
<td>1800.27</td>
<td>52.36</td>
<td>1904.80</td>
<td>134.15</td>
<td>104.53</td>
<td>126.16</td>
</tr>
<tr>
<td>RISE</td>
<td>1800.38</td>
<td>50.10</td>
<td>1902.75</td>
<td>128.83</td>
<td>102.37</td>
<td>119.58</td>
</tr>
<tr>
<td>SOAR</td>
<td>1800.24</td>
<td>49.20</td>
<td>1870.09</td>
<td>130.09</td>
<td>69.85</td>
<td>119.57</td>
</tr>
</tbody>
</table>

Note. RISE = Reading Intervention through Strategy Enhancement; SOAR = School Offered Accelerated Reading.

Because the average initial status scores across interventions slightly varied within each risk group, Glass’s adjusted delta was used to determine the percentage of nonoverlap in the gain score distributions (Cooper & Hedges, 1994). When the proportion of nonoverlap between the distributions for a two-group comparison is less than 14.7%, the difference is considered to be practically unimportant (Cohen, 1988). Results for interventions in the high risk group revealed that the range of nonoverlap was between 0.0% and 7.7%, small enough to ensure that groups were not reliably different from each other. Similarly, students across the four intervention groups in the moderate risk varied very little in their initial status with no practically important nonoverlap (0.0–5.2%).

Effects of Interventions on Students’ FCAT Reading Gains

A $2 \times 4$ (Risk Level $\times$ Intervention Group) linear mixed model with random coefficients was used to model students’ gains in FCAT reading developmental scale scores using PROC MIXED in SAS 9.1. Because students’ scores were nested within classrooms, failure to account for the nested nature of the data could lead to misestimated standard errors. Although one might conceive that data in our study could be further nested within schools, the nature of the data is such that there are relatively few schools ($n = 7$) with which to examine between-school variance in the outcome. Analyzing the data in such a way would have resulted in estimates that were likely unreliable and could not have been extended in inference to a larger population of schools.

Because the FCAT is a criterion referenced test that has gone through IRT-based cross-grade scaling, we computed and used as our outcome students’
FCAT DSS gain score. The simple difference score between the posttest (2006–2007) and the pretest (2005–2006) FCAT DSS score was calculated to observe the amount of progress in achievement made by students in the various treatment conditions (see Table 2). Under many conditions, a residualized change score may be used to covary preexisting variability in the pretest score. Although covarying the pretest is appropriate in designs where randomization occurs (Jamieson, 1999; Rogosa, 1988), when large pretest variability is not observed, it is not necessary to use a residualized change score and the gain score is a viable estimator to use (Rogosa, Brandt, & Zimowski, 1982; Zumbo, 1999).

Moreover, using gain scores allowed us to compare gain scores for the students in the four intervention groups with statewide gains for all students in Florida for the same year. The average statewide reported mean gain from the end of eighth to the end of ninth grade for students was 66 points (Florida Department of Education, 2008) for the 2006–2007 school year. A DSS gain of 77 points served as the benchmark for adequate yearly growth from end of eighth grade to end of ninth grade for all performance levels (Florida Department of Education, 2006) for the same year.

Beginning with an unconditional mixed model (see Equation 1),

\[ Y_{ij} = \gamma_{00} + u_{0j} + r_{ij} \]  

which is described, such that \( Y_{ij} \) is the average FCAT DSS gain score for student \( i \) in class \( j \), \( \gamma_{00} \) represented the grand-mean gain score for students, and \( u_{0j} \) and \( r_{ij} \) were the random components associated with the classroom and student residuals. The ICCs for the unconditional and conditional two-level models were calculated using

\[ \frac{\tau_{00}}{\tau_{00} + \sigma^2}. \]  

Then the other variables were included (see Equations 3 and 4).

\[ Y_{ij} = \gamma_{00} + \gamma_{01}(Risk)_{ij} + u_{0j} + u_{1j}(Risk) + r_{ij} \]  

\[ Y_{ij} = \gamma_{00} + \gamma_{01}(Rise)_j + \gamma_{02}(Reach)_j + \gamma_{03}(Read - 180)_j + \gamma_{10}(Risk)_{ij} + \gamma_{11}(Rise \times Risk)_j + \gamma_{12}(Reach \times Risk)_j + \gamma_{13}(Read - 180 \times Risk)_j + u_{0j} + u_{1j}(Risk) + r_{ij} \]  

The student covariate model (see Equation 3) is an extension of the unconditional model, where students’ group mean gains were a function of either \( \gamma_{00} \) (i.e., moderate risk group mean) or \( \gamma_{10} \) (i.e., high risk group mean), where for the variable \( risk \), students in the high risk group are coded 1 students in the moderate risk group are coded 0. The addition of \( u_{1j}(Risk) \) denoted the random effect of risk status. Last, the classroom covariate model (see Equation 4)
introduced the dummy codes for the main of group assignment main effect of group assignment (i.e., $\gamma_{01} - \gamma_{03}$), where each group was dummy coded, with SOAR, the school-designed intervention, as the fixed reference group. $\gamma_{11} - \gamma_{13}$ represented the interactions between risk status and intervention group. Dummy codes for each school (with School 1 as the fixed referent) were added to the classroom model but did not significantly predict student outcomes and so were trimmed from the model to preserve parsimony. Using these models we compared the effect of each of the four interventions for the two student groups by using contrast statements in PROC MIXED.

Unconditional and Student Covariate Modeling

From the unconditional model (Equation 1), the results suggested that across all students and conditions, the average class gain was approximately 96 points ($SE = 6.4, p < .001$) from year to year. Further, 24% of the variance in FCAT DSS gains was between classrooms and was statistically significant ($\chi^2 = 107.88, p < .01$). This suggested that modeling covariates to explain the variance in FCAT gain was warranted. When covarying level of student risk, as described in Equation 3, the mean classroom gain for students that were at a moderate level of risk before the intervention began was 85 points. Mean classroom gains for the most struggling, high-risk students, were 118, 33 points greater than the students who were at moderate risk ($SE = 2.70, p < .01$). The amount of variance between classes dropped 14% by modeling student level of risk; however, a significant portion of variability in gain still existed ($\chi^2 = 92.64, p < .01$). With FCAT DSS gains still varying after controlling for student characteristics, Equation 4 was tested to estimate the fixed and random components for the main effect of the intervention as well as its interaction with student risk. Results from the conditional model showed that a small statistically significant effect was found for the interaction ($F = 4.05, p < .05$), suggesting that the treatments were differentially effective across students’ pretest level of risk.

Effect of Interventions for Students in the High-Risk Group

Results from the complete model revealed that, on average, all four interventions were equally effective although the fitted mean gain score for students in SOAR classrooms was larger (170.42) than gain scores for the other interventions groups (range = 123.92 – 160.32). Effect size differences showed that the percentage of nonoverlap between SOAR and READ 180 was 19.3%. This corresponded to an effect size difference of $adj. \Delta = -0.27$, which is small.

Comparing FCAT DSS gains for the high-risk ninth-grade students in the four intervention groups to the statewide average gain of 66 points revealed that intervention students’ FCAT performance gains were substantially greater. Students in the high risk group in all four interventions also demonstrated growth that exceeded the annual expected gain of 77 DSS points. Indeed,
students in the SOAR and RISE groups demonstrated improvements that were more than twice the magnitude of the state benchmark for expected annual growth.

Effect of Intervention for Students in the Moderate Risk Group

Comparisons across interventions within the moderate risk group (Level 2 on the FCAT) revealed that students in READ 180 classes made significantly greater gains compared to students in SOAR classes ($F = 6.31, p < .05, \text{adj. } \Delta = .30$). Similarly, students in RISE classrooms significantly outperformed students in SOAR classrooms ($F = 5.15, p < .05, \text{adj. } \Delta = .27$). A summary of the mixed model results is reported in Table 3. There was, however, no significant difference between the performance of students in the SOAR and the REACH classrooms. However, students in READ 180, RISE, and REACH classrooms showed greater average gains in FCAT DSS points relative to ninth-grade students statewide, whose average DSS gain was 66 points. Similarly, the average gain for students in this group for these three interventions exceeded the expected annual gain of 77 DSS points.

DISCUSSION

This study provides two important kinds of information about the instructional programs for struggling readers that were implemented across seven high schools. First, it provides a rigorous examination of the relative impact of the four interventions that were implemented in 90-min instructional sessions
throughout the year. The conclusions with regard to this question are relatively straightforward. For Level 1 (high-risk) students reading below the fourth-grade level, there were no statistically reliable differences in outcome across the four interventions. Of interest, the “control” intervention that focused a bit more directly on teaching skills specifically required for the FCAT (SOAR) produced the largest overall growth in performance from eighth to ninth grade. In light of the dip in performance often encountered during the first 2 to 3 years of new instructional approaches (Borman, Gamoran, & Bowden, 2008; Fullan, 1991), one explanation is that the extra year of experience SOAR teachers had with that curriculum led to higher quality implementation. Although summary data provided by the district indicated that fidelity and quality of implementation were relatively strong across all four conditions, more detailed information would have provided an opportunity to explore this and other assertions more fully. Although one of the experimental conditions (RISE) produced effects quite similar to SOAR, the effect size for the difference between SOAR and the least effective instructional condition for high-risk students (READ 180) was not found to be statistically reliable, primarily because of substantial variability in the treatment effect within all conditions.

In contrast to the lack of significant differences among intervention conditions for the high-risk students, among Level 2 (moderate-risk) students, those in the READ 180 and RISE conditions performed significantly better than students in the SOAR condition. The effect sizes for the READ 180 and RISE programs were .30 and .27, respectively.

It is interesting to note that the READ 180 program was associated with the smallest reading gains for the high-risk students and the largest gains for the moderate-risk students. Although the RISE program was relatively effective for both groups of students, the READ 180 program seemed much more attuned to the needs of the students reading above the fourth-grade level than it was to students with weaker reading skills. In contrast, the SOAR program was particularly effective with the weakest readers and least effective with the stronger readers. For READ 180 in particular, these results are consistent with descriptions of the program’s instructional focus provided by its developers. Although READ 180 does provide some instruction and practice support for the development of word-level skills, including reading accuracy and decoding, its primary focus is on building fluency and reading comprehension. This focus seems particularly well suited to the needs of Level 2 readers, who generally have well-developed basic reading skills but need additional instruction and support for the development of higher level comprehension skills.

The second purpose of the study was to provide some descriptive information about the size of the instructional impacts relative to state-level averages in growth on the FCAT that were obtained during the same year. Some aspects of this data are of particular interest to schools in Florida, whereas other aspects contribute to our general knowledge base about reading interventions for struggling readers in high school.
During the year of this study, ninth-grade students in Florida improved their DSS scores by 66 points. The average gain for the district’s ninth-grade population, 78 points, was slightly more than the expected FCAT-DSS gain score, and exceeded the state average by 10 points. In contrast, students in our study in the two strongest conditions made improvements of 160 and 171 DSS points, or more than twice the state average. Of course, it is not possible to attribute this difference solely to the interventions that were provided, as we also provided professional development for selected content area teachers to improve their instruction and support for the use of reading comprehension strategies. Further, there may have been other differences between schools in this particular district and the rest of the schools in Florida that affected the outcomes. However, given that the demographic characteristics of the students in this study are fully described, the growth of Level 1 students provides a useful “benchmark” against which other schools might evaluate the effectiveness of their own programs for seriously impaired readers in ninth grade.

The gains of students who performed at Level 2 on the FCAT in eighth grade in this study were also substantially above the state average for improvement of students in this category. Whereas the average gain of students across the state was 66 DSS points, students in the two most effective conditions in this study (READ 180 and RISE) improved by 105 and 102 DSS points, respectively. Although the difference between these latter scores and the state average is not as dramatic as that obtained for Level 1 students, it nevertheless provides a higher standard of achievement that is associated with clearly described instructional interventions.

It is also possible to draw implications from this study that contribute to the general knowledge base about instructional impacts for struggling readers in high school. The instructional conditions in this study were able to accelerate reading development substantially above both the average rate of growth across the state and the “expected” growth for adequate yearly progress as assessed by the FCAT. However, as a group, neither Level 1 nor Level 2 students made sufficient progress to move to a lower level of risk in ninth grade than they were in eighth grade. For example, in eighth grade, the average DSS score of students in the SOAR condition was 1559, which was 137 points away from the criterion for Level 2 performance. After the substantial acceleration in growth achieved by students in this group during ninth grade, their average score of 1729 placed them still 43 points below the criterion for designation as moderate risk in the Level 2 category. In fact, 56% of the eighth-grade Level 1 students in the most effective condition (SOAR) remained at Level 1 at the end of ninth grade.

A similar pattern emerged for the Level 2 Students. Those in the READ 180 group obtained an average DSS score of 1800 in eighth grade and an average score of 1905 in ninth grade. In eighth grade, these students fell 82 DSS points below the grade-level standard, and in ninth grade they fell 67 points below
the standard. Only 30% of the students in this group who scored at Level 2 in eighth grade scored at Level 3 or higher in ninth grade.

These results are consistent with the pattern that emerged in a recent report of a national study of supplemental reading interventions for ninth-grade students that was supported by the National Center for Education Evaluation and Regional Assistance (Kemple et al., 2008). This study, involving 2,916 students from 34 high schools in different regions of the United States, evaluated the impact of two instructional programs that were offered in supplemental classes for an average of 225 min a week for the school year. Although the pooled impact of the interventions on reading comprehension scores was statistically reliable when compared to a group that did not receive the interventions, it was quite small ($ES = .09$). On average, the students in the intervention groups began the study reading at the 16th percentile nationally and finished the study reading at the 25th percentile. Seventy-six percent of the students in the intervention groups were still reading more than 2 years below grade level at the conclusion of the study.

The findings from both of these studies indicate that most students who enter high school reading substantially below grade level will require more than 1 year of relatively intensive reading intervention to make significant progress toward the grade-level standard in reading. Both studies show that it is possible to accelerate reading development in high school struggling readers so that the gap between their skills and the grade-level standard is narrowed, but they both also suggest that the instructional conditions that accelerate reading growth may need to be extended across several years to make significant progress toward actually closing the gap.

These results are also consistent with a recent review of smaller scale studies that provided interventions to children with reading difficulties in late elementary and middle school (Torgesen, 2005b). The studies all demonstrated that increasing the intensity of high-quality instruction is able to substantially accelerate growth in students who have struggled in learning to read, but none of them was able to demonstrate the instructional conditions that are necessary (which likely involve high-quality, intensive instruction sustained over sufficient time) to actually bring struggling readers to grade-level performance in reading. This review also pointed out that most intervention studies with older readers have not employed outcome measures that are similar to the kind of formal, group-administered tests of reading comprehension that are currently used to measure the progress of schools and districts in teaching all students to read well. Rather, they have tended to use measures of reading comprehension that often involve reading much shorter passages than are typically found on reading accountability measures given by states and that are also given in a supportive, individually administered format.

This latter observation introduces one of the strengths of our study, in that the impacts of the interventions were studied in the realistic context of Florida’s accountability measure for reading. The FCAT is a challenging test of
reading comprehension (Schatschneider et al., 2004) that requires students in high school to read passages averaging 1,000 words in length and then to answer questions about the content of the passages that have a relatively high degree of cognitive complexity. The test is strongly correlated with nationally normed standardized tests of reading comprehension (Schatschneider et al., 2004), and the grade-level performance standard is reasonable in comparison to standards for basic reading proficiency established by the National Assessment Governing Board for the National Assessment of Educational Progress (National Center for Education Statistics, n.d.). It is important that we begin to establish the effectiveness of research-based instructional conditions by studying their impact on the measures that are being used to evaluate the progress of schools nationally to remediate reading problems in struggling readers.

Although the use of the FCAT to evaluate reading outcomes in this study can be considered a strength in one sense, we also recognize that it would have been useful to have an additional reading outcome such as a widely used standardized assessment of reading comprehension. We were not able to include such a measure because the funds available for the study (which came from the ongoing budgets of the organizations involved) were insufficient. Although this must remain a weakness of the current study, we would also argue that, in the context of current intervention research, one of this study’s strengths is that it provides unique insight into the impact of widely used interventions within the context of realistic state level accountability programs.

REFERENCES


Effectiveness of Reading Interventions


APPENDIX

Conceptual Logic Model

- Sample: Moderate and High Risk 9th graders
- Intervention: Participation in one of four interventions
- Proximal Outcomes: Improved instruction in reading
- Improved reading skills
- Distal Outcomes: Increased Reading Achievement
- Systematic evidence for others