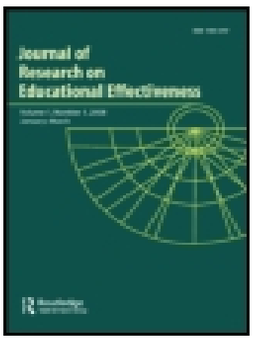

Routledge
Taylor & Francis Group

Journal of Research on Educational Effectiveness

ISSN: 1934-5747 (Print) 1934-5739 (Online) Journal homepage: http://www.tandfonline.com/loi/uree20

Examining the Effects and Feasibility of a Teacher-Implemented Tier 1 and Tier 2 Intervention in Word Reading, Fluency, and Comprehension

Emily J. Solari, Carolyn A. Denton, Yaacov Petscher & Christa Haring

To cite this article: Emily J. Solari, Carolyn A. Denton, Yaacov Petscher & Christa Haring (2017): Examining the Effects and Feasibility of a Teacher-Implemented Tier 1 and Tier 2 Intervention in Word Reading, Fluency, and Comprehension, Journal of Research on Educational Effectiveness, DOI: 10.1080/19345747.2017.1375582

To link to this article: https://doi.org/10.1080/19345747.2017.1375582

JOURNAL OF RESEARCH ON EDUCATIONAL EFFECTIVENESS
2017, VOL. 0, NO. 0, 1–29
https://doi.org/10.1080/19345747.2017.1375582

Examining the Effects and Feasibility of a Teacher-Implemented Tier 1 and Tier 2 Intervention in Word Reading, Fluency, and Comprehension

Emily J. Solari[a], Carolyn A. Denton[b], Yaacov Petscher[c], and Christa Haring[d]

ABSTRACT

This study investigates the effects and feasibility of an intervention for first-grade students at risk for reading difficulties or disabilities (RD). The intervention was provided by general education classroom teachers and consisted of 15 min whole-class comprehension lessons (Tier 1) and 30 min Tier 2 intervention sessions in word reading, comprehension, and text reading. First-grade teachers ($n = 21$), with 4–5 students at risk for reading difficulties and potential reading disability were randomly assigned to treatment or typical practice comparison conditions. Significant group differences were detected on all measures of word reading, decoding, and fluency. Effect sizes were educationally important for all measures of word reading, decoding, and reading comprehension; however, effects on standardized measures were smaller than those in prior studies with similar students in which intervention was typically provided outside the regular classroom. An exploratory analysis indicated that students at different parts of the pretest and posttest distributions responded more and less positively to the intervention, providing insights that may help guide future revisions. The study provides preliminary evidence of the intervention's promise for positively impacting student outcomes.

KEYWORDS

elementary
intervention
reading

First-grade students who demonstrate difficulties with early reading acquisition are likely to have continuing reading difficulties, and are at higher risk of being identified as reading disabled (RD) as they progress through school (Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996; Torgesen & Burgess, 1998). Later-developing reading problems may be reduced if the gap between typically developing readers and students at risk for reading problems is addressed aggressively in the early grades (Torgesen, 2004; Vellutino Scanlon, Small, & Fanuele, 2006). Based on evidence of the general effectiveness of early reading interventions (Elbaum, Vaughn, Hughes, & Moody, 2000; Torgesen, 2004; Wanzek & Vaughn, 2007), some schools implement Response to Intervention models for the prevention of RD in the early grades (Berkeley, Bender, Peaster, & Saunders, 2009; Spectrum K12 School

CONTACT Emily J. Solari Ejsolari@ucdavis.edu School of Education, University of California, Davis, 1 Shields Ave., Davis, CA 95616, USA.
[a]Reading and Academic Development Center, M.I.N.D. Institute, University of California, Davis, California, USA
[b]Children's Learning Institute, University of Texas Health Science Center Houston, Houston, Texas, USA
[c]Florida Center for Reading Research, Florida State University, Tallahassee, Florida, USA
[d]Developmental Pediatrics, University of South Florida, Tampa, Florida, USA
Supplemental data for this article can be accessed on the publisher's website.

Solutions, 2009); these models are also referred to as Multi-Tier Systems of Support (MTSS). MTSS models are comprehensive frameworks providing increasingly intensive academic and/or behavioral support according to documented student needs. Preventative MTSS models typically consist of three tiers of intervention, with each tier increasing the level of instructional support (Gersten et al., 2009). In an MTSS framework, all students receive evidence-based classroom reading instruction, or Tier 1. When Tier 1 is deemed insufficient for an individual reader, Tier 2 provides additional support, usually in the form of small-group intervention aimed at ensuring that students at risk for RD meet grade-level reading benchmarks. Students with inadequate response to sufficiently intensive, evidence-based Tier 2 interventions receive more individualized and intensive Tier 3 interventions.

Many studies of MTSS have been conducted in the early grades in an attempt to prevent reading difficulties with interventions that address early word reading (e.g., McMaster, Fuchs, Fuchs, & Compton, 2005; Vellutino et al.; Wanzek & Vaughn, 2006); all of these studies have advanced the understanding of early reading problems and prevention models. However, very few studies have examined early preventative models of reading intervention that address early risk in both word reading abilities and comprehension skills. Additionally, very few previous studies have investigated the effectiveness and feasibility of Tier 2 supplemental instruction implemented by regular classroom teachers in general education classrooms. The current study attempts to address both of these gaps in the existing literature.

Decoding and Comprehension: Two Important Skills in Early Reading Intervention

The effectiveness of explicit and systematic early word reading interventions has been well-documented (e.g., Ehri, Nunes, Stahl, & Willows, 2001; Wanzek & Vaughn, 2007), but considerably less attention has been paid to systematic early intervention in comprehension in the primary grades, specifically early comprehension development through listening comprehension instruction. There is evidence that many students are impaired in both word reading and comprehension (Catts, Hogan, & Fey, 2003; Leach, Scarborough, & Rescorla, 2003) and that some are impaired specifically in reading comprehension but not word reading (Catts, Compton, Tomblin, & Bridges, 2012). There is also empirical evidence to suggest that adequate word reading is not, by itself, sufficient for successful reading comprehension (Aarnoutse, Leeuwe, Voeten, & Oud, 2001). Consequently, students who have difficulties with comprehension early are at a significant disadvantage. Empirical data suggests that there is a relation between early listening comprehension and later reading comprehension; previous studies have indicated that listening comprehension abilities may be an important early precursor to later reading comprehension (Carlisle & Felbinger, 1991; Catts, Hogan, & Fey, 2003; Duke, Pressley, & Hilden, 2004) and that early listening comprehension predicts later reading comprehension outcomes (Aarnoutse, van den Bos, & Brand-Gruwel, 1998; Garner & Bochna, 2004; Nation & Snowling, 2004). Additionally, studies have suggested a significant relation between listening comprehension and decoding ability in populations of children who have reading disabilities (Wise, Sevcil, Morris, Lovett, & Wolf, 2007) and those who are typically developing (Nation & Snowling, 2004).

From a theoretical standpoint, if empirical data demonstrates a significant relation between early listening comprehension and both later reading comprehension and

decoding, it would make sense that explicit comprehension instruction be included in early reading programs and interventions even before children are skilled readers. Arguably, this is more important of early readers who demonstrate risk on one or both of these areas of reading development. Shanahan et al. (2010), in a guide sponsored by the Institute of Education Sciences' What Works Clearinghouse, suggested that "students who read with understanding at an early age gain access to a broader range of texts, knowledge, and educational opportunities, making early reading comprehension instruction particularly critical" (p. 5). The development of accurate and fluent word reading has a critical role in reading comprehension; however, skilled text comprehension also depends on the development of other skills. The practice guide recommends instruction in oral language, vocabulary knowledge, broad world knowledge, an understanding of text structures, and a repertoire of comprehension strategies that can be applied to understand and remember ideas in text (Shanahan et al., 2010). Shanahan et al.'s comprehensive review of research provided strong support for the provision of explicit comprehension instruction in kindergarten through third grade. Because many young readers, especially those who are struggling with developing reading-related skills, do not fluently decode in the early grades, it is reasonable to assume that the bulk of early comprehension skill development would occur through explicit listening comprehension instruction.

Supplemental Small-Group Interventions That Target Decoding and Comprehension For At-Risk Readers

Few studies have investigated the provision of small-group supplemental interventions emphasizing simultaneous explicit, systematic instruction in both word study and comprehension in the early primary grades. The small number of studies that have been conducted show promise for the approach. Allor and McCathren (2004) tested a supplemental first-grade reading intervention with at-risk students who were randomly assigned to receive the intervention or assigned to the contrast condition; intervention sessions were conducted by volunteers and covered phonological awareness (PA), letter-sound correspondence, word study, and simple comprehension questions. Significant differences were seen between treatment and control groups in phonemic awareness and measures of decoding; no significant differences were seen on measures of comprehension. Similarly, Baker, Gersten, and Keating (2000) tested the effects of a volunteer-implemented reading program with first- and second-grade at-risk readers. As seen in the previous study, significant differences in word-level measures favor the students who were randomly assigned to the reading program; however, no differences were seen between groups on a measure of passage comprehension. Berninger et al. (2003) provided at-risk second-grade readers with interventions addressing word recognition only, reading comprehension only, or a combination of the two. All conditions were compared to a control group that engaged in text-reading practice with no instruction. Comprehension intervention included directed discussion guided by teacher questions as well as explicit instruction. In the directed discussions, questions prompted students to apply a variety of reading skills and strategies, and the explicit instruction segment was directed at literal comprehension. Students who received the combined comprehension and word study intervention performed significantly better on one

measure of phonological decoding relative to the group that received word study intervention alone and to controls, but no other significant differences were found. Although comprehension was taught in each of these studies, it appears that the main concentration was on word study; comprehension was not systematic and explicit, nor was there a clear scope and sequence. Given this, it is not surprising that there were not discernable effects on comprehension.

There are a handful of studies that have seen positive effects on measures of comprehension after explicit comprehension instruction in early primary school years. Solari and Gerber (2008) compared the effects of two interventions for kindergarten English learners (ELs); one included instruction in PA, letter knowledge, and listening comprehension, while the second included only PA and letter knowledge. Both at-risk and typically developing EL readers who received the combined intervention had significantly better outcomes on story retell, listening comprehension, and PA than students who received only word-level instruction. In this study, the comprehension instruction was systematic and explicit, with comprehension skills development following a scope and sequence, from easier comprehension skills to more difficult. Additionally, the instruction covered many of the instructional elements in the *Shanahan 2010 IES Practice Guide*, including setting a purpose for reading and explicit instruction in comprehension strategies. Denton, Solari, Ciancio, Hecht, and Swank (2010) implemented a small-group instruction reading program in comprehension and word reading in a summer school program for students about to enter first grade, reporting significant positive effects on word reading and listening comprehension. Similarly, Schacter and Jo (2005) reported positive effects on decoding and comprehension outcomes for a kindergarten and first-grade summer school program that addressed both word reading and comprehension-related skills.

Classroom Teacher-Delivered Tier 2 Reading Intervention

In efficacy studies of supplemental Tier 2 interventions for at-risk readers, the treatment is more often implemented by the researchers themselves, graduate students, or a small number of teachers or paraprofessionals trained by the researchers (e.g., Allor & McCathren, 2004; Baker et al., 2000; Ball & Blachman, 1991; Berninger et al., 2003; Mathes et al., 2005; Solari & Gerber, 2008). Although there have been studies that employ regular classroom teachers as the interventionists, these have been limited to studies that have asked classroom teachers to provide instruction in studies of whole-class, or Tier 1, instructional approaches (e.g., Williams, 2005) and in studies of peer-mediated interventions (e.g., Mathes, Grek, Howard, Babyak, & Allen, 1999). Few of the peer-tutoring interventions implemented by classroom teachers have examined teacher-directed Tier 2 reading interventions within their own classrooms. In one such study, Fuchs et al. (2001) examined a PA intervention delivered by kindergarten teachers, comparing its effects to those of the same intervention combined with peer tutoring and a control group; the students in the two treatment groups showed greater improvement in PA. The PA intervention was designed for delivery 1–3 days per week in 5–15 min sessions for 20 weeks. Teachers generally implemented the intervention on this schedule, and they demonstrated adequate fidelity; however, the time demands of this intervention were fairly minimal.

From a resources standpoint, it can be argued that the most efficient model of Tier 2 supplemental reading instruction would occur in classroom settings, with the general education teacher implementing supplemental instruction for the most at-risk readers. This is because in typical MTSS models, up to 25% of a classroom is eligible for Tier 2 supplemental reading services; this accounts for a large number of students, too many for many schools to provide these services through specialist reading teachers or special education teachers. However, very few studies have empirically tested a model in which Tier 2 instruction is implemented in general education settings with the regular classroom teachers in providing the supplemental instruction. To our knowledge, the only teacher-implemented, Tier 2 reading intervention that explicitly taught both decoding (word-level) and comprehension skills with first-grade students is the aforementioned Denton et al. (2010) study. Though the results of this study were promising, the intervention period was brief—it only occurred over the course of a short summer school session. Additionally, the teachers in the study did not have the competing demands on instructional time as is typically seen during the regular school year; instead, implementing the reading program was their main focus during the summer school period. Therefore it is difficult to determine whether classroom teacher-implemented Tier 2 instruction is feasible in academic year classroom settings when teachers have competing instructional and student demands.

Purpose of the Current Study and Research Questions

Despite evidence suggesting the importance of early comprehension instruction, few studies have evaluated reading interventions that address both word reading and comprehension in first grade, and there are few programs designed for the provision of systematic, explicit instruction in comprehension as well as word reading to primary-grade students who need supplemental small-group Tier 2 support. There are very few studies that investigate the effectiveness and feasibility of teacher-delivered Tier 2 supplemental reading instruction. Furthermore, the majority of previous intervention studies utilize analytical approaches that treat at-risk students as homogenous groups of students. It is possible that more comprehensive findings that have potential to better inform targeted intervention development and implementation could result from studies that employ analytical approaches that investigated subgroups of at-risk readers. To build on the current literature base and address these gaps in the literature, this study implemented a comprehensive, Tier 2 reading intervention, implemented by classroom teachers. We conducted a series of data analyses that investigate student responsiveness to intervention, differential response based on reading proficiency and pretest and posttest, and an examination of the feasibility of implementation. Specifically, we addressed the following three research questions:

(1) Do first-grade students at risk for reading difficulties who receive the experimental intervention perform significantly better than similar students who receive typical reading instruction on measures of phonemic awareness, phonemic decoding, word reading, reading fluency, listening comprehension, and reading comprehension?
(2) To what extent do students who receive the experimental intervention outperform the comparison group on the outcome measures depending on: (a) their initial baseline performance or (b) varying points of the outcome distribution?

(3) Are first-grade classroom teachers able to implement the intervention with adequate fidelity (i.e., 80% or higher)? Are teachers able to implement the intervention regularly on the prescribed schedule?

Method

Schools and Context

This study was conducted in eight elementary schools in a large urban center in the Southwest. Students in the schools were primarily Hispanic (65%) and African American (27%). On average, 88% of students attending these schools were economically disadvantaged (i.e., qualified for the federal free or reduced-price lunch program; range = 67%–98%). The study was approved by the university's institutional review board and the school district's research department.

Participants and Design

Study participants were 21 first-grade teachers and four to five students from each of their classrooms ($n = 98$ students). Once district- and school-level agreements to participate were secured, the research team met with school principals and first-grade teachers to describe the study. Teachers who agreed to participate in the study signed letters of informed consent; these letters indicated that teachers would be randomly assigned to participate in the curriculum intervention or serve as controls. Parental informed consent and verbal student assent were obtained for all student participants. The identification of eligible teachers and students was a multistep process. After initial recruitment, students were screened for eligibility (process described below). Teachers were eligible for the study if they had at least four students who met the requirements for risk status; every classroom met this criteria. After students were screened for risk status and classroom eligibility was determined, teachers and the students in their classrooms were assigned to condition.

At-Risk Student Identification Process

Students were identified as at risk for RD using teacher nomination in combination with screening measures of word reading and listening comprehension. First, teachers nominated the eight students in their classes that they identified as the most impaired readers, just over one third of the typical class of 21 or 22 students; these students were subsequently screened to identify the most at-risk students in each classroom. Research staff gave the nominated students a brief word-reading screen comprising simple words that students are typically exposed to by the beginning of first grade; this measure has been used successfully to identify at-risk first graders in previous research (Denton et al., 2010; Mathes et al., 2005); if students read less than four words correctly, they were deemed at risk for decoding. Next, to identify students with impaired comprehension, we used the scores from the comprehension subtest of the Texas Primary Reading Inventory (TPRI; Children's Learning Institute & Texas Institute for Measurement, Evaluation, and Statistics, 2010), which had been administered by the classroom teachers as part of normal district procedures. The TPRI is designed to identify students in Grades K–3 at risk for reading difficulties and as a diagnostic instrument to assist in planning instruction. In the comprehension task, students answer literal and inferential

questions about text. In first grade, this test can be administered as either a reading comprehension or a listening comprehension task (i.e., the text is read aloud to the student). All students in this study received the listening comprehension administration. If students scored in the at-risk range on the listening comprehension subtest, they qualified for the study. Using scores on the two screens, we selected four to five students in each classroom who were at risk for RD. Three identified students were removed by school personnel prior to the beginning of intervention—two due to severe behavior problems and one because the student was being evaluated for special education; these students were replaced by the students in these classrooms with the next-lowest screening scores. In addition, one student who was originally selected was moved by the school into a nonstudy classroom prior to the beginning of the intervention, resulting in our study sample of 98 students.

Teacher Assignment and Sample Characteristics

Because our original intent was to conduct an exploratory investigation of the effects of coaching on RR outcomes, after student screening and classroom eligibility was determined, we randomly assigned teachers, with clusters of four to five of their students, to one of three conditions: an RR group that received coaching (RR-C group; $n = 6$ teachers), an uncoached RR group (RR-UC group; $n = 7$ teachers), or a comparison group in which students received their typical school-provided reading instruction and intervention but did not receive RR (Typical Reading Instruction, or TRI Group; $n = 8$ teachers). There were no significant differences on any baseline or outcome measure between the coached and uncoached RR groups; therefore, the two RR groups were combined into a single treatment group for the current study (RR group).

The sample for the current study consisted of 13 teachers and 61 of their students who were randomly assigned to the RR group and 8 teachers and 37 of their students who were randomly assigned to the TRI group. In both conditions, 43% of the students were female. RR students were 31% African American, 60% Hispanic, 4% White, and 1% other ethnicities, while students in the TRI group were 37% African American, 48% Hispanic, 5% White, and 3% other ethnicities. Ninety percent of RR students and 100% of TRI students participated in the federal free or reduced-price lunch program. Percentages of students with limited English proficiency were 32% for the RR group and 21% for the TRE group. In the RR group, 4% of students were served by special education, primarily for speech and language disorders; in the TRE group this proportion was 11%. Eight students withdrew during the intervention period, six in the RR group and two in the TRI group. All moved away from their schools except one RR student, who was removed by the teacher prior to the onset of intervention because she did not feel the child was appropriate for the study. Thus, 90 students completed the study, 55 in the RR group and 35 in the TRI group.

All of the 21 teacher participants were female. Ten were White (four in the RR group), nine were African American (five in the RR group), one was Asian (in the TRE group), and one was Hispanic (in the RR group). One teacher had a doctoral degree (TRE group), and seven had master's degrees (three in the RR group), while 13 had Bachelor's degrees (six in the RR group). All had state teaching credentials in elementary or early childhood education. Two were also certified as reading specialists (one in the RR group), and nine were certified to teach English as a second language (five in the RR group). The average number of years of teaching experience of RR and TRI teachers was 10.5 and 7 years, respectively.

Description of the Intervention

The RR intervention included systematic instruction in both word reading and comprehension. Instruction took place in both Tier 1 and Tier 2 settings; both Tier 1 and Tier 2 instruction is supplemental to the basal English language arts curriculum. Tier 1 comprehension instruction was delivered to the entire class in 15-minute lessons; the purpose of this brief Tier 1 instruction is to introduce students to comprehension instruction through a whole-group read-aloud. Students at risk for RD received an additional 30-minute Tier 2 supplemental small-group lesson that consisted of 10 min of additional comprehension instruction, 10 min of word study instruction, and 10 min of connected text reading, including fluency instruction. Both the Tier 1 and Tier 2 components were provided four days per week for 17 weeks.

The intervention incorporated instructional design principles based on characteristics of effective instruction for students with RD. These included the provision of explicit instruction (Swanson, 2000; Swanson & Hoskyn, 1998; Torgesen, 2004) with objectives sequenced and systematically addressed (Coyne, Kame'enui, Simmons, & Harn, 2004; Swanson, 2000). The RR Teacher Guides were designed to be sufficiently standardized to support implementation with high fidelity and decrease demands on teacher time, thereby increasing feasibility. When new skills or activities were introduced, instruction was highly scripted, but the level of scripting decreased as teachers and students became familiar with the instructional routines that were repeated throughout the program. RR was also designed to promote (a) active student engagement with specific teacher feedback and instructional scaffolding (Vaughn, Gersten, & Chard, 2000), (b) adaptation of instruction based on student data (Stecker, Fuchs, & Fuchs, 2005), and (c) extended practice, including application of skills and strategies in connected text (Snow, Burns, & Griffin, 1998). A distinguishing feature of the program was the inclusion of specific procedures for instructional scaffolding for each comprehension and word study activity. Teachers also received professional development in providing scaffolding and feedback during oral reading to promote reading accuracy and fluency, self-monitoring, and self-correction of errors.

Comprehension Instruction

The design of the RR comprehension lessons was guided by research indicating that students with reading difficulties benefit when the difficulty of comprehension tasks is controlled (Gersten, Fuchs, Williams, & Baker, 2001) and when direct instruction is provided (Gajria, Jitendra, Sood, & Sacks, 2007). Each comprehension strategy was taught and practiced within an integrated unit of study for a minimum of three weeks. The intervention progressed systematically from easier to more challenging strategies. The unit topics were (a) activating background knowledge and making personal connections with text; (b) direct recall of text information and literal comprehension, (c) recognition of narrative text elements and story retelling, (d) question generation, (e) making inferences, and (f) identifying the topic and main idea of a text segment.

Each comprehension unit began with one or more Tier 1 anchor lesson(s) in which the teacher provided explicit instruction in the new strategy and modeled its application through a "think-aloud," and students received guided practice in its application. Students were taught hand signals that represented each comprehension strategy and used the hand signals as they practiced the strategies. The anchor lessons were followed by read-aloud lessons in

which teachers introduced a problem or guiding question that was strategically planned to guide students toward comprehension of important themes or elements. While reading the story aloud, the teacher stopped at predetermined points to provide explanations of vocabulary words and ask questions designed to lead students toward answering the primary guiding question. After reading, teachers and students engaged in text-based discussions related to the guiding question. In Tier 2 comprehension instruction, students engaged in activities targeting the same objectives as in the Tier 1 lessons, but designed to provide extended practice with increased opportunities for students to respond with teacher feedback. Small-group instruction frequently included the use of manipulatives and other strategies designed to promote active student involvement.

Word Study Instruction

Word study instruction was provided only in the Tier 2 intervention. Instructional activities in the word study component followed five strands: PA; letter-sound instruction; sound analysis and spelling; phonemic decoding; and high-frequency word-reading. Within each strand, activities progressed systematically from easier to more challenging objectives. Students learned to spell words with each sound-spelling pattern as well as read them. Instructional activities provided direct modeling and explanation, guided practice with teacher feedback, and independent practice, including extended practice in multiple formats to develop firm and fluent application of skills. Word study instruction promoted active engagement by using instructional materials such as letter tiles, magnetic letters, individual whiteboards, Elkonin sound boxes, and other manipulatives.

Each unit in the word study program introduced one sound-spelling pattern and one or more new high-frequency words, and teachers were provided with three to four lesson plans for each unit. One of these lessons was designed to be used only for reteaching if students did not master the unit objectives, as measured by mastery tests. The tests were designed to be administered by teachers for initial placement and after every two units of instruction. Guidelines were provided to assist teachers in the use of mastery test data to make decisions to (a) reteach previous lessons, (b) continue to the next lesson but integrate extra practice in items students had missed, or (c) continue to the next lesson. Mastery test data could also be used to determine whether some students were progressing rapidly and could skip some lessons, thereby accelerating progress through the program.

Text Reading

During the 10-min text reading segment, students applied both reading comprehension and word reading skills and strategies with teacher modeling, scaffolding, and feedback. The teacher provided a very brief book introduction to focus attention on a guiding question related to the day's comprehension lesson. The guiding question was revisited partway through the book and after reading. Students were taught only one three-part strategy for identification of unknown words: (a) look for parts you know (i.e., check for familiar sound-spelling patterns, word endings, or other familiar patterns); (b) sound the word out; and (c) check it (i.e., put the newly decoded word into the sentence and read to be sure it makes sense). Nondecodable words that were not recognized by students were provided by the teacher. Students were also taught through teacher modeling and prompting to attend closely to the meaning of the text and to monitor for and self-correct decoding errors that impeded meaning.

Student text, which was written by a children's author to follow the RR decoding scope and sequence, was fully decodable using previously taught letter-sounds, sound-spelling patterns, and high-frequency words. Unlike much decodable text, the RR student books included engaging content, beautiful illustrations, and recognizable plot structures, even at the beginning levels. Students encountered increasingly sophisticated vocabulary and language as they progressed through the series. Student texts included narrative and informational books, as well as poetry.

To promote text reading fluency, teachers used a systematic progression of instructional strategies. In the earliest lessons, teachers provided examples and nonexamples of fluent reading. Later, they explicitly taught the meaning of punctuation marks, modeled their interpretation, and guided students to apply them to read expressively. In addition, teachers modeled phrased reading and provided practice in reading in phrases rather than word-by-word. Students also practiced fluent reading in familiar text with partners.

Intervention Implementation

Teachers were asked to implement the RR intervention four days per week over the course of 17 weeks, from October through March. The entire intervention was delivered by classroom teachers within their regular classrooms. To our knowledge, they were not assisted by paraprofessionals or other persons. When teachers were working with their small Tier 2 groups, other students in the classrooms either wrote in response to writing prompts related to the whole-group comprehension lessons or engaged in teacher-designed small-group center activities or other independent work.

Professional Development

RR teachers received two days of professional development prior to implementation and two additional days during the course of the study. Topics included: (a) overview of tiered, data-driven intervention; (b) evidence-based instruction for students with reading difficulties; (c) implementation of the RR lesson components with role-played practice; (d) mastery test administration, scoring, and use of data to make instructional modifications, and (e) managing small-group instruction in the classroom.

School-Provided Reading Instruction

We were unable to directly observe the supplemental reading instruction provided to the TRI group; however, at the end of the study we conducted structured interviews of participating teachers to (a) document the amount of classroom reading instruction provided to participating students and (b) document any additional reading tutoring or intervention provided outside of our study. Data were collected for 96 of the 98 participating students.

Classroom Reading Instruction

Teachers in both the RR and TRI groups reported that the majority of participating students received 90 to 120 min of regular classroom reading instruction each day, while about 14% of the students received more than 120 min of daily reading instruction. The majority of students received instruction that primarily utilized the district-adopted core reading program (59% RR, 76% TRI), while about 20% of the students in each group received guided reading lessons as their primary reading instruction. One student in each group received all daily

reading instruction in a resource room setting, and about 12% of students in each group received some instruction from a reading interventionist during their regular classroom reading instruction time.

Additional Reading Instruction

Teachers also reported that about 60% of the students in each randomized group received some kind of school-provided reading intervention or tutoring in addition to regular classroom instruction, besides the RR instruction provided through the study. This supplemental instruction ranged from occasional tutoring to reading with community volunteers to specialized interventions delivered outside the classroom. Some students received this instruction from their regular classroom teachers (27% RR, 35% TRI), while others received instruction from a reading specialist (25% RR, 19% TRI) and still others received tutoring from community volunteers or other certified teachers. Teachers estimated the number of minutes per week students received additional school-provided reading intervention as an average of 28.41 min ($SD = 29.15$) for RR group students and 21.58 min ($SD = 25.44$) for TRI group students.

Measures

A pretest-posttest battery measured phonemic awareness, word recognition, phonological decoding, text reading fluency, listening comprehension, and reading comprehension. Videotapes of classroom instruction were also collected. Assessments were administered by testers who were experienced in data collection and trained to criteria by a research assistant whose primary focus was data collection. All student measures were administered in quiet locations in students' home schools outside of their classrooms.

Phonemic Awareness

The Yopp-Singer Test of Phoneme Segmentation (Yopp-Singer; Yopp, 1995) was used to measure phonemic awareness. In this task, students listen to a word presented orally by the administrator and then pronounce each of the phonemes in the word. The raw score, which was analyzed for this study, is the number of words segmented correctly. Yopp (1988) reported internal consistency reliability for the measure of .95.

Word Reading and Phonological Decoding

Two subtests of the Woodcock-Johnson Tests of Achievement (WJ III; Woodcock, McGrew, & Mather, 2001) were administered to measure word reading and decoding. The Letter-Word Identification subtest assesses the ability to identify letters and read words presented in a list format. The Word Attack subtest measures phonological decoding through pseudoword reading. Both measures are untimed. Split half reliability is .98–.99 for Letter-Word Identification and .87 for Word Attack for the age of interest. WJ III *W* scores were utilized in data analyses. The *W* scale is an equal interval scale that represents both an individual's ability and item difficulty, allowing accurate description of an individual's growth in a measured trait and comparison of proficiency for individuals across multiple ages on a single continuous scale (Jaffe, 2009).

Two researcher-developed measures of word reading fluency were utilized as proximal measures of intervention effects. In Sight Word Fluency students read a list of

high-frequency nondecodable words taught in the research intervention and commonly taught in first-grade classrooms. In Decodable Word Fluency students read a list of three-letter decodable words composed of letter-sounds taught in the experimental program. In both measures, all students were presented with all items, and time to completion was recorded. The score was the number of words read correctly per minute. Test-retest correlations calculated for students in an untreated comparison group in a previous pilot study were .80 for Sight Word Fluency and .69 for Decodable Word Fluency.

Reading Fluency

The DIBELS Oral Reading Fluency (DORF) subtest (Good & Kaminski, 2002) was administered as a measure text reading fluency. DORF is an individually administered assessment of fluent and accurate reading of connected text. At pretest and posttest, students read two first-grade passages for one minute each. The score is the number of words read correctly per minute. The score utilized in the analyses was the average of the scores for the two passages. The National Center on Intensive Intervention (2015) reviewed technical adequacy data for the DORF from multiple sources, reporting alternate-form reliability estimates ranging from .89 to .96 for DORF at Grade 2, and citing several studies that reported reliability estimates exceeding .90 for DORF at Grade 1. To control for form effects, all students in the current study read the same two passages at pretest and at posttest. This decision was based on the fact that students could only read a few words of these passages at pretest; moreover, any effects from rereading the same passages would be presumed to affect the randomized groups equally.

A proximal researcher-developed measure of Decodable Sentence Fluency in which students read sentences composed of common high-frequency words and decodable words was administered. Students read for two minutes, and the score was the number of words read correctly in two minutes. We measured concurrent validity for this measure using pretest scores in the present study, finding a correlation of 0.94 with DIBELS ORF.

Comprehension

Two subtests of the WJ III to measure comprehension—Oral Comprehension and Passage Comprehension—were collected. The two tests have parallel formats in which students provide missing words to complete brief sentences or passages. In Oral Comprehension, these items are read to the student, while the student reads independently in the Passage Comprehension test. Split half reliability is .85 for Oral Comprehension and .89 for Passage Comprehension. For data analyses, *W* scores were utilized.

Implementation Fidelity

Observational data documenting teachers' fidelity and quality of implementation of curricular components was collected. For most teachers, data were collected at three to four points over the course of the intervention; one teacher had data for only two time points due to scheduling conflicts. Research personnel videotaped RR group teachers as they implemented the intervention. Tapes were later viewed and coded by two individuals who were familiar with reading instruction but had not been associated with the study. These coders were trained using videotapes collected the previous year and practiced coding lessons on these tapes until they reached inter-rater reliability of 91% based on absolute agreement.

Each instructional activity in the taped lessons was coded on a three-point Likert-type scale for (a) oral presentation, (b) implementation of instructional activities as prescribed, (c) providing appropriate scaffolding and feedback, and (d) accuracy of student responses. Teachers were also rated on a three-point scale for each segment of the intervention (i.e., large-group comprehension, small-group comprehension, word study, text reading) for pacing/use of instructional time, being prepared and organized, and whether students were exhibiting engaged and on-task behavior. Global ratings were assigned for the quality of instruction and behavior management for the entire lesson.

Data on dosage of the intervention received by each student and the number of RR lessons taught by each teacher were collected using weekly teacher logs, documenting the comprehension and word study units completed and student attendance.

Analytic Approach

The primary impact analyses were addressed by a series of mixed model analyses to estimate the extent to which the treatment resulted in significant effects across the nine measures. Although a more parsimonious approach might be to leverage the common variances across measures in order to create latent composites, (a) the iterative design nature of the treatment was such that individual analyses might yield different effects according to the measures, (b) the classrooms were unbalanced in size between treatment and comparison, and (c) the number of classrooms (i.e., 26) was fewer than that typically desired for a multilevel structural equation model (Muthén & Muthén, 2002).

The structure of the data reflected a three-level model as students were nested within classrooms and schools. In order to provide symmetry in the statistical models between the primary impact and the exploratory models (discussed below), and due to the primary random assignment and treatment delivery occurring at the classroom level, we opted to run a two-level model with a sandwich estimator (White, 1980) used to cluster-correct the standard errors for school-level effects.[1] An unconditional model (i.e., Model 1) was generated in order to decompose the variances in posttest scores due to between-student and between-classroom differences. Model 2 tested the impact of the intervention on the nine posttest scores between RR and TRI groups (i.e., a dummy-code covariate indicating treatment status at the classroom level), controlling for student pretest scores. Statistically significant treatment effects across the outcomes were subjected to a linear step-up correction in order to guard against the false-discovery rate (Benjamini & Hochberg, 1995). All models were estimated using PROC MIXED in SAS 9.3.

Hedges's *g* effect sizes were computed for all outcomes. The magnitude of these effects and their substantive importance were evaluated using two types of indicators. First, following the recommendations of the What Works Clearinghouse (2014), effect sizes were calculated for all measures regardless of statistical significance, and any effect larger than .25 was considered to be substantively important, particularly for students at risk for reading difficulties. Second, following recommendations in Hill, Bloom, Black, and Lipsey (2008) magnitude of the effects was evaluated on standardized measures by comparing them to empirical

[1] A three-level unconditional model was estimated to evaluate the school-level intraclass correlations for Word Attack (.25), Letter-Word Identification (.25), Sight Word Fluency (.20), Yopp-Singer (.05), Decodable Word Fluency (.16), Decodable Sentence Fluency (.15), Oral Comprehension (.06), Oral Reading Fluency (.09), and Passage Comprehension (.15).

benchmarks derived from other studies in which reading intervention was provided to at-risk first-grade students. These benchmarks are, as suggested by Hill et al., "relevant to the intervention, target population, and outcome measure being considered," (p. 172). For comparison, we examined the synthesis of early reading intervention studies by Wanzek and Vaughn (2007) and identified the first-grade studies in which WJ III Letter-Word Identification, Word Attack, and Passage Comprehension were administered. We calculated the mean effect sizes for these measures as reported by Wanzek and Vaughn, along with the mean effect sizes reported in a large-scale study of small-group reading intervention previously conducted (Denton et al., 2010) The average effect sizes for these studies were .69 for WJ III Word Attack, .68 for Letter-Word Identification, and .62 for Passage Comprehension.

In order to take a fine-grained look at our outcomes to evaluate the effects of the intervention for different subgroups of students, our secondary exploratory analyses investigated the extent to which the impact of the intervention may have been differentially effective conditional on either pretest or posttest score. Tests of moderation were estimated with interactions between the pretest and treatment variables. In evaluating interaction terms, which frequently require larger sample sizes than main effects to detect statistically significant effects, p values of $<.10$ were explored via simple slopes analysis (Preacher, Curran, & Bauer, 2006). By using the dichotomous variable indicating treatment status (treatment = 1, comparison = 0) as well as selected cut-points on the pretest variable (i.e., ± 1 standard deviation of the sample mean), it is possible to identify regions of significance for where the treatment and control groups may differ in estimated posttest scores conditional on different levels of the pretest.

The second set of exploratory analyses entailed the use of linear quantile mixed models. Unlike fixed and mixed effect models that estimate regression coefficients conditional on the mean of the posttest distribution, quantile models estimate regression coefficients conditional on many points of the posttest distribution (Koenker & Bassett, 1978; Petscher, 2016). A quantile may be viewed as conceptually similar to a percentile or fractile, and quantile regression may be viewed as a special case of median regression (Koenker & Bassett Jr, 1978). Whereas traditional mixed effects models that are based on the conditional mean produce coefficients that reflect the average relation between a set of covariates and a selected outcome, this approach may be potentially limited due to the lack of strict normality in the measured variables or circumstances where one has a hypothesis that the relationship between the independent and dependent variables differs for students at one end of the distribution compared to another. In such instances, quantile regression is a useful analysis to estimate the conditional relationship between the independent and dependent variables at selected points of the outcome distribution due to its lack of assumptions about the shape of the distribution of the outcome (Koenker & Bassett, 1978), as well as its ability to estimate individual conditional effects.

A useful mechanism for understanding the similarities and differences between the traditional conditional means mixed model and a quantile regression approach may be viewed through a formulaic expression of each. A sample mixed model for testing primary impacts in a two-level model is

$$Y_{ij} = \gamma_{00} + \gamma_{10}(\text{Pretest}_{ij}) + \gamma_{20}(\text{Treatment}_{ij}) + e_{ij} + r_{0j}$$

where Y_{ij} is the posttest score on measure Y for student i in classroom j, γ_{00} is the conditional mean posttest score for the control group, $\gamma_{20}(\text{Treatment}_{ij})$ is the fitted deflection of

the treatment group mean from the control on the posttest conditional mean, controlling for the student pretest (i.e., $\gamma_{10}(\text{Pretest}_{ij})$), e_{ij} is the student-level residual, r_{0j} is the class-level residual, and each residual has a respective variance. In a similar manner, a multilevel quantile regression equation for the primary impact analysis is:

$$Y_{ij_\tau} = \gamma_{00_\tau} + \gamma_{10_\tau}(\text{Pretest}_{ij}) + \gamma_{20_\tau}(\text{Treatment}_{ij}) + e_{ij_\tau} + r_{0j_\tau}$$

Note the equivalence between the two expressions such that each contains terms for the parameters in the model with subscripts noting clustering units, coefficients for means of the control group, pretest, and experimental group, and residuals with associated variances. The primary difference between the two equations is that the quantile expression includes the subscript τ, denoting the quantile at which the coefficients are estimated. A particular benefit of quantile regression models is that they do not make assumptions about the distribution of the outcome or predictor variables. The estimation process for intercept and slope coefficients in multilevel quantile is similar to traditional multilevel models in that it uses a loss function; the chief difference between the two approaches is that the latter conditions only on the mean whereas quantile regression conditions on as many points of the posttest as the user is interested in testing. Traditional multilevel models would require splitting posttest data into quartiles, quintiles, deciles, or the like in order to evaluate associations conditional on posttest performance, yet these methods have been shown to be problematic (Petscher, 2016) due to a truncated range of scores and diminished power for estimating associations. Quantile regression models do not truncate the sample or create groups of individuals, but rather use the full data matrix with differential weights to estimate the intercept and regression coefficients at each quantile. Subsequently, there is no loss of power or truncation of the range of the dependent variable. Quantile regression has been used in literacy research to more comprehensively understand the impact of form effects in oral reading fluency in predicting reading comprehension (Petscher & Kim, 2011) and the relationship between measures of universal screening with proximal and distal outcomes (Catts, Petscher, Schatschneider, Bridges, & Mendoza, 2009) and more recently for treatment effects (Wanzek et al., 2015).

Because the present design was a cluster randomized trial, it was necessary to use a linear quantile mixed model (lqmm; Geraci & Bottai, 2014), which estimates unique random and fixed effects for each specified quantile (see Geraci & Bottai, 2014 for technical details). Analyses were performed with the lqmm package (Geraci, 2012) for the R environment (R Development Core Team, 2012); however, the package only allows for two-level models. As such, the standard errors were cluster-corrected to account for school-level effects. Prior to running the lqmm, it was necessary to specify the number of quantiles to estimate. Although it is possible to specify as many quantiles as you have equal to the range of scores on the dependent variable, the specification should be based on the sample size and number of parameters in the number (Cade & Noon, 2003). Given the relatively smaller cluster size and overall participant pool, we opted to select five specific quantiles (i.e., .10, .25, .50, .75, and .90 quantiles) in order to reflect approximately how treatment effects may vary along lower levels of the conditional posttest distribution (i.e., .10 and .25) at the median of the conditional distribution (.50) and at higher levels of the conditional distribution (i.e., .75 and .90).

Results

Descriptive statistics for all pretest and posttest measures by group are found in Table 1. For the WJ III subtests, we provide descriptive statistics for *W* scores, which were utilized in the analyses, and we also include grade-based standard scores for ease of interpretation. Grade-based scores indicated that the RR group demonstrated positive growth across the school year on all measures, while scores for the TRI group remained essentially unchanged on WJ III Word Attack and declined on WJ III Passage Comprehension. RR intervention students made greater pre-post gains than TRI students on every measure except the Yopp-Singer test and WJ III Oral Comprehension, on which gains by the two groups were virtually identical.

Preliminary Analysis

A series of random effects mixed models were run to evaluate the extent to which the RR and TRI groups significantly differed on the baseline scores for each of the administered measures. The mixed model was set up as a simple multilevel regression of the pretest score on the dichotomous indicator of group. No differences were observed for any of the measures:

Table 1. Pre- and posttest means and standard deviations.

		Group			
		RR		TRI	
Measure	Time	Mean	*SD*	Mean	*SD*
Yopp- Singer[a]	Pre	10.39	6.21	9.76	7.24
	Post	16.89	3.50	15.57	6.37
WJ III Letter-Word Identification[b]	Pre	382.57	18.24	382.32	29.98
	Post	419.25	20.65	407.06	35.07
WJ III Letter-Word Identification[c]	Pre	90.21	10.84	89.02	17.73
	Post	98.76	10.16	92.26	17.05
WJ III Word Attack[b]	Pre	435.44	23.43	431.58	29.81
	Post	461.64	20.74	449.94	28.04
WJ III Word Attack[c]	Pre	99.07	12.17	96.05	16.40
	Post	101.89	11.34	95.46	15.48
Sight Word Fluency[d]	Pre	7.49	6.49	11.74	14.55
	Post	36.58	17.47	27.89	24.36
Decodable Word Fluency[d]	Pre	3.11	3.27	3.70	6.24
	Post	12.34	8.72	10.13	12.16
Decodable Sentence Fluency[e]	Pre	11.97	9.60	18.24	22.84
	Post	59.23	28.92	45.03	37.41
DIBELS Oral Reading Fluency[d]	Pre	4.52	4.79	8.73	13.17
	Post	28.66	23.39	24.71	23.68
WJ III Passage Comprehension[b]	Pre	419.03	16.09	423.22	16.80
	Post	447.38	15.69	439.54	20.38
WJ III Passage Comprehension[c]	Pre	87.15	10.41	89.51	10.86
	Post	91.32	11.30	85.82	13.90
WJ III Oral Comprehension[b]	Pre	462.82	13.81	462.16	14.77
	Post	468.98	13.28	467.54	14.10
WJ III Oral Comprehension[c]	Pre	93.61	12.16	91.48	15.23
	Post	93.72	12.57	92.26	13.25

Notes. [a]Score is total items correct; [b]*W* Scores; [c]Grade-Based Standardized Scores; [d]Score is words correctly read in 1 min; [e]Score is words correctly read in 2 min; RR = Reading RULES; TRI = Typical Reading Instruction; WJ III = Woodcock-Johnson III Tests of Achievement; Yopp-Singer = Yopp-Singer Test of Phoneme Segmentation; DIBELS = Dynamic Indicators of Basic Early Literacy Skills.

WJ III Word Attack [$t(19) = -0.67$, $p = .512$, $g = 0.17$], WJ III Letter-Word Identification [$t(19) = 0.16$, $p = .877$, $g = 0.06$]; Sight Word Fluency [$t(19) = 0.09$, $p = .927$, $g = 0.03$]; Yopp-Singer [$t(19) = 0.30$, $p = .766$, $g = 0.09$], Decodable Word Fluency [$t(19) = 0.43$, $p = .669$, $g = 0.15$]; Decodable Sound Fluency [$t(19) = -0.81$, $p = .426$, $g = 0.33$]; WJ III Oral Comprehension [$t(19) = 0.19$, $p = .852$, $g = 0.04$]; DIBELS Oral Reading Fluency [$t(19) = -0.95$, $p = .356$, $g = 0.41$]; or WJ III Passage Comprehension [$t(19) = -0.84$, $p = .409$, $g = 0.25$].

Impact Analysis

The initial unconditional model (Table 2) reflected the average score for each of the nine outcomes, as well as the random effects. With the exception of WJIII Oral Comprehension, where 3% of the respective variance in outcomes was due to classroom effects, the random effects for the other seven outcomes-indicated classrooms represented a large source of individual differences in the outcomes, ranging from 19% for DIBELS ORF to 47% for WJ III Letter-Word Identification.

Results for the impact analyses are reported in Table 3. Statistically significant results suggest that the intervention positively impacted student outcomes on WJ III Word Attack, WJ III Letter-Word Identification, Sight Word Fluency, Decodable Word Fluency, Decodable Sentence Fluency, and DORF. Although WJ III Passage Comprehension demonstrated marginally significant effects, it was found to be nonsignificant when the linear step-up was applied to control for Type 1 error. Outcomes on WJ III Oral Comprehension and the Yopp-Singer Test of Phoneme Segmentation did not differ between the groups.

Hedges's g effect sizes computed for each of the outcomes demonstrated that the intervention was associated with no impact on WJ III Oral Comprehension ($g = 0.04$), and small to moderate impacts on Yopp-Singer Test ($g = 0.19$), WJ III Word Attack ($g = 0.41$),

Table 2. Results from posttest unconditional mixed effects models.

Fixed effects	Intercept	*SE*	*df*	*t* value	*p* value
WJ III Word Attack	456.77	3.70	20	123.46	<.001
WJ III Letter-Word ID	413.83	4.61	20	89.71	<.001
Sight Word Fluency	29.66	1.22	20	24.40	<.001
Yopp-Singer TPS	16.21	0.56	20	28.77	<.001
Decodable Word Fluency	25.11	1.73	20	14.53	<.001
Decodable Sentence Fluency	53.08	5.13	20	10.34	<.001
WJ II Oral Comprehension	468.40	1.39	20	337.68	<.001
DIBELS Oral Reading Fluency	17.50	2.11	20	8.31	<.001
WJ III Passage Comprehension	443.81	2.83	20	156.67	<.001
Random effects	**Class variance**	**SE**	**Student residual**	**SE**	**Class ICC**
WJ III Word Attack	197.75	90.73*	386.30	64.98*	0.34
WJ III Letter-Word ID	353.11	139.37*	402.72	67.67*	0.47
Sight Word Fluency	22.80	9.71*	35.35	5.94*	0.39
Yopp-Singer	1.22	2.26*	23.74	3.98*	0.08
Decodable Word Fluency	44.03	19.71*	80.30	13.50*	0.35
Decodable Sentence Fluency	400.05	171.54*	659.33	110.48*	0.38
WJ III Oral Comprehension	0.00	.	179.50	26.47*	0.00
DIBELS Oral Reading Fluency	47.26	28.55*	198.33	32.98*	0.19
WJ III Passage Comprehension	123.19	52.54*	195.04	32.73*	0.39

Notes. WJ III = Woodcock-Johnson III Tests of Achievement; Yopp-Singer = Yopp-Singer Test of Phoneme Segmentation; DIBELS = Dynamic Indicators of Basic Early Literacy Skills.
*$p < .05$

Table 3. Results from conditional mixed effects models examining the impact of the intervention.

Outcome	Effect	Estimate	*SE*	*df*	*t* value	*p* value	Hedges's *g*
Woodcock-Johnson III	Intercept	217.99	33.50	19	6.51	<.001	
Word Attack	Treatment	9.59	4.08	19	2.35	.023	0.41
	Pretest	0.54	0.08	70	6.96	<.001	
Woodcock-Johnson III	Intercept	89.66	31.65	19	2.83	.016	
Letter-Word Identification	Treatment	12.21	5.13	19	2.38	.021	0.45
	Pretest	0.83	0.08	70	10.06	<.001	
Sight Word Fluency	Intercept	18.07	1.70	19	10.62	<.001	
	Treatment	5.24	1.62	19	3.24	.004	0.71
	Pretest	0.50	0.07	70	7.45	<.001	
Yopp-Singer Test of	Intercept	12.21	1.03	19	11.87	<.001	
Phoneme Segmentation	Treatment	0.96	0.95	19	1.00	.329	0.19
	Pretest	0.34	0.07	70	4.80	<.001	
Decodable Word Fluency	Intercept	15.49	1.91	19	8.10	<.001	
	Treatment	6.38	2.16	19	2.96	.008	0.60
	Pretest	0.51	0.08	70	6.09	<.001	
Decodable Sentence Fluency	Intercept	21.61	5.54	19	3.90	.001	
	Treatment	23.03	5.87	19	3.92	<.001	0.72
	Pretest	1.22	0.17	70	7.17	<.001	
Woodcock-Johnson III	Intercept	180.46	35.94	19	5.02	.001	
Oral Comprehension	Treatment	0.52	2.43	19	0.21	.834	0.04
	Pretest	0.62	0.08	70	8.00	<.001	
DIBELS Oral Reading Fluency	Intercept	8.21	2.63	19	3.12	.006	
	Treatment	6.66	2.96	19	2.25	.037	0.42
	Pretest	1.36	0.23	70	5.87	<.001	
Woodcock-Johnson III	Intercept	261.71	40.81	19	6.41	<.001	
Passage Comprehension	Treatment	10.16	4.87	19	2.09	0.05	0.57
	Pretest	0.42	0.10	70	4.35	<.001	

Notes. DIBELS = Dynamic Indicators of Basic Early Literacy Skills; All treatment effects $p < .05$ were statistically significant after applying linear step-up to the treatment effect coefficients. Adjusted critical alphas: Sight Word Fluency = .005, Decodable Sentence Fluency = .011, Decodable Word Fluency = .016, WJ III Letter-Word Identification = .022, WJ III Word Attack = .028, DIBELS Oral Reading Fluency = .039.

DIBELS Oral Reading Fluency ($g = 0.42$), WJ III Letter-Word Identification ($g = 0.45$), WJ III Passage Comprehension ($g = 0.57$), Decodable Word Fluency ($g = 0.60$), Sight Word Fluency ($g = 0.71$), and Decodable Sentence Fluency ($g = 0.72$). Effect sizes for all outcomes except Oral Comprehension and the Yopp-Singer phoneme segmentation measure met the What Works Clearinghouse (2014) standard for substantive importance (≥.25). Effect sizes for WJ III Letter-Word Identification and Word Attack were smaller than expected based on a comparison with average effect sizes on these measures in a group of previous studies of first-grade intervention for at-risk readers (i.e., .68 and .69, respectively; Denton et al., 2010; Wanzek & Vaughn, 2007), but our findings for Passage Comprehension were consistent with results from the same literature base (i.e., .62 effect size; Denton et al., 2010; Wanzek & Vaughn, 2007) .

Investigation of Intervention Impact Based on Pretest and Posttest Scores

In order to examine the effects of the intervention for different segments of the group of at-risk readers, we conducted secondary analyses to determine if intervention effects were differential based on conditional effects of the pretest (via moderation and simple slopes analysis) and posttest (via linear quantile mixed model) scores. Tests of moderation indicated that significant interaction terms were observed for WJ III Letter-Word identification ($p = .073$), Sight Word Fluency ($p = .003$), and Yopp-Singer ($p = .002$; see supplemental

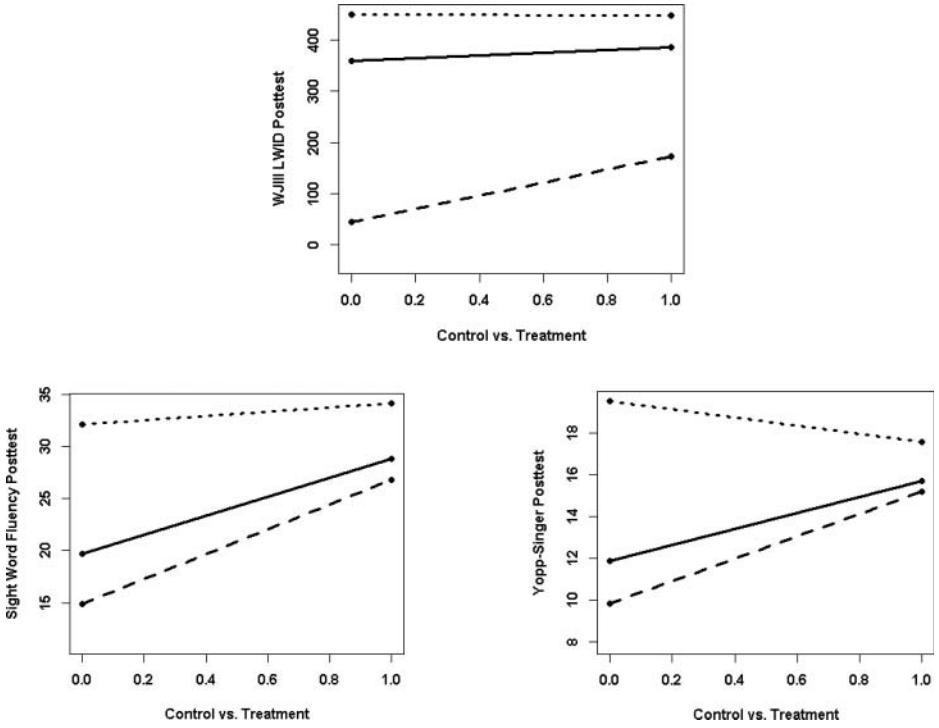

Figure 1. Pretest moderation of treatment effects for letter-word identification (top center), sight word fluency (lower left), and Yopp-Singer (lower right) for pretest scores at −1 *SD* (dashed line), the mean (solid line), and 1 *SD* (dotted line).

online materials for model coefficients). A simple slopes analysis (Figure 1) for WJ III Letter-Word Identification revealed that for students with baseline scores at or below one standard deviation below the sample mean, students in the treatment group outperformed the control by approximately 25 points ($t = 2.03$, $p = .046$, $g = 1.31$) but no effects were observed at one standard deviation above the mean ($t = -1.49$, $p = .928$). Sight Word Fluency moderation demonstrated a similar trend whereby significant effects in favor of the treatment group were found at one standard deviation below the sample mean ($t = 4.58$, $p < .001$, $g = 0.82$) but not at one standard deviation above the sample mean ($t = 1.09$, $p = .276$). The same pattern held for the Yopp-Singer for individuals below the sample mean ($t = 3.02$, $p = .004$, $g = 0.58$) and above the sample mean ($t = -1.62$, $p = .111$).

Results for the lqmm analyses are reported in Table 4. Although the primary impact analyses demonstrated a significant effect for WJ III Word Attack, the lqmm showed that the impact may be more strongly associated with the individuals who were the lowest performers on the measure at the end of the study. That is, for those students who ended the intervention study at approximately the 10th percentile of WJ III Word Attack (.10 quantile), intervention students outperformed the control students by approximately 27 points when controlling for the pretest. No other statistically significant effects were found for the intervention on Word Attack across the remaining quantiles. When considering WJ III Letter-Word Identification, the effect was significant for individuals who were at the .25 and .50 quantiles. For Sight Word Fluency, significant effects were observed across all estimated quantiles, and no substantive difference in the magnitude of treatment coefficients was

Table 4. Linear quantile mixed model results.

		.10 quantile		.25 quantile		.50 quantile		.75 quantile		.90 quantile	
Posttest	Parameter	Coeff	*SE*	Coeff	*SE*	Coeff	*SE*	Coeff	*SE*	Coeff	*SE*
WA	Intercept	194.11	77.33	215.14	53.85	270.68	48.2	289.6	33.2	343.55	42.86
	Pretest	0.51	0.17	0.51	0.11	0.42	0.1	0.4	0.07	0.28	0.1
	Treatment	27.38	11.42*	10.99	9.9	4.29	6.5	4.0	5.97	14.99	8.69
LWID	Intercept	66.99	48.62	59.87	4.65	68.37	50.56	140.31	52.82	149.22	58.81
	Pretest	0.83	0.13	0.88	0.11	0.87	0.13	0.73	0.14	0.74	0.15
	Treatment	9.86	6.82	11.76	4.97*	14.89	4.92**	11.78	6.86	6.70	7.90
SWF	Intercept	16.70	2.96	17.79	2.83	18.07	2.80	18.84	2.76	19.62	2.59
	Pretest	0.37	0.14	0.44	0.11	0.51	0.10	0.56	0.11	0.56	0.13
	Treatment	4.48	1.69*	4.89	1.62**	5.07	1.59**	5.59	1.6**	5.97	1.70***
Yopp	Intercept	4.75	3.37	9.99	2.65	13.67	1.19	16.99	1.19	19.99	1.12
	Pretest	0.50	0.23	0.34	0.15	0.38	0.10	0.19	0.08	0.03	0.05
	Treatment	1.00	2.44	1.57	1.90	−0.66	0.64	0.56	0.64	0.03	1.01
DWF	Intercept	11.65	3.02	13.92	3.07	15.66	2.45	17.85	2.70	20.13	2.60
	Pretest	0.41	0.16	0.52	0.12	0.54	0.06	0.53	0.11	0.53	0.13
	Treatment	4.20	2.94	5.30	2.75	6.33	2.48*	7.55	2.54**	9.10	2.60**
DSF	Intercept	10.71	5.68	15.27	5.93	20.91	5.14	22.69	5.02	46.49	5.83
	Pretest	0.99	0.34	0.98	0.18	1.06	0.22	1.95	0.35	1.27	0.62
	Treatment	16.43	5.80**	19.51	5.95**	22.96	5.88***	25.49	6.01***	47.67	7.03***
OC	Intercept	142.78	44.38	112.48	55.10	186.38	64.19	241.50	64.91	289.36	63.26
	Pretest	0.68	0.09	0.75	0.11	0.61	0.13	0.50	0.13	0.40	0.13
	Treatment	−4.84	3.02	−3.62	3.58	0.27	3.62	2.50	3.32	4.00	3.25
DORF	Intercept	0.66	2.01	4.05	2.51	6.26	2.30	10.80	1.98	16.51	2.76
	Pretest	1.29	0.20	1.24	0.24	1.45	0.14	1.46	0.22	1.49	0.69
	Treatment	2.56	1.88	3.99	1.82*	5.49	1.98**	9.86	2.29***	10.59	2.42***
PC	Intercept	234.45	54.59	176.02	47.75	232.59	43.66	265.61	51.08	294.51	48.73
	Pretest	0.44	0.13	0.60	0.10	0.48	0.10	0.44	0.11	0.35	0.11
	Treatment	8.11	8.38	7.32	5.55	10.28	5.18	11.00	7.65	20.07	5.76**

Notes. WA = WJ III Word Attack; LWID = WJ III Letter-Word Identification; SWF = Sight Word Fluency, Yopp = Yopp-Singer Test of Phoneme Segmentation; DWF = Decodable Word Fluency; DSF = Decodable Sentence Fluency; OC = WJ III Oral Comprehension; DORF = DIBELS Oral Reading Fluency; PC = WJ III Passage Comprehension.

observed. Such findings indicated that, regardless of where in the Sight Word Fluency distribution they fell, students in the RR Group consistently outperformed those in the TRI Group, and the magnitude of the effect was the same. This can be contrasted with Decodable Sentence Fluency, which also demonstrated significant treatment effects across all quantiles, but showed an increasing magnitude of effects; at the .10 quantile, the estimated difference between treatment and control groups was 5.80 points, compared with 7.03 at the .90 quantile. Treatment effects were also found for DORF (.25, .50, .75, .90 quantiles), suggesting that the intervention had a greater impact for those students who were not the weakest in their DORF skills at the end of the study. Though WJ III Passage Comprehension did not yield a significant treatment effect in the primary impact analysis, the lqmm revealed that a significant effect existed for those students who performed at the .90 quantile at posttest. The Yopp-Singer Test of Phoneme Segmentation and WJ III Oral Comprehension yielded no significant treatment effects across the quantiles, largely due to the lack of variance at the cluster level.

Implementation Fidelity and Dosage

Fidelity and quality of implementation were collected for the RR teachers; these data were coded on a Likert-type scale. The ratings were to a value representing the percentage of a

potential perfect rating for each observation, and then we calculated mean scores for each teacher across observations. Our a priori criterion for adequate fidelity of implementation was 80% of a possible perfect score.

Aggregated across observations and across teachers, the mean fidelity of implementation rating was 93% (SD = 5%; range = 82%–98%). Average fidelity ratings for all teachers exceeded 80%. For the individual lesson components, the average ratings were 95% (SD = 6%) for whole-group comprehension, 95% (SD = 4%) for small-group comprehension, 90% (SD = 7%) for word study, and 90% (SD = 9%) for text reading. Aggregated across observations and across teachers, the mean global quality of implementation rating was 89% (SD = 14%; range = 50%–100%). Quality ratings exceeded 80% for 12 of the 15 RR group teachers.

Another aspect of fidelity of implementation is the dosage of the intervention actually received by participants (Dane & Schneider, 1998). We calculated dosage in two ways, both based on information from weekly teacher logs. First, we calculated the mean number of days on which students received the intervention, and then we calculated the mean number of printed RR word study and comprehension lessons students covered during the intervention period. In both of these calculations, we excluded students who moved away from their schools during the study and thus received abbreviated programs. If teachers provided the program as requested, four days per week, it would have been possible to provide intervention on a total of 64 days. According to teacher report, students in the RR group received some instruction using the program on an average of 42 days (SD = 11), which is about 66% of the possible days. Students progressed through an average of 38 comprehension lessons, counting both the whole- and small-group components (SD = 13), and 35 word study and text reading lessons (SD = 18).

Discussion

This study was an initial evaluation of a first-grade supplemental reading intervention that provided systematic instruction in both word reading and comprehension, delivered by general education classroom teachers. The study had three main goals. First, the study examined effects, or student responsiveness to the supplemental intervention, based on reading outcomes. Second, through secondary exploratory analyses, we attempted to determine if student response to the RR intervention was dependent upon reading proficiency at pretest and posttest. These types of analyses have important implications for the development and implementation of reading intervention with the most at-risk students. Often, data analyses stop short of examining differential impact simply because all students who qualify for this type of intervention study are already deemed at risk for reading failure, implying that they are performing similarly on early reading tasks. However, there is much heterogeneity in at-risk samples, and it is possible that responsiveness to intervention could be differential based on levels of proficiency. Finally, we investigated the feasibility of the intervention when implemented in authentic first-grade classrooms.

Research Questions 1 & 2: Impacts on Student Outcomes

Word reading and decoding

The RR intervention demonstrated promise for supporting word reading and decoding in at-risk first-grade readers. The intervention was associated with significant positive impacts on

standardized assessments of word reading and phonological decoding and on researcher-constructed measures of sight word reading fluency and decodable word reading fluency. No significant differences or substantively important effects were detected in PA, and students in the RR and typical practice classrooms made about the same amount of progress in this domain. PA instruction is emphasized in most first-grade reading programs and is generally implemented consistently by first-grade teachers, so it is not surprising that both groups made progress in this area.

The magnitude of the effects on both standardized and researcher-developed measures of word reading and decoding can be considered substantively important in educational research (What Works Clearinghouse, 2014); however the effect sizes on the standardized measures of phonemic decoding and word reading were smaller than have been observed on the same measures in previous studies with similar students (Denton et al., 2010; Wanzek & Vaughn, 2007). Although we can only speculate, this may be explained by the fact that, for the most part, the interventions in the previous studies were implemented in locations outside of the regular classroom by specially trained teachers, and some were delivered in one-to-one formats. There are trade-offs to having Tier 2 interventions provided by classroom teachers in their own classrooms rather than by specialists in an alternate location. More students can be served by classroom teachers, and this model is less taxing on school resources, but effects may not be as strong as when intervention is provided outside of the classroom by specially trained interventionists. Classrooms can be active places, which may make it difficult for some at-risk students to attend to the teacher during small-group instruction and for the teacher to sustain his or her focus on the intervention group. A reading interventionist may be able to implement interventions with fewer distractions and interruptions. However, if a school has few resources available for designated reading interventionists, the current study suggests that it is possible for classroom teachers to implement a Tier 2 intervention that makes a positive impact on students' ability to decode and read words.

When we examined the effects for subgroups at pretest, for measures related to word study, there were significant impacts on students who were most at risk for reading difficulties when they received the RR intervention. Differential impact was noted for students performing at or below one standard deviation from the study sample, when compared to students performing at a similar level at pretest who did not get the RR intervention. This finding indicates that the word study portion of the intervention may be especially impactful for students who are the most at risk for reading difficulties.

Results from the lqmm found that the impact of the intervention on phonemic decoding was greatest for students who ended the study with the lowest scores on this variable (i.e., .10 quantile), while the impact on word reading was greatest for low-average students (i.e., those at the .25 and .50 quantiles). Although the lowest-performing students had significantly better outcomes in word reading and decoding when provided with the RR intervention relative to typical instruction, they remained in the lower part of the distribution at the end of the study. These students may benefit from a similar intervention delivered with increased intensity (i.e., more instructional time, smaller groups), characteristic of Tier 3 interventions. Effects of the intervention on sight word fluency were uniform for all students, regardless of where they fell in the posttest distribution on this measure. Similarly, the intervention's impact on fluent reading of decodable words was significant across the distribution, although it was stronger at the higher than at the lower quantiles.

Fluency

Students who received RR had significantly better outcomes than those who received typical instruction in the fluent reading of decodable sentences and nondecodable text (i.e., DORF). However, despite randomization, the TRI group had significantly higher DORF pretest scores than the RR group, so regression to the mean may have influenced study results. Secondary analyses demonstrated that there were significant posttest differences between the RR and TRI groups across the entire DORF posttest distribution except for students with the lowest fluency scores (i.e., at the .10 quantile), a finding that argues against regression to the mean as a threat to internal validity. Regardless, despite significantly better outcomes for the RR group on the DORF measure, the effect size did not meet our a priori benchmark for substantive educational importance ($\geq$.25; What Works Clearinghouse, 2014). This may be due in part to the intervention's relative brevity and teachers' relatively slow rates of progress through the RR program. Some students may not have progressed far enough in the decoding component to sufficiently automatize word recognition to the degree necessary to support strong growth in fluent reading of nondecodable text.

Comprehension

The lack of significant main effects for listening and reading comprehension was unexpected, given the emphasis in RR on explicit comprehension instruction and practice. Listening comprehension outcomes were virtually identical for treatment and comparison groups. Listening comprehension was measured with a norm-referenced test, and standard scores that account for grade level showed little change from pretest to posttest; thus, both the treatment and comparison groups grew at the expected rate during the study.

Effects of RR on reading comprehension approached significance, and the effect size met the criteria for substantive importance. Our secondary, exploratory lqmm analysis indicated that RR was associated with significantly better comprehension than TRI for one subgroup of students—those at the highest end of the distribution at posttest, indicating that the program should be strengthened for students who are more severely impaired in comprehension. In a study of a whole-class comprehension intervention similar to that provided in RR (Baker et al., 2013), students who were at risk for both language and literacy difficulties had weaker intervention responses than students who were only at risk for literacy difficulties. We do not have oral language assessments for participants in the current study, but we speculate that RR comprehension instruction may not have been sufficiently robust for children with impaired oral language. Because we had only 10–15 min of small-group comprehension lessons, vocabulary instruction was relatively brief. Enhancement of the vocabulary instruction in RR may result in stronger comprehension outcomes, especially for students with oral language challenges.

A second possible explanation for the weaker-than-expected outcomes in comprehension was the fact that the classroom teachers did not deliver the large-group or small-group comprehension instruction with the prescribed regularity. Systematic instruction in comprehension strategies may be more challenging for classroom teachers to implement than word study instruction. Comprehension instruction is, essentially, teaching children how to think as they read. To deliver explicit comprehension instruction, teachers must learn to "think aloud," to model their thought processes related to a variety of comprehension strategies, and they must learn to scaffold student responses to a variety of questions and tasks (e.g., summarization). Future implementations of the intervention should include more time

devoted to professional development in comprehension instruction and additional oral language instruction, and this component of the program could be revised to make it easier and more intuitive for teachers to implement.

Research Question 2: Implementation Fidelity and Dosage

On average, classroom teachers implemented RR with acceptable fidelity to program procedures, and the majority delivered it with acceptable quality; however, as a group, they did not implement it on the prescribed schedule. Classroom teachers have many demands on their time and energy, and reading interventions they implement need to be both effective and efficient. The RR intervention was developed through an iterative process over two years, with considerable input and feedback from first-grade classroom teachers who field-tested each component. As a result, the intervention became more effective and more "user-friendly"; however, further refinement appears warranted.

It is also likely that RR teachers need more professional development and support in using the program's mastery tests to adjust the rate of progress through the program for students who are able to move more quickly. If students can make accelerated growth, they should not be held back because a teacher progresses methodically through a set of lesson plans. A drawback to small-group instruction is that faster-progressing students are sometimes held back so that instruction is appropriate for those in the group who need extended practice. It may be beneficial for two or more teachers to deliver Tier 2 instruction simultaneously so that students can be grouped and regrouped across classrooms with students who progress at similar rates.

Summary

In sum, this pilot study demonstrated that the newly developed intervention has promise for supporting student outcomes in word reading and potentially in comprehension. Reading fluency outcomes, though significantly better than with typical practice, were weaker than expected. Classroom teachers are able to implement the program with sufficient fidelity and quality; however, they may not implement it on the prescribed schedule. This appeared to be most problematic for the comprehension portion of the program. We conclude that the intervention merits further study, but only after revisions designed to strengthen the oral reading fluency component and make the comprehension component more practical to implement, and after revisions in the professional development designed to encourage the use of mastery test data to accelerate student progress through the program when results indicate that they have mastered objectives and do not need further instruction on them.

Limitations

Because randomization occurred at the level of the classroom, the study was limited by a small sample size ($n = 21$ classrooms). Inadequate power due to a smaller sample size for a design study compared to a larger efficacy trial may have caused us to fail to detect intervention effects when they were present, especially in reading comprehension. A second limitation is that we were unable to directly observe the reading intervention provided to some comparison group students; thus, we cannot document the extent of overlap of that school-provided instruction with the intervention provided to the RR group. Last, practically

important baseline differences with $g > 0.20$ were observed in this study. The implication of practically important pretest differences is that the moderation analyses (e.g., DIBELS Oral Reading Fluency) may influence the finding of conditional effects. Future research is needed to replicate the findings of this study.

Implications for Practice

This study demonstrated that first-grade classroom teachers can implement a supplemental Tier 2 reading intervention within their own classrooms with high fidelity and quality, but that they may not implement it with the same level of intensity and consistency as might be typical when students receive intervention from a reading specialist, or researcher-trained interventionists. Regardless, it is important to note that not all children who are at risk for reading difficulties in first grade have access to small-group instruction from a reading specialist; for some schools, the resources are not available to provide supplemental small-group instruction outside of the general classroom setting. If many children respond positively to intervention provided within their own classrooms, the overall impact could be considerable. Developers of interventions that are designed for implementation by classroom teachers must ensure that the interventions are streamlined and efficient as well as effective to increase the likelihood of consistent implementation.

Conclusion

This study demonstrated that first-grade classroom teachers are able to implement a comprehensive supplemental reading intervention in the context of their own classrooms with fidelity and have positive impacts on at-risk readers' skill development. We also demonstrated that when investigating effects of intervention on reading outcomes, it is important to consider the differential impact dependent upon pretest and posttest reading proficiency; at-risk readers represent a heterogeneous group of students with responsiveness dependent upon differential proficiency.

Funding

This research was supported by the Institute of Education Sciences, U.S. Department of Education, through grant R324A100129 to the University of Texas Health Science Center at Houston. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Institute of Education Sciences or the U.S. Department of Education. The authors would like to thank Rebecca Beegle, Kelly Millner, Kelli Jimenez, and Sarah Stanley for their assistance.

ARTICLE HISTORY

Received 6 May 2016
Revised 29 May 2017
Accepted 28 August 2017

References

Aarnoutse, C. A., Van den Bos, K. P., & Brand-Gruwel, S. (1998). Effects of listening comprehension training on listening and reading. *The Journal of Special Education*, *32*(2), 115–126.

Aarnoutse, C., Van Leeuwe, J., Voeten, M., & Oud, H. (2001). Development of decoding, reading comprehension, vocabulary and spelling during the elementary school years. *Reading and Writing*, *14* (1–2), 61–89.

Allor, J., & McCathren, R. (2004). The efficacy of an early literacy tutoring program implemented by college students. *Learning Disabilities Research & Practice*, *19*(2), 116–129.

Baker, S., Gersten, R., & Keating, T. (2000). When less may be more: A 2-year longitudinal evaluation of a volunteer tutoring program requiring minimal training. *Reading Research Quarterly*, *35*(4), 494–519.

Baker, S. K., Santoro, L. E., Chard, D. J., Fien, H., Park, Y., & Otterstedt, J. (2013). An evaluation of an explicit read aloud intervention taught in whole-classroom formats in first grade. *Elementary School Journal*, *113*, 331–358.

Ball, E. W., & Blachman, B. A. (1991). Does phoneme awareness training in kindergarten make a difference in early word recognition and developmental spelling? *Reading Research Quarterly, 26*(1), 49–66.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological), 57*, 289–300.

Berkeley, S., Bender, W. N., Peaster, L. G., & Saunders, L. (2009). Implementation of response to intervention: A snapshot of progress. *Journal of Learning Disabilities*, *42*(1), 85–95.

Berninger, V. W., Vermeulen, K., Abbott, R. D., McCutchen, D., Cotton, S., Cude, J., & Sharon, T. (2003). Comparison of three approaches to supplementary reading instruction for low-achieving second-grade readers. *Language, Speech, and Hearing Services in Schools*, *34*(2), 101–116.

Cade, B. S., & Noon, B. R. (2003). A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment*, *1*(8), 412–420.

Carlisle, J. F., & Felbinger, L. (1991). Profiles of listening and reading comprehension. *The Journal of Educational Research*, *84*(6), 345–354.

Catts, H. W., Compton, D., Tomblin, J. B., & Bridges, M. S. (2012). Prevalence and nature of late-emerging poor readers. *Journal of Educational Psychology*, *104*(1), 166–181.

Catts, H. W., Hogan, T. P., & Fey, M. E. (2003). Subgrouping poor readers on the basis of individual differences in reading-related abilities. *Journal of Learning Disabilities*, *36*(2), 151–164.

Catts, H. W., Petscher, Y., Schatschneider, C., Bridges, M. S., & Mendoza, K. (2009). Floor effects associated with universal screening and their impact on early identification. *Journal of Learning Disabilities*, *42*(2), 163–176.

Children's Learning Institute and Texas Institute for Measurement, Evaluation, and Statistics (2010). *Technical report: Texas Primary Reading Inventory (2010–2014 Edition)*. Retrieved from https://www.tpri.org/resources/documents/20102014TechnicalReport.pdf

Coyne, M. D., Kame'enui, E. J., Simmons, D. C., & Harn, B. A. (2004). Beginning reading intervention as inoculation or insulin: First-grade reading performance of strong responders to kindergarten intervention. *Journal of Learning Disabilities*, *37*(2), 90–104.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, *18*(1), 23–45.

Denton, C. A., Solari, E. J., Ciancio, D. J., Hecht, S. A., & Swank, P. R. (2010). A pilot study of a kindergarten summer school reading program in high-poverty urban schools. *The Elementary School Journal*, *110*(4), 423–439.

Duke, N. K., Pressley, M., Hilden, K., Stone, C. A., & Silliman, E. R. (2004). Difficulties with reading comprehension. In A. E. Farstrup & S. J. Samuels (Eds.), *Handbook of language and literacy: Development and disorders*, (501–520). New York: Guilford.

Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of educational research*, *71*(3), 393–447.

Elbaum, B., Vaughn, S., Hughes, M. T., & Moody, S. W. (2000). How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. *Journal of Educational Psychology*, *92*(4), 605–619.

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology*, *88*(1), 3–17.

Fuchs, D., Fuchs, L. S., Otaiba, S. A., Thompson, A., Yen, L., Mcmaster, K. N., ...& Yang, N. J. (2001). K-PALS helping kindergartners with reading readiness: Teachers and researchers in partnerships. *Teaching Exceptional Children, 33*(4), 76–80.

Gajria, M., Jitendra, A. K., Sood, S., & Sacks, G. (2007). Improving comprehension of expository text in students with LD: A research synthesis. *Journal of Learning Disabilities*, *40*(3), 210–225.

Garner, J. K., & Bochna, C. R. (2004). Transfer of a listening comprehension strategy to independent reading in first-grade students. *Early Childhood Education Journal*, *32*(2), 69–74.

Geraci, M. (2013). *lqmm: Linear quantile mixed models* (R package version 1.03). Retrieved from http://CRAN.R-project.org/package=lqmm

Geraci, M., & Bottai, M. (2014). Linear quantile mixed models. *Statistics and Computing, 24*(3), 461–479.

Gersten, R., Fuchs, L. S., Williams, J. P., & Baker, S. (2001). Teaching reading comprehension strategies to students with learning disabilities: A review of research. *Review of Educational Research*, *71* (2), 279–320.

Gersten, R. M., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W. D. (2009). *Assisting students struggling with reading: Response to intervention and multi-tier intervention in the primary grades*. Washington, DC: US Department of Education, National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences.

Good, R. H., III, & Kaminski, R. A. (Eds.). (2002). *Dynamic indicators of basic early literacy skills (DIBELS)* (6th ed.). Eugene: University of Oregon, Institute for the Development of Educational Achievement.

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, *2*(3), 172–177.

Jaffe, L. E. (2009). *Development, interpretation, and application of the W score and the relative proficiency index* (Woodcock-Johnson III Assessment Service Bulletin No. 11). Rolling Meadows, IL: Riverside.

Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society*, *46*, 33–50.

Leach, J. M., Scarborough, H. S., & Rescorla, L. (2003). Late-emerging reading disabilities. *Journal of Educational Psychology*, *95*(2), 211–224.

Mathes, P. G., Denton, C. A., Fletcher, J. M., Anthony, J. L., Francis, D. J., & Schatschneider, C. (2005). The effects of theoretically different instruction and student characteristics on the skills of struggling readers. *Reading Research Quarterly*, *40*(2), 148–182.

Mathes, P. G., Grek, M. L., Howard, J. K., Babyak, A. E., & Allen, S. H. (1999). Peer-assisted learning strategies for first-grade readers: A tool for preventing early reading failure. *Learning Disabilities Research & Practice*, *14*(1), 50–60.

McMaster, K. L., Fuchs, D., Fuchs, L. S., & Compton, D. L. (2005). Responding to nonresponders: An experimental field trial of identification and intervention methods. *Exceptional Children*, *71*(4), 445–463.

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, *9*(4), 599–620.

Nation, K., & Snowling, M. J. (2004). Beyond phonological skills: Broader language skills contribute to the development of reading. *Journal of Research in Reading*, *27*(4), 342–356.

National Center on Intensive Intervention (2015). *Academic Progress Monitoring GOM Tool Chart*. Retrieved from http://www.intensiveintervention.org/chart/progress-monitoring/13077#relps

Petscher, Y. (2016). Do our means of inquiry match our intentions? *Frontiers in Psychology, 7*:1048.

Petscher, Y., & Kim, Y. S. (2011). The utility and accuracy of oral reading fluency score types in predicting reading comprehension. *Journal of School Psychology*, *49*(1), 107–129.

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interaction effects in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*, *31*(4), 437–448.

R Core Development Team (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from: http://www.R-project.org/

Schacter, J., & Jo, B. (2005). Learning when school is not in session: A reading summer day-camp intervention to improve the achievement of exiting first-grade students who are economically disadvantaged. *Journal of Research in Reading*, *28*(2), 158–169.

Shanahan, T., Callison, K., Carriere, C., Duke, N. K., Pearson, P. D., Schatschneider, C., & Torgesen, J. (2010). *Improving reading comprehension in kindergarten through 3rd grade: IES practice guide* (NCEE 2010–4038). Washington DC: National Center for Education Evaluation and Regional Assistance, Institute of Education, U.S. Department of Education. Retrieved from https://ies.ed.gov/ncee/wwc/

Snow, C. E., Burns, S., & Griffin, P. (Eds.) (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Solari, E. J., & Gerber, M. M. (2008). Early comprehension instruction for Spanish-speaking English language learners: Teaching text-level reading skills while maintaining effects on word-level skills. *Learning Disabilities Research & Practice*, *23*(4), 155–168.

Spectrum K12 School Solutions (2009). *Response to intervention (RTI) adoption survey 2009*. Retrieved from http://www.spectrumk12.com

Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools*, *42*(8), 795–819.

Swanson, H. L. (2000). What instruction works for students with learning disabilities? Summarizing the results from a meta-analysis of intervention studies. In R. M. Gersten, E. P. Schiller, & S. Vaughn (Eds.), *Contemporary special education research: Syntheses of the knowledge base on critical instructional issues* (pp. 1–30). Mahwah, NJ: Erlbaum.

Swanson, H. L., & Hoskyn, M. (1998). Experimental intervention research on students with learning disabilities: A meta-analysis of treatment outcomes. *Review of Educational Research*, *68*(3), 277–321.

Torgesen, J. K. (2004). Lessons learned from research on interventions for students who have difficulty learning to read. In P. McCardle & V. Chhabra (Eds.), *The voice of evidence in reading research* (pp. 355–382). Baltimore, MD: Brookes.

Torgesen, J. K., & Burgess, S. R. (1998). Consistency of reading-related phonological processes throughout early childhood: Evidence from longitudinal-correlational and instructional studies. In J. Metsala & L. Ehri (Eds.), *Word recognition in beginning reading*. Hillsdale, NJ: Erlbaum.

Vaughn, S., Gersten, R., & Chard, D. J. (2000). The underlying message in LD intervention research: Findings from research syntheses. *Exceptional Children*, *67*(1), 99–114.

Vellutino, F. R., Scanlon, D. M., Small, S., & Fanuele, D. P. (2006). Response to intervention as a vehicle for distinguishing between children with and without reading disabilities: Evidence for the role of kindergarten and first-grade interventions. *Journal of Learning Disabilities*, *39*(2), 157–169.

Wanzek, J., Petscher, Y., Al Otaiba, S., Kent, S. C., Schatschneider, C., Haynes, M., … Jones, F. G. (2015). Examining the average and local effects of a standardized treatment for fourth graders with reading difficulties. *Journal of Research on Educational Effectiveness*, *9*(S1), 1–22.

Wanzek, J., & Vaughn, S. (2007). Research-based implications from extensive early reading interventions. *School Psychology Review*, *36*(4), 541–561.

What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education (2014). *WWC procedures and standards handbook (Version 3.0)*. Retrieved from http://ies.ed.gov/ncee/wwc/documentsum.aspx?sid=19

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, *48*(4), 817–838.

Williams, J. P. (2005). Instruction in reading comprehension for primary-grade students: A focus on text structure. *The Journal of Special Education*, *39*(1), 6–18.

Wise, J. C., Sevcik, R. A., Morris, R. D., Lovett, M. W., & Wolf, M. (2007). The relationship among receptive and expressive vocabulary, listening comprehension, pre-reading skills, word

identification skills, and reading comprehension by children with reading disabilities. *Journal of Speech, Language, and Hearing Research*, *50*(4), 1093–1109.
Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III tests of achievement*. Itasca, IL: Riverside Publishing.
Yopp, H. K. (1988). The validity and reliability of phonemic awareness tests. *Reading Research Quarterly*, *23*(2), 159–177.
Yopp, H. K. (1995). A test for assessing phonemic awareness in young children. *Reading Teacher*, *49*(1), 20–29.