Title

Effects From a Randomized Control Trial Comparing Researcher and School-Implemented Treatments With Fourth Graders With Significant Reading Difficulties

Journal

Journal of Research on Educational Effectiveness, 9(sup1)

ISSN

1934-5747

Authors

Vaughn, Sharon
Solís, Michael
Miciak, Jeremy
et al.

Publication Date

2016-10-03

DOI

10.1080/19345747.2015.1126386

Peer reviewed

HHS Public Access

Author manuscript

J Res Educ Eff. Author manuscript; available in PMC 2017 May 08.

Published in final edited form as:

J Res Educ Eff. 2016 ; 9(Suppl 1): 23–44. doi:10.1080/19345747.2015.1126386.

Effects From a Randomized Control Trial Comparing Researcher and School-Implemented Treatments With Fourth Graders With Significant Reading Difficulties

Sharon Vaughn,
The University of Texas at Austin, Austin, Texas, USA

Michael Solís,
University of Virginia, Charlottesville, Virginia, USA and The University of Texas at Austin, Austin, Texas, USA

Jeremy Miciak,
University of Houston, Houston, Texas, USA

W. Pat Taylor, and
University of Houston, Houston, Texas, USA

Jack M. Fletcher
University of Houston, Houston, Texas, USA

Abstract

This study examined the effectiveness of a researcher-provided intervention with 4th-graders with significant reading difficulties. The intervention emphasized multi-syllable word reading, fluent reading of high frequency words and phrases, vocabulary, and comprehension. To identify the participants, 1,695 fourth grade students were screened using the Gates MacGinitie Reading Test, and those whose standard score was 85 or lower were included in the study (N=485). Participants were randomly assigned (2:1) to receive either researcher-provided intervention (n=324) or intervention provided by school personnel (business as usual, BAU) (n=161). Findings revealed no statistically significant differences between students in the researcher-provided intervention and BAU groups. Using effect sizes as an indicator of impact, students in the researcher implemented treatment generally outperformed students in the school implemented treatment (BAU). Examining growth in standard scores, both groups made significant gains in reading outcomes with standard score growth from pretest to posttest of 3 standard score points on decoding, 5 on fluency,, and 2.0 to 7 standard score points on reading comprehension measures.

Keywords

reading intervention; reading comprehension; reading difficulties; fourth grade

Address correspondence to Sharon Vaughn, The Meadows Center for Preventing Educational Risk, The University of Texas at Austin, 1912 Speedway D4900, Austin, TX, 78712-1284, USA. srvaughn@austin.utexas.edu.

Fourth grade is a critical time in the development of reading proficiency. Chall and Jacobs (1983) first identified the “fourth grade slump”, a phenomenon in which many 3rd grade students who were previously reading on grade level experienced a drop in normative reading scores in 4th grade. They suggested that these students were not going “backwards” in reading, but were instead failing to meet grade-level expectations as text complexity increased and the reading task became more difficult. Compton, Fuchs, Fuchs, Elleman, and Gilbert (2008) described students whose reading performance was relatively typical prior to 4th grade but then demonstrated significant reading problems in grade 4 as students with late-emerging reading disability. Thus, the upper elementary grades (4th and 5th) may be particularly good targets for reading interventions because these grades are the last in which formal reading instruction traditionally occurs and also are grades in which reading to learn and understand complex texts is initially most apparent.

Considering the importance of the upper elementary grades as targets for students with reading difficulties, there has been relatively little research on the efficacy of reading approaches in these grades (4 and 5) (see for review, Wanzek, Wexler, Vaughn, & Ciullo, 2010). Only nine experimental studies and four quasi-experimental studies of reading interventions for struggling 4th and 5th graders over a thirty-year period were located in a synthesis conducted by Wanzek and colleagues (2010). Based on these studies, outcomes for vocabulary and comprehension interventions were moderate-to- high. However, all studies utilized researcher-developed, proximal measures that typically yield higher effects than standardized measures (Swanson, Hoskyn, & Lee, 1999). Word study treatments were associated with small to moderate effects. The two studies that examined multi-component interventions (e.g., word study and comprehension) also did not use standardized measures of reading comprehension. In sum, the authors recommended that research on populations in 4th grade and above utilizing randomized control trial designs and standardized outcome measures was necessary because of the limited number of rigorous studies investigating the effects of multi-component interventions.

Two studies relate closely to our proposed study, addressing multi-component reading interventions for students in 4th and 5th grades with reading problems. Therrien, Wickstrom, and Jones (2006) addressed fluency and comprehension and reported significant differences favoring the treatment group on fluency but not on passage comprehension. In a multi-component intervention using content texts in science, Ritchey and colleagues (Ritchey, Silverman, Montanaro, Speece, and Schatschneider, 2012) provided 24 sessions of treatment to 4th graders with low reading comprehension. They report mixed results with the treatment condition outperforming the comparison on near transfer tasks highly aligned with the treatment but no differences on standardized measures of comprehension.

Considering the possible influence of reading interventions on older students, we examined findings from two recent syntheses investigating the effectiveness of reading interventions for students with reading difficulties in grades 4 through 12 (Scammacca, Roberts, Vaughn, & Stuebing, 2013; Wanzek et al., 2013). Scammacca and colleagues utilized meta-analytic techniques to estimate mean effect sizes for reading interventions for students in grades 4–12. Analyzing over 82 study-wise effect sizes, Scammacca and colleagues reported an overall mean effect (d) of 0.49 for all measures and 0.21 for standardized measures. For all

studies, vocabulary interventions had a significantly larger mean effect sizes than all other types of interventions (e.g., word study, comprehension), but this effect was confounded by the use of researcher-developed measures to estimate vocabulary acquisition. Scammacca and colleagues considered grade level as a moderator and determined that there were no significant differences for studies that addressed 4th–5th graders, 6th–8th graders, and 9th–12th graders.

The Wanzek et al., (2013) synthesis differed from the Scammacca et al., (2013) synthesis in its focus on extensive interventions (i.e., interventions providing 75 or more sessions) in Grades 4–12. Yielding considerably fewer studies and effects, the mean effect sizes ranged from 0.10 to 0.16 with no significant differences in student outcomes for grade level of intervention. The authors noted that based on findings from a previous synthesis of extensive interventions in kindergarten through 3rd grade (Wanzek & Vaughn, 2007), there were considerably fewer studies conducted in grades 4th through 12th and the effects for older grades were considerably smaller. The explanation for smaller effects as students get older may be a result of more complex text associated with older grades including more difficult language and concepts, and older students demonstrating more intractable reading difficulties. Authors of both syntheses called for additional intervention studies with students beyond grade three that utilized rigorous designs.

The findings from these studies on interventions with older students may be considered in light of several important issues that are not adequately reflected in aggregated effect sizes. First, the effect sizes favoring treatment students may have been inflated if the comparison students were not participating in any reading instruction. Unlike in elementary school where all students receive reading instruction, reading instruction at the middle school may not be formal and may be represented as part of occasional vocabulary or comprehension activities in the content area. Second, most of the interventions represented in the syntheses were relatively short in duration (less than 2 months). Finally, insufficient data were available from the studies to determine whether the interventions improved student outcomes relative to grade-level expectations. Also, while students in treatment are making gains they have not closed the gap relative to typically developing peers, which is particularly noteworthy with older students because these students are more likely to be multiple grade levels behind the normative sample.

Study Purpose

The purpose of this study was to implement a randomized control trial to evaluate the outcomes of a comprehensive, researcher-provided intervention, contrasted with school implemented intervention, for students with severe reading difficulties in Grade 4. Responding to the need for rigorous research designs using standardized outcome measures (Scammacca et al., 2013), we designed the study to address effects from an extensive year-long intervention using reliable and valid measures to determine program efficacy. This study is part of a large-scale, multiyear study designed to examine the efficacy of intensive interventions for upper elementary students with significant reading difficulties. The study reported here presents an intervention and its implementation format that was specifically designed to be feasible, given the realities of elementary schools.

Our primary research question was: what are the effects of a researcher implemented intervention contrasted with school-implemented intervention on the reading related outcomes of individuals with reading difficulties? Based on previous reviews of interventions with upper elementary students with reading difficulties (e.g., Wanzek et al., 2013), we hypothesized that both interventions would be associated with modest gains over time and that the researcher-provided intervention would result in improved outcomes for students relative to other students at risk for reading difficulties.

Framework for Reading Intervention

The Simple View of Reading (SVR) (Gough & Tunmer, 1986; Hoover & Gough, 1990) served as the overarching framework guiding the development of the researcher developed intervention. The SVR defines reading comprehension as the product of two complex, yet distinct skills: word recognition and listening comprehension. Word recognition refers to decoding and reading printed words regardless of context; listening comprehension is the ability to understand language. Automaticity of word reading is essential for comprehension because it minimizes the need to focus on lower-level processes, allowing readers to devote resources to higher-level comprehension processes (Perfetti, 2011). The SVR asserts that the decoding and automaticity (word-level skill) aspect of the model is specific to reading and that comprehension is shared in both written and oral language. Research on the SVR model indicates that word reading automaticity and language comprehension can account for 40%–85% of variance in reading comprehension at various grade levels (e.g., Catts, Adlof, Hogan, & Weismer, 2005; Cirino et al., 2013; Johnston & Kirby, 2006; Joshi & Aaron, 2000; Savage, 2006).

Our prototype reading intervention was built upon and extended treatments designed and evaluated through prior research (Author). Applying the SVR and recognizing that reading problems have multiple sources and involve the interaction of lower-level (e.g., decoding) and higher-level (e.g., inference instruction) processes, we designed daily lessons to target word reading automaticity including high frequency words and strategies for decoding multi-syllable words, vocabulary acquisition through essential word instruction as well as morphology, building background knowledge by presenting lower level texts and then higher level texts on the same topic, and text processing and inference making. We organized these components into three working constructs: (1) *word reading* (automaticity in reading high-frequency and multi-syllable words), (2) *world knowledge* (vocabulary and background knowledge), and (3) *text-processing strategies* (including mental models and inference-making). The focus on (2) and (3) specifically targets the language comprehension component within the overarching framework of the SVR, whereas the focus on (1) targets the decoding component.

We derived procedures for enhancing world knowledge (vocabulary/comprehension) from vocabulary instruction research (Beck, McKeown, & Kucan, 2013; Elleman, Lindo, Morphy, & Compton, 2009; Scammacca et al., 2007) and background knowledge building reading practices (Vaughn et al., 2014). Such practices provide initial instruction and depth of knowledge through activities that include definitions, illustrations, word building, word associations, contextual clues for word meaning and text understanding, and use of

vocabulary in original sentences (Baumann, Edwards, Boland, Olejnik, & Kame'enui, 2003; Beck et al., 2013; Kamil et al., 2008).

Method

Participants

School sites—This study was conducted in one large urban district (8 elementary schools) and two near urban school districts (9 elementary schools) in the southwestern United States, with approximately half the sample from the large urban district. These districts were selected to provide geographic and demographic diversity across sites. Within participating districts, schools were invited to participate if they were rated academically acceptable and the school's schedule permitted the 35-minute intervention period. The goal was to identify schools that would yield a demographically diverse student sample that would approximate the demographics of urban districts in the larger region. Each participating school individually agreed to the procedures of the research study. The mean enrollment of the 17 participating schools was 697 students (range 425–1140 students). All schools included a significant population of students qualifying for free or reduced lunch with a mean percentage of 81.6% (range 46.1%–98.4%).

Selection of participants—We administered the Gates MacGinitie Reading Test (MacGinitie, MacGinitie, Maria, Dreyer, & Hughes, 2000)) to all fourth graders and then identified students whose standard score was 85 or lower as participants in the sample. Students were excluded from screening only if: (a) they were enrolled in an alternative curriculum, (i.e., life skills class), or (b) they were identified as having a significant sensory or intellectual disability that interfered with participation in the study (e.g., blindness, deafness). Students with mild disabilities such as learning disabilities or speech and language impairments were included in the screening sample as long as they participated in a general education English Language Arts course.

Student participants—The preliminary sample included 1,695 fourth-grade students who attended one of the designated elementary schools. A total of 487 students scored below the cut point. Of these 487 students, three were not randomized because: (a) a parent refused participation ($n = 1$); (b) the student did not complete the screening test and the score was considered invalid ($n = 1$); and (c) the student demonstrated significant intellectual deficits unknown at the time of screening ($n = 1$). The 484 eligible struggling readers were randomly assigned within school in a 2:1 ratio to a researcher provided treatment ($n = 323$) or a school provided treatment, business as usual (BAU) ($n = 161$). Since all students had significant reading difficulties and it was not ethically possible to ask the schools to withhold treatment, school personnel chose to use their resources to provide treatment to students in the BAU condition. The mean age of the sample was 9.8 years (range 8.8–12.0 years). Participants were 44% female and 87% of the sample qualified for free or reduced lunch. Approximately 14% of the sample had been previously identified for special education. The racial/ethnic composition of the sample was 68% Hispanic, 22% African-American, 2% Caucasian, and 8% other (including students listing two or more race/ethnicities). There were no statistically significant associations between treatment assignment and age ($t(479) = 1.15$, $p > .05$), free

or reduced lunch status (χ^2 (1) = 0.00, $p > .05$), special education status (χ^2 (1) = 0.34, $p > .05$), or race/ethnicity (χ^2 (1) = 3.20, $p > .05$).

Thirteen of the 323 students initially assigned to receive the intervention were unable to be assigned because of scheduling conflicts at their school. An additional 17 moved away from their schools during the school year. Seventeen students were withdrawn from the study by administrators at their school because of scheduling conflicts or concerns about behavior and 7 parents asked that their child be withdrawn. One student was withdrawn because of significant visual impairments that were unknown at the time of randomization. Another two students were withdrawn for unknown reasons. Students who attrited were compared to students who remained to investigate the effects of attrition. Attrited students did not differ from students who remained at pre-test on the Gates-MacGinitie Reading Test, $t(482) = 0.05$, $p > .05$. There was no association between attrition and site (urban site vs. near urban site; χ^2 (1) = 2.50, $p > .05$) or attrition and treatment condition (treatment vs control; χ^2 (1) = 0.76, $p > .05$).

Measures

Decoding and spelling—We assessed word reading accuracy and spelling with the Letter-Word Identification and Spelling subtests of the Woodcock-Johnson III Tests of Achievement (WJ-III; Woodcock, McGrew, & Mather, 2001). Published test-retest reliability coefficients for students aged 8–13 range from .84 to .85.

Fluency—The Sight Word Efficiency subtest from the Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1999) assessed word list fluency for real words and pseudowords. Alternate-forms reliability of this well-standardized test exceed .90 (Torgesen, Wagner, & Rashotte, 1999). The Test of Sentence Reading Efficiency and Comprehension (TOSREC; Wagner, Torgesen, Rashotte, & Pearson, 2010) is a 3-minute, group-based assessment that was also used to assess reading fluency. Students are presented with a series of short sentences and are required to assess their veridicality. Published corrected alternate-form reliability coefficients for students in 4th grade range from .82 to .86.

Comprehension—The WJ-III Passage Comprehension subtest (Woodcock et al., 2001), a cloze-based assessment in which students read a passage and fill in missing words, was also used to assess comprehension. Published test-retest reliabilities for children aged 8–13 range from .76 to .86. Gates MacGinitie Reading Test–Fourth Ed (GM-RT), (MacGinitie et al., 2000). This test is a timed, group-administered assessment of reading comprehension. It consists of expository and narrative passages ranging in length from 3 to 15 sentences. Students answer three to six multiple-choice questions related to the most recently read passage. Published internal consistency reliability coefficients range from .91 to .93, and alternate form reliability is reported as .80 to .87.

Progress monitoring—AIMSweb Curriculum-Based Measure: Reading (CBM-R; Shinn, 2002) probes for 4th grade were administered to students in the researcher-implemented treatment (procedures described below). The oral reading fluency (ORF) probes require the

student to read aloud for one minute yielding a Words Correct per Minute score, which was utilized for progress monitoring purposes. The progress monitoring measures were used to inform instruction only and are not reported as part of the evaluation of intervention effectiveness because no posttest was administered.

Intervention

Reading intervention instruction was delivered to small groups of four to five students for 35-min sessions five times per week for approximately 16 weeks from November through April. Lessons were organized into two-week thematic units consisting of ten lessons, including an end of unit Maze activity for lesson 10. See Table 1 for a sequence and organization of the instruction. The text aligned with what was taught in the students' social studies classes. We chose social studies content because it allowed us to include both expository and narrative texts and provided multiple opportunities to build background knowledge. The standardized researcher-provided intervention lessons consisted of three components presented daily: word and concept building (i.e., vocabulary knowledge; 3 to 10 minutes), text reading of either a narrative or expository passage (15 to 20 minutes), and word study (i.e., decoding; 6 to 10 minutes).

Vocabulary—Vocabulary instruction occurred for 3 to 10-mins for nine of the ten lessons in each unit. Within each ten-lesson unit, six vocabulary words related to key concepts for the unit's social studies texts were explicitly taught. Selection of the six words within each unit was the result of a collaborative process, in which each member of the research team individually identified potential vocabulary words from that unit's selected texts. Members of the research team then convened to share their vocabulary word choices and discuss until a consensus was reached about which six words to teach for each unit. Words were selected based on multiple considerations, including: (a) the utility of the word across multiple contexts, (b) the word's relation to key social studies concepts, and (c) the likelihood that students would know the word. Words were introduced to students through an explicit instruction routine of presenting simplified definitions, visual representations, and synonyms followed by turn-and-talk questions that students answered with a partner. Questions were designed to encourage application of each word as it related to the text and to personal life experiences. Two words were introduced during the first lesson followed by review of the words during lessons two and three. This three-day pattern that included word introduction and two lessons of review was repeated three times during the ten day unit. On day 10, students completed a MAZE activity as a curriculum-based measure to check for word understanding and inform future instruction.

Text-based Reading—The readings for each unit were comprised of two types of text: (1) stretch-text, (2) fluency-text. Stretch-text instruction (i.e., texts on grade-level not reading level) occurred for 15 to 20-mins during lessons four through seven in each unit. The stretch-text was grade-level expository text of social studies topics (i.e., the American Revolution) that was adapted slightly to improve the readability for struggling readers. Stopping points were identified every three to four paragraphs based on the organization of the text and its ideational development. At each stopping point, students were prompted to explain in their own words the meaning of the text (i.e., what is this part of the text about?).

Additionally, tutors asked text-based questions that required students to synthesize information from different sections of the text. When students had difficulty answering a question, tutors scaffolded the activity by identifying a shorter section of text in which the answer could be found and asking students to re-read the section to find the answer. This scaffold was repeated with increasingly smaller sections of text until the student identified the correct answer. During stretch-text reading (e.g., texts on grade level not reading level), tutors utilized a variety of reading routines dependent upon the decoding and fluency skills of individual students and the group, including repeated reading, choral reading, partner reading, or individual silent reading.

Fluency-text instruction occurred for 15 to 20-mins during lessons one through three and lessons eight and nine. The fluency-texts were passages from the QuickReads program (Hiebert, 2003). Quickreads passages were designed to reduce the cognitive load required to complete lower-level reading tasks and facilitate comprehension by regulating the linguistic content and limiting the number of unfamiliar words. We attempted to identify QuickReads related to topics in the stretch-text (e.g., texts on grade level not reading level); however, identifying the appropriate instructional-level for students took priority over content of readings. Prior to reading the fluency-text students were prompted to think about their prior-knowledge of the topic, skim the passage, and ask for clarification of unknown words. Students read the text silently for 1–2 minutes and were asked to tell what the passage was about, similar to the stretch-text activity. Following a brief discussion, students engaged in repeated readings to build reading fluency. The routine progressed from teacher-modeled reading or choral reading routines to independent reading with or without a partner. After completing fluency activities, students completed a “Does it Make Sense?” activity, which required students to carefully read a sentence or group of sentences and consider its syntax and semantics to determine if it made sense. If a student determined a sentence did not make sense, he or she was asked to underline the word(s) in the sentence that disrupted meaning. Tutors provided feedback to students through discussion during the activity. On day 10, students re-read passages from the unit for additional practice and to check for understanding.

Word Study—The word study component of the intervention addressed phonics skills with multi-syllabic words that included opportunities to practice patterned word reading at the word, phrase, and sentence level. Based on the individual needs of students, different word lists were taught. These lists included multi-syllabic words, high-frequency sight words, and word patterns. Teachers would model word reading and students would have multiple opportunities to read and re-read the lists with feedback from the teacher and from other students. Tutors and students documented list mastery and progressed to more difficult lists as previous lists were mastered (i.e., read with automaticity). On day 10, students re-read word lists from the unit for additional practice and to check for accuracy. Assigned lists were consistently updated based on individual progress over the course of the 16-week intervention.

Progress Monitoring—Student progress was monitored throughout the researcher-implemented intervention using 4th grade level passages from AIMSWeb CBM-R (Shinn,

2002) and the 4th grade form of the Test of Silent Reading Efficiency and Comprehension (TOSREC; Wagner et al., 2010). Fluency was assessed monthly. Two passages were administered in alternating fashion to help control for form effects and minimize potential practice effects. We calculated individual student slopes after each administration using least squares regression and a report was generated for tutors and coaches to evaluate student growth in reading and adjust instruction if necessary. Benchmark growth rates were identified based on published ORF growth rates (Hasbrouck & Tindal, 2006) and data from previous intervention studies. At each measurement point, student growth in ORF was calculated and categorized as on target, at some risk, or at risk. For students at risk, tutors and coaches discussed ways to adapt instruction to improve student growth in reading.

Intervention Implementation—Nineteen tutors (18 female) were hired and trained by the research team to provide the intervention to students in groups of approximately 4–5 students for approximately 35 minutes per school day. All interventionists but one had at least an undergraduate degree, six interventionists had a master's degree, and one held a doctorate in education. Thirteen had a teaching certificate in reading, special education, or a related field.

The research team provided the interventionists with approximately 10 hours of professional development prior to teaching. This training included sessions related to implementation of the key elements of the intervention, strategies to promote active engagement, as well as other features of effective instruction and behavior management. They also received an additional 8 hours of professional development related to the intervention throughout the year and participated in biweekly staff development meetings with ongoing on-site feedback and on site coaching (approximately once every 2 to 3 weeks).

The average total amount of research intervention received for the students in researcher implemented intervention at the end of the year was 23.4 hours (*SD* = 17.6, range 0.0 to 42.0) at the near urban site, 24.4 hours (SD = 18.4, range 0.0 to 46.1) at the suburban site, and 26.5 hours (*SD* = 19.0, range 0.0 to 47.8) at the large urban site.

Comparison Group Reading Intervention—All students in the treatment and comparison conditions demonstrated significant reading problems. School personnel elected to provide treatment to the students in the comparison condition. We interviewed all classroom teachers and also asked all relevant educators (e.g., teachers, special education personnel) to complete an alternate reading inventory form to determine the amount and type of school-based reading intervention provided to students. The information was then categorized into different types of instruction: test preparation, basic word reading interventions, fluency interventions, inclusion support, and RTI/Resource instruction. Teachers reported that students received phonics and word-reading interventions, which employed the use of commercially published products including SRA Corrective Reading (Engelmann, 1988), SRA Reading Mastery (Engelmann & Bruner, 1995), and the Basic Language Skills program (Vickery, Reynolds, & Cochran, 1987), and the computer-based programs Read 180 (Scholastic, Inc., 2013) and Istation (Istation, 2011). Teachers reported students received fluency-based interventions with the use of the commercially published products Fast ForWord (Scientific Learning Corporation, 1997), Reading Plus (Rasinski,

Samuels, Hiebert, Petscher, & Feller, 2011), and Read Naturally (Hasbrouck, Ihnot, & Rogers, 1999). Students were also provided with test preparation instruction consisting of students being exposed to practice test items of the same type and receiving feedback from teachers shortly after answering the questions. For example students would read several short passages followed by main idea multiple-choice questions. Students continued to work on specific questions types with teacher feedback provided until there was evidence of mastery. Additional school-provided intervention was typically delivered by certified teachers in groups ranging in size from one to 15 students for 2 to 5 days per week in sessions of time ranging between 10 to 60 minutes per session. The proportion of students receiving additional instruction was similar across sites ($p > .05$).

Intervention Fidelity

Tutors in the researcher implemented intervention audio-recorded intervention lessons daily. A subset of audio-recorded lessons were then randomly selected by blocking on reading group and school within each tutor to identify a total of eight lessons per tutor to be coded for fidelity. Prior to independent fidelity coding, two members of the research team at each site randomly chose an audio recording and independently coded the lesson. Upon completion, the researchers met and discussed any discrepancy in scores. This process was repeated until comparison of code sheets reached agreement of 90% or higher in adherence to the gold standard method (Gwet, 2001). The researchers who established the gold standard then trained an additional five coders on use of the code sheet. Lessons were coded until scores of 90% or above were obtained for all researchers coding fidelity as a means of establishing coder reliability. To protect against rating drift, a second reliability check was conducted after each researcher had completed half of all assigned independent coding. Similar to the initial reliability check, all coders had to achieve 90% reliability to proceed. The mean reliability score between coders was 95%.

Fidelity was coded by rating each of the instructional components on a 4-point Likert-type rating scale ranging from 1 (low), 2 (mid-low), 3 (mid-high), and 4 (high). A score of 4 (high) was coded when the interventionist implemented all of the required elements and procedures. A score of 3 (mid-high) was coded when nearly all of the required elements and procedures were completed, and a score of 2 (mid-low) when some of the required elements and procedures were completed. A score of 1 was coded if less than half of the required elements and procedures were completed for a given component of the lesson. If a component was not expected during the lesson, a score of *N* was coded indicating that the component was not expected during that particular lesson and was not included in the fidelity score calculation. Global observations of quality and fidelity were also coded on 4-point Likert type rating scale ranging from 4 (highest quality), 3 (mid-high), 2 (mid-low) and 1 (lowest quality). Considerations for global observations of quality were pacing, active engagement, group management, clarity in instructions and feedback, and behavior management. Global fidelity was based on a holistic evaluation of intervention implementation and the degree of success in implementing all components as designed..

The mean implementation score across components and across interventionists was 3.71 (*SD* = .24, range 3.45 to 4.00). The mean global quality score across components and across

interventionists was 3.71 (*SD* = 0.50, range 2.00 to 4.00). The mean global fidelity ranking was 3.48 (*SD* = 0.55, range 2.00 to 3.82).

Additional Reading Intervention Inventory

In addition to collecting information about the fidelity of instruction for the researcher-provided intervention, we collected school-reported information on school-provided additional reading intervention (ARI) that students in both the treatment and comparison conditions may have received across the year. Data were collected in two waves. First, teachers completed a form indicating which study participants had received additional interventions in reading. The classroom teacher or instructor of record also completed a second form describing supplemental reading instruction for all students who had received ARI. This second form collected data on the number of intervention sessions, minutes per session, as well as qualitative information about the type of intervention. These data were utilized to calculate two additional variables for analysis: ARI (total school-provided reading intervention) and total reading intervention (ARI + researcher-provided intervention time). Students in the control condition did not receive any researcher-provided intervention. Thus, for the control condition ARI is equal to total supplemental reading intervention.

Results

Data Analysis

A one-way analysis of covariance (ANCOVA) was conducted for each outcome measure. In each analysis, the posttest served as the dependent variable and treatment condition served as the independent variable. To control experiment-wise α across the six outcome measures, statistical significance was tested against an adjusted α of .008. Standardized effect sizes for differences between pretest and posttest were calculated using model predicted posttest standard score means and observed posttest standard deviations (Bloom, Hill, Black, & Lipsey, 2008). In addition, standardized effect sizes were calculated for treatment effects using model-adjusted means. Based on the suggestion of the United States Department of Education's What Works Clearinghouse (2011), we examined effect sizes for all measures, regardless of statistical significance.

Preliminary Analyses

Preliminary analyses examined the need to include school effects in subsequent models. A random intercepts model was fit to each outcome measure with students as the first level and schools as the second level. ICCs calculated from the variance components of these models ranged from 0.00 to 0.05. As a result of these low values, we did not include school as a "level" in subsequent models. These tests were repeated for teacher and tutor level effects. ICCs calculated from the variance components of these models ranged from 0.00 – 0.08. Unconditional random intercept models were fit to evaluate the need to utilize multilevel models. The variance of the random intercepts did not differ statistically from zero for any of the outcomes when students were nested within school or tutor. There were significant differences from zero (p = 0.04) for the TOWRE and WJ3 Spelling tests when students were clustered within teacher. When pretest and treatment were added to the models the variance of the random intercept was no longer significantly different from zero for the TOWRE.

Given that only two of the eighteen models showed significant variation in random intercepts and that the pattern of results would not differ from a non-hierarchical model it was decided to exclude these random effects from the remainder of analyses.

The primary analyses relied on the assumptions of independence of covariates and treatment effect as well as homogeneity of regression slopes. The independence of covariates and treatment effect was confirmed using t-test ($p > .05$). We checked the assumption for homogeneity of regression slopes by testing interaction terms involving the treatment variable. The interaction terms were not statistically significant ($p > .05$) for each dependent variable.

Overview of Results

Table 2 shows the observed means and standard deviations for the pretest and posttest assessments by condition. Table 2 also shows the effect size associated with change from pretest to posttest. Based on posttest – pretest differences, the researcher implemented treatment condition outperformed the comparison condition on all measures except the comprehension measures. The results for group comparisons of the model-adjusted means from ANCOVA analyses are presented in Table 3 with effect sizes. This table shows the same pattern of results for the model-adjusted means.

ANCOVA Results

The same pattern of results was found for every outcome variable. Specifically, the pretest measure was a significant predictor of posttest performance (All measures: $p < .0001$) but the tests of treatment effect were not significant (WJ-III Letter Word ID: $F(1,400) = 3.31$, $p = 0.07$; WJ-III Spelling: $F(1,399) = 2.24$, $p = 0.14$; WJ-III Passage Comprehension: $F(1,400) = 0.03$, $p = 0.86$; Gates Comprehension: $F(1,405) = 0.69$, $p = 0.41$; TOWRE: $F(1,399) = 0.01$, $p = 0.92$; TOSREC: $F(1,414) = 0.32$, $p = 0.57$).

Additional Analyses

Two major questions arose from results of the primary analyses. First, there appeared to be strong normative growth for both control and treatment groups. For two of the measures, Gates MacGinitie and TOSREC, standardized effect sizes ranged from 0.67 – 0.98. Given that the growth in standard scores was not different across groups it seemed plausible that these effects were due to regression toward the mean. Second, the fact that no group differences were found despite the intensive intervention was perplexing. To better understand these null results, we conducted a series of analyses related to the fidelity and duration of reading interventions provided by both the research team and the school. For example, it is common for struggling readers such as those in this study to receive additional reading intervention (ARI). It was hypothesized that schools may have allocated ARI disproportionately to struggling readers who were not in the intervention group. A second possibility is that the existence of considerable school-based ARI may have obscured differences between the treatment and control conditions. Finally, we sought to understand whether variations in fidelity (treatment group only) or differences in the amount of time in intervention were related to differential treatment results.

In order to address these hypotheses, a series of analyses were performed. First, to address whether the observed standard score gains could plausibly be attributed to gains in proficiency throughout the year, we evaluated change in W scores on each WJ-III measures (Woodcock et al., 201). W scores represent an equal interval metric and permit an analysis of "expected" growth through a designated timeframe. We report W score improvements compared to expected W score improvements based on the WJ-III norming sample. Second, we investigated whether differences at pre and post could be attributed to regression to the mean. The original analyses were repeated but the observed values at pretest were replaced by estimated true scores in order to control for effects from regression to the mean (Cronbach, Gleser, Nanda, and Rajaratnam, 1972). The next set of analyses investigated the distribution of ARI across groups (treatment vs. control) and whether controlling for ARI revealed different results. We then evaluated whether treatment fidelity and treatment duration predicted differential outcomes. Given the exploratory nature of these analyses, statistical significance was evaluated with α set at 0.05.

Analysis of Proficiency Gains

The original analyses were replicated for the WJ-III measures using W scores and the same pattern of results was obtained for Spelling and Passage Comprehension. For Letter Word ID, treatment was significant ($F(1,400) = 5.06$, $p = 0.03$) as was pretest ($F(1,400) = 597.7$, $p < 0.0001$). Of particular interest was the gains made by each group from pretest to posttest. Table 4 displays the WJ-III W score version of the results shown in Table 2.

The expected gains in proficiency for the WJ-III tasks based on the normative sample is 12.1 for Letter Word ID, 8.4 for Spelling, and 5.3 for Passage Comprehension. The increase in proficiency shown by the treatment group equaled expected growth for Letter Word ID and slightly exceeded expectations for Passage Comprehension. The control group performed a little lower than expectations for Letter Word and also slightly exceeded expectations for Passage Comprehension. Figure 1 displays W score gains for the treatment and control conditions, as well as the WJ-III norming sample.

Regression toward the mean

Table 5 displays information similar to that found in Table 2. The difference between the two tables is that values in Table 5 are based on the estimated true scores for pretest values. Effect sizes stayed consistent with prior analyses. The results still show that the treatment group had better gains on all measures except the comprehension measures.

ANCOVA results with estimated true pretest scores were similar to the original analyses in that the same pattern of results was found for every outcome variable. For each measure the treatment by pretest interaction was not statistically significant ($p > 0.05$). As in the prior analyses, the pretest measure was a significant predictor of posttest performance (All measures: $p < .0001$) but the tests of treatment effect were not significant (WJ-III Letter Word ID: $F(1,400) = 3.19$, $p = 0.07$; WJ-III Spelling: $F(1,399) = 2.27$, $p = 0.13$; WJ-III Passage Comprehension: $F(1,400) = 0.03$, $p = 0.87$; Gates Comprehension: $F(1,405) = 0.71$, $p = 0.40$; TOWRE: $F(1,399) = 0.02$, $p = 0.89$; TOSRE: $F(1,415) = 0.25$, $p = 0.62$.

Additional Reading Intervention (ARI)

Table 6 displays the amount of ARI in minutes by school for each group. Only one school showed statistically significant differences in the amount of ARI between groups. School 1 provided more ARI to the control group. Overall, while not statistically significant, schools provided more ARI to control students than to treatment students as seen on the Total line. Initial ANCOVA models were fit for each of the outcome measures utilizing pretest score and ARI as covariates. There were no significant interactions involving the treatment effect and those terms were dropped from the models.

Letter-Word Identification—The test of treatment effect was not significant for WJ-III Letter Word ID ($F(1,3400) = 1.99$, $p = 0.16$) but the pretest ($F(1,400) = 482.4$, $p < 0.0001$) and ARI ($F(1400) = 7.7$, $p < 0.01$) were both significant. The regression coefficient for the pretest was positive while the ARI coefficient was negative. For the obtained values of the regression coefficients the negative ARI term suggests that increasing values of ARI result in lower outcomes.

Spelling—The test of treatment effect was not significant for WJ-III Spelling ($F(1,399) = 0.06$, $p < 0.81$) The effect of pretest ($F(1,399) = 909.8$, $p < 0.0001$) was significant but the effect of ARI was not significant ($F(1,101) = 2.93$, $p = 0.09$).

Comprehension—The test of the treatment effect was not significant for WJ-III Passage Comprehension ($F(1,393) = 0.84$, $p = 0.36$) or for Gates Comprehension ($F(1,405) = 0.25$, $p = 0.62$). The effect for pretest was significant for both measures ($p < 0.01$) but the effect for ARI was only significant for Gates Comprehension ($p < 0.01$). The effect of ARI on Gates Comprehension was negative but very small such that an additional 580 hours of ARI would be necessary to reduce the outcome by 15 points.

Fluency—The test of the treatment effect was not statistically significant for the TOWRE ($F(1,400) = 3.51$, $p = 0.06$). The effect of pretest was significant ($F(1,400) = 421.2$, $p < 0.0001$) as was the effect of ARI ($F(1,400) = 4.86$, $p = 0.03$).

Fidelity of Implementation

To test the effect of fidelity, new models were fit for each of the outcome measures with pretest and a measure of fidelity as predictors. The first set used a global rating of instruction and the next set used a global rating of implementation as the fidelity predictors. The control group did not receive any tutoring so there was no fidelity data available for that group. As a result these analyses were limited to the treatment group. In all cases, the pretest was significant ($p < 0.0001$) but the fidelity measure was not ($p > 0.05$).

Total Researcher-Provided Intervention Time

To test the effects of the total amount of researcher-provided intervention time, additional models were fit for each of the outcome measures with pretest and total researcher-provided intervention time as predictors. The control group did not receive any researcher-provided intervention, so these analyses were performed with the treatment group only. In all cases,

the pretest was significant ($p < .0001$), but the amount of researcher-provided intervention time was not significant ($p < .05$). All coefficients were positive.

Total Reading Intervention Time

The final set of models replaced the treatment variable with total reading time. This value was the sum of the time spent in the experimental treatment and the amount of time indicated by the ARI. A model was fit for each outcome measure with total reading time and pretest as predictors. The pretest was significant ($p < 0.0001$) for all outcome measures while the effect of total reading time was not significant for any outcomes. For each outcome the regression coefficient for total reading time was positive, ranging from 0.0001 to 0.0007. For these values total reading time required to increase an outcome by 15 points would be extreme. Even the largest regression coefficient for TOWRE of 0.0007 would require 377 hours to raise the outcome score by 1 standard deviation, or 15 points.

Discussion

This study examined the efficacy of a fourth-grade reading intervention for students with significant reading difficulties (i.e., more than one standard deviation below the mean on a reading comprehension measure). Students were randomly assigned to either a researcher implemented reading treatment or a school provided reading treatment (2 to treatment for each 1 to comparison condition). Findings revealed that students in the researcher-implemented treatment did not differ significantly from students in the school-implemented condition. Examining growth in standard scores as well as W scores from all WJ-III measures (WJ-III; Woodcock, McGrew, & Mather, 2001) both groups made gains in reading outcomes. To illustrate, students' standard score growth from pretest to posttest was 3 standard score points on decoding, 5 on fluency and between 2–7 on reading comprehension. Evaluating main effects with W scores for all WJ-III measures (equal interval metrics are not available for the other measures), we examined proficiency in reference to age or grade. These analyses demonstrated that gains in proficiency exceeded or matched expected gains in proficiency based on the WJ-III norming sample in reading comprehension and decoding. Considering the very low reading levels of students in this study (reading comprehension pretest mean standard scores between 77–81), this provides additional support for claims of strong proficiency growth in both treatment and comparison conditions. These gains are substantial when compared to standard score gains from previous interventions with students in upper elementary grades (Wanzek et al., 2013) and recent studies with this grade group (Ritchey et al., 2012).

The findings from this study revealed that regardless of whether students were in the researcher provided treatment or the school provided treatment, they made progress between the beginning of 4th grade and end of fourth grade. These findings contrast with previous intervention studies with middle school students in which standard score growth was not realized in a year-long intervention (Vaughn et al., 2010; Vaughn et al., 2012). Wanzek and Roberts (2012) conducted a similar study with 4th graders assigned to one of 3 treatment conditions (e.g., comprehension, word reading, individualized) or a control condition.

Students in the treatment conditions did not score significantly higher than students on the comparison condition on outcome measures in reading.

While much is known about effective instruction to assist young students' transition from nonreaders to beginning readers, less is known about how to effectively remediate struggling readers after second grade. The challenge of improving students' language and background knowledge so that they can adequately understand increasingly complex texts is likely to require instruction that goes well beyond traditional reading interventions (Compton, Miller, Elleman, & Steacy, 2014; Vaughn & Fletcher, 2012).

The effect sizes reported in this study are in line with the more recent syntheses of interventions with students in 4th grade and above (Scammacca et al., 2013; Wanzek et al., 2013) suggesting that small effects are likely to result from extensive interventions for students with significant reading problems. There are several other considerations when interpreting the findings from this study. First, after students were identified with significant reading comprehension problems and were randomized to treatment and comparison conditions, the schools decided to provide their own interventions to students in the comparison condition. Since students were so far behind, it was unethical to ask them not to provide the intervention particularly since the study was scheduled throughout the 4th grade year. Second, the sample size in this study is larger than most intervention studies with struggling readers; larger sample sizes are associated with smaller effect sizes. Third, the intervention was provided for an extended period of time (8 months) and findings from previous research suggest that interventions of shorter duration report higher effects than interventions of longer duration (Elbaum, Vaughn, Hughes, & Moody, 2000; Scammacca et al., 2013), perhaps due to an initial boost in learning from the addition of instruction or even the novelty of the intervention. Based on findings from efficacy studies reported by the Institute for Education Sciences (i.e., Kemple et al., 2008), large-scale, long, and school-based interventions are likely to be associated with lower effects.

Most students in this study were able to decode single syllable words at a basic level but many lacked word-reading skills for more complex word types (i.e., multi-syllable words), but the efficiency with which they read and comprehended text varied widely. Using the simple view of reading (Hoover & Gough, 1990) as a theoretical model to guide our instruction, we focused both on building students' word reading skills by teaching multi-syllable word reading as well as focusing on automaticity through mastery of high frequency words. Additionally, we wanted to increase students language development broadly within the context of increasingly complex texts selected from topics that related to history and social studies. We don't know whether effects would have been different had we focused more on narrative text, however, our rationale for the texts types we chose was to further build background knowledge and vocabulary.

Implications

From this study, we can form hypotheses about more effective ways to remediate upper elementary students with reading difficulties. It may be necessary to provide even more intensive intervention for some students (e.g., longer time, smaller groups, intervention even more specifically focused to meet students' needs). For example, it is possible to consider

ways to integrate text reading throughout the school day and thus relying less on the reading intervention time to compensate for the vocabulary, background knowledge, and text-based challenges these students demonstrate. We also think it may be important to consider ways to provide individualized treatments to better align with the individual learning needs of students. This might be accomplished with even smaller groups (e.g., one-on-one, pairs), but these approaches may be less feasible for schools to implement. We are also examining ways to increase impact by considering mechanisms for improving attention and self-regulation within the reading intervention treatment.

Limitations

As might be expected in any school-based intervention study, there were several limitations to consider. The most significant limitation was the challenge with contrasting the researcher-provided treatment with a business as usual condition. Schools elected to provide treatment to comparison students requiring us to adjust our contrast as between a researcher-provided treatment and school-provided treatment. We have confidence in the amount of treatment provided to students in the researcher-provided treatment because we documented daily attendance and the amount of time in treatment. We have considerably less confidence in the teacher-reported data on amount of treatment. The study was also limited by the amount of time and access the schools provided the researchers to instruct students. Ideally, daily sessions of closer to 50 minutes per day would have been provided. In summary, while the treatments were minimally different from each other on reading outcome measures, growth for students in both treatment conditions (as determined by standard score increases) was strong and aligned with promising findings for improving reading outcomes for students in 4th grade with significant reading comprehension problems.

Acknowledgments

Funding

This research was supported by grant P50 HD052117 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Eunice Kennedy Shriver National Institute of Child Health and Human Development or the National Institutes of Health.

References

Baumann JF, Edwards EC, Boland EM, Olejnik S, Kame'enui EJ. Vocabulary tricks: Effects of instruction in morphology and context on fifth-grade students' ability to derive and infer word meanings. American Educational Research Journal. 2003; 40(2):447–494.

Beck, IL., McKeown, MG., Kucan, L. Bringing words to life: Robust vocabulary instruction. Guilford Press; 2013.

Bloom HS, Hill CJ, Black AB, Lipsey MW. Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. Journal of Research on Educational Effectiveness. 2008; 1(4):289–328. DOI: 10.1080/19345740802400072

Catts HW, Adlof SM, Hogan T, Weismer SE. Are specific language impairment and dyslexia distinct disorders? Journal of Speech, Language, and Hearing Research. 2005; 48(6):1378.doi: 10.1044/1092-4388(2005/096)

Chall JS, Jacobs VA. Writing and reading in the elementary grades: Developmental trends among low SES children. Language Arts. 1983; 60(5):617–626.

Cirino PT, Romain MA, Barth AE, Tolar TD, Fletcher JM, Vaughn S. Reading skill components and impairments in middle school struggling readers. Reading and Writing. 2013; 26(7):1059–1086. DOI: 10.1007/s11145-012-9406-3 [PubMed: 24000271]

Compton DL, Fuchs D, Fuchs LS, Elleman AM, Gilbert JK. Tracking children who fly below the radar: Latent transition modeling of students with late-emerging reading disability. Learning and Individual Differences. 2008; 18(3):329–337. DOI: 10.1016/j.lindif.2008.04.003

Compton DL, Miller AC, Elleman AM, Steacy LM. Have we forsaken reading theory in the name of "quick fix" interventions for children with reading disability? Scientific Studies of Reading. 2014; 18(1):55–73. DOI: 10.1080/10888438.2013.836200

Cronbach, LJ., Gleser, GC., Nanda, H., Rajaratnam, N. The dependability of behavioral measurements: Theory of generalizability for scores and profiles. New York: Wiley; 1972.

Elbaum B, Vaughn S, Hughes MT, Moody SW. 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. 2000; 92(4):605–619. DOI: 10.1037/0022-663.92.4.602

Elleman AM, Lindo EJ, Morphy P, Compton DL. The impact of vocabulary instruction on passage-level comprehension of school-age children: A meta-analysis. Journal of Research on Educational Effectiveness. 2009; 2(1):1–44. DOI: 10.1080/19345740802539200

Engelmann, S. Corrective reading: Decoding strategies. SRA Macmillan/McGraw-Hill; 1988.

Engelmann, S., Bruner, E. Reading mastery I/II: fast cycle. SRA Macmillian/McGraw-Hill; 1995.

Gough P, Tunmer W. Decoding, reading, and reading disability. Remedial and Special Education. 1986; 7:6–10. DOI: 10.1177/074193258600700104

Gwet, K. Handbook of inter-rater reliability: How to estimate the level of agreement between two or multiple raters. Gaithersburg, MD: STATAXIS Publishing Company; 2001.

Hasbrouck JE, Ihnot C, Rogers GH. "Read naturally": A strategy to increase oral reading fluency. Literacy Research and Instruction. 1999; 39(1):27–37.

Hasbrouck J, Tindal GA. Oral reading fluency norms: A valuable assessment tool for reading teachers. The Reading Teacher. 2006; 59(7):636–644. DOI: 10.1598/RT.59.7.3

Hiebert, EH. QuickReads – A research-based fluency program. Parsippany, NJ: Pearson; 2003.

Hoover WA, Gough PB. The simple view of reading. Reading and Writing. 1990; 2(2):127–160. DOI: 10.1007/BF00401799

Istation. Computer software. Dallas, TX: Istation; 2011.

Johnston TC, Kirby JR. The contribution of naming speed to the simple view of reading. Reading and Writing. 2006; 19(4):339–361. DOI: 10.1007/s11145-005-4644-2

Joshi RM, Aaron PG. The component model of reading: Simple view of reading made a little more complex. Reading Psychology. 2000; 21(2):85–97.

Kamil, ML., Borman, GD., Dole, J., Kral, CC., Salinger, T., Torgesen, J. Improving adolescent literacy: Effective classroom and intervention practices: A practice guide. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education; 2008. NCEE#2008-4027Retrieved from http://ies.ed.gov/ncee/wwc

Kemple, JJ., Corrin, W., Nelson, E., Salinger, T., Herrmann, S., Drummond, K. The enhanced reading opportunities study: Early impact and implementation findings. Washington, DC: U.S. Department of Education, Institute of Education Sciences: National Center for Education Evaluation and Regional Assistance; 2008.

MacGinitie, WH., MacGinitie, RK., Maria, K., Dreyer, LG., Hughes, KE. Gates-MacGinitie reading tests. 4. Iasca: Riverside; 2000.

Perfetti, CA. Dyslexia across languages: Orthography and the brain-gene-behavior link. Baltimore: Brookes Publishing; 2011. Reading processes and reading problems: Progress toward a universal reading science.

Rasinski TV, Samuels SJ, Hiebert E, Petscher Y, Feller K. The effects of silent reading fluency instructional protocol on students' reading comprehension and achievement in an urban school setting. Reading Psychology. 2011; 32(1):75–97. DOI: 10.1080/02702710903346873 [PubMed: 26347186]

Ritchey KD, Silverman RD, Montanaro EA, Speece DL, Schatschneider C. Effects of a tier 2 supplemental reading intervention for at risk fourth-grade students. Exceptional Children. 2012; 78(3):318–334. [PubMed: 22685347]

Savage R. Reading comprehension is not always the product of nonsense word decoding and linguistic comprehension: Evidence from teenagers who are extremely poor readers. Scientific Studies of Reading. 2006; 10(2):143–164. DOI: 10.1207/s1532799xssr1002_2

Scammacca, N., Roberts, G., Vaughn, S., Edmonds, M., Wexler, J., Reutebuch, CK., Torgesen, J. Intervention for adolescent struggling readers: A meta-analysis with implication for practice. Portsmouth, NH: RMC Research Corporation, Center on Instruction; 2007.

Scammacca NK, Roberts G, Vaughn S, Stuebing KK. A meta-analysis of interventions for struggling readers in Grades 4–12: 1980–2011. Journal of Learning Disabilities. 2013; Advance online publication. doi: 10.1177/0022219413504995

Scholastic, Inc. Read 180. New York: Scholastic; 2013.

Scientific Learning Corporation. Fast ForWord [Computer software]. Oakland, CA: Author; 1997.

Shinn, MR. AIMSweb training workbook: Progress monitoring strategies for writing individualized goals in general curriculum and more frequent formative evaluation. 2002. Retrieved August, 2014 from: http://www.cnyric.org/tfiles/folder1052/Progress%20Monitoring%20Guide.pdf

Swanson, HL., Hoskyn, M., Lee, C. Interventions for students with learning disabilities. New York: Guilford Publishing; 1999.

Therrien WJ, Wickstrom K, Jones K. Effects of a combined repeated reading and question generation intervention on reading achievement. Learning Disabilities Research & Practice. 2006; 21(2):89–97. DOI: 10.111/j.1540-5826.2006.00209.x

Torgesen, JK., Wagner, RK., Rashotte, CA. Test of word reading efficiency. San Antonio, TX: PRO-ED; 1999.

Vaughn S, Cirino PT, Wanzek J, Wexler J, Fletcher JM, Denton CD, … Francis DJ. Response to intervention for middle school students with reading difficulties: Effects of a primary and secondary intervention. School Psychology Review. 2010; 39(1):3–21. DOI: 10.1002/pits.20481 [PubMed: 21479079]

Vaughn S, Fletcher JM. Response to intervention with secondary students with reading difficulties. Journal of Learning Disabilities. 2012; 45(3):241–253. DOI: 10.1177/0022219412442157

Vaughn S, Roberts G, Wexler J, Vaughn MG, Fall A-M, Schnakenberg JB. High school students with reading comprehension difficulties: Results of a randomized control trial of a two-year reading intervention. Journal of Learning Disabilities. 2014; Advanced Online Publication. doi: 10.1177/0022219413515511

Vaughn S, Wexler J, Leroux A, Roberts G, Denton C, Barth A, Fletcher J. Effects of intensive reading intervention for eighth-grade students with persistently inadequate response to intervention. Journal of Learning Disabilities. 2012; 45(6):515–525. DOI: 10.1177/0022219411402692 [PubMed: 21512102]

Vickery KS, Reynolds VA, Cochran SW. Multisensory teaching approach for reading, spelling, and handwriting, Orton-Gillingham based curriculum, in a public school setting. Annals of Dyslexia. 1987; 37(1):189–200. DOI: 10.1007/BF02648066 [PubMed: 24234994]

Wagner, RK., Torgesen, JK., Rashotte, CA., Pearson, NA. Test of sentence reading efficiency and comprehension (TOSREC). Austin, TX: PRO-ED; 2010.

Wanzek J, Roberts G. Reading interventions with varying instructional emphases for Fourth Graders With Reading Difficulties. Learning Disability Quarterly. 2012; 35(2):90–101.

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

Wanzek J, Vaughn S, Scammacca N, Metz K, Murray C, Roberts G, Danielson L. Extensive reading interventions for older struggling readers: Implications from research. Review of Educational Research. 2013; 83:163–195. DOI: 10.3102/0034654313477212

Wanzek J, Wexler J, Vaughn S, Ciullo S. Reading interventions for struggling readers in the upper elementary grades: A synthesis of 20 years of research. Reading and writing. 2010; 23(8):889–912. DOI: 10.1007/s11145-009-9179-5 [PubMed: 21072128]

What Works Clearinghouse. WWC procedures and standards handbook, version 2.1. 2011. Retrieved February 25, 2014 from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v2_1_standards_handbook.pdf

Woodcock, RW., McGrew, KS., Mather, N. Woodcock-Johnson III tests of achievement. Itasca, IL: Riverside; 2001.

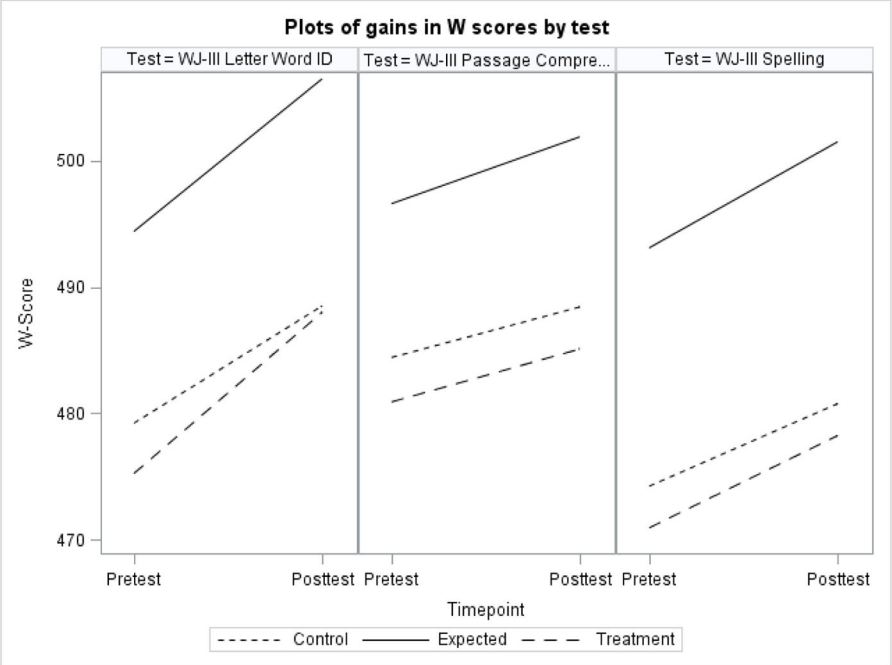

Figure 1.

W score gains by test for treatment and control groups, as well as the WJ-III norming sample

Table 1

Intervention scope and sequence

Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10
Vocabulary words 1 and 2 (10 min)	Vocabulary review (5 min)	Vocabulary review (5 min)	Vocabulary words 3 and 4 (10 min)	Vocabulary review (5 min)	Vocabulary review (5 min)	Vocabulary words 5 and 6 (10 min)	Vocabulary review (5 min)	Vocabulary review (5 min)	CBM (20 min)
Fluency-text (15 min)	Fluency-text (15 min)	Fluency-text (20 min)	Stretch-text (15 min)	Stretch text (20 min)	Stretch text (20 min)	Stretch-text (15 min)	Fluency-text (15 min)	Fluency-text (20 min)	IR (15 min)
Word study (10 min)	Word study (15 min)	Word study (10 min)	Word study (10 min)	Word study (10 min)	Word study (10 min)	Word study (10 min)	Word study (15 min)	Word study (10 min)	

Note. CBM = Curriculum-based Measure; IR = Independent Reading

Table 2

Pretest and posttest means, standard deviations, and effect sizes

				Pre		Post		
Measure	***Construct***	***Group***	***n***	**Mean**	**SD**	**Mean**	**SD**	***d (CI)***
WJ-III Letter Word ID	Decoding	Treatment	296	89.16	11.25	92.34	10.40	0.29 (0.13, 0.46)
		Control	149	90.40	11.01	92.46	10.78	0.19 (−0.04, 0.42)
WJ-III Spelling	Spelling	Treatment	292	86.40	10.27	86.79	13.77	0.03 (−0.13, 0.20)
		Control	148	88.11	10.56	88.85	12.69	0.06 (−0.17, 0.30)
WJ-III Passage Comprehension	Comprehension	Treatment	296	81.37	8.89	83.37	9.09	0.22 (0.06, 0.39)
		Control	149	82.77	8.87	84.81	8.67	0.23 (0.00, 0.47)
Gates McGinitie Passage Comprehension	Comprehension	Treatment	323	77.25	6.09	84.07	8.00	0.97 (0.80, 1.14)
		Control	161	77.14	6.17	84.53	8.92	0.98 (0.74, 1.22)
TOWRE Sight Word Efficiency	Fluency	Treatment	295	79.88	12.19	85.13	12.36	0.43 (0.26, 0.59)
		Control	150	80.95	12.07	84.46	12.09	0.29 (0.06, 0.52)
TOSRE	Fluency and Comprehension	Treatment	321	10.80	5.38	15.03	7.21	0.67 (0.51, 0.83)
		Control	157	11.16	5.35	15.99	6.38	0.83 (0.59, 1.06)

Note. WJ-III = Woodcock-Johnson III Tests of Achievement (WJ-III; Woodcock, McGrew, & Mather, 2001). TOWRE = Test of Word Reading Efficiency (Torgesen, Wagner, & Rashotte, 1999). TOSREC = Test of Sentence Reading Efficiency and Comprehension (Wagner, Torgesen, Rashotte, & Pearson, 2010)

Note. Effect sizes are posttest – pretest.

Table 3

Posttest group comparisons and effect sizes calculated using model-predicted means

Measure	Construct	Comparison	t	p	ES (CI)
WJ-III Letter Word ID	Decoding	Treatment - Control	1.80	0.07	0.11 (−0.10, 0.32)
WJ-III Spelling	Spelling	Treatment - Control	0.23	0.82	0.01 (−0.19, 0.22)
WJ-III Passage Comprehension	Comprehension	Treatment - Control	−0.85	0.39	−0.06 (−0.26, 0.15)
Gates McGinitie Passage Comprehension	Comprehension	Treatment - Control	−0.44	0.66	−0.04 (−0.25, 0.16)
TOWRE Sight Word Efficiency	Fluency	Treatment - Control	2.03	0.04	0.13 (−0.08, 0.34)
TOSRE	Fluency and Comprehension	Treatment - Control	−1.11	0.27	−0.10 (−0.30, 0.10)

Table 4

W score pretest and posttest means, standard deviations, and effect sizes

				Pre		Post		
Measure	***Construct***	***Group***	***n***	**Mean**	**SD**	**Mean**	**SD**	***d (CI)***
WJ-III Letter Word ID	Decoding	Treatment	296	475.2	22.57	488.0	19.77	0.60 (0.43, 0.77)
		Control	149	479.2	21.63	488.5	20.70	0.44 (0.20, 0.67)
WJ-III Spelling	Spelling	Treatment	292	480.9	14.91	485.1	19.91	0.24 (0.07, 0.41)
		Control	148	484.4	14.64	488.4	17.58	0.25 (0.02, 0.48)
WJ-III Passage Comprehension	Comprehension	Treatment	296	470.9	12.77	478.2	12.44	0.58 (0.41, 0.75)
		Control	149	474.2	12.04	480.8	11.41	0.56 (0.32, 0.79)

Note. WJ-III = Woodcock-Johnson III Tests of Achievement (WJ-III; Woodcock, McGrew, & Mather, 2001

Note. Effect sizes are posttest – pretest.

Table 5

Estimated true score pretest and observed posttest means, standard deviations, and effect sizes

				Estimated True Pre		Observed Post		
Measure	***Construct***	***Group***	***n***	**Mean**	**SD**	**Mean**	**SD**	***d (CI)***
WJ-III Letter Word ID	Decoding	Treatment	296	90.89	9.45	92.34	10.40	0.15 (−0.02, 0.31)
		Control	149	91.93	9.24	92.46	10.78	0.05 (−0.18, 0.29)
WJ-III Spelling	Spelling	Treatment	292	88.57	8.62	86.79	13.77	−0.16 (−0.32, 0.01)
		Control	148	90.02	8.87	88.85	12.69	−0.11 (−0.34, 0.13)
WJ-III Passage Comprehension	Comprehension	Treatment	296	84.91	7.20	83.37	9.09	−0.19 (−0.35, −0.02)
		Control	149	86.04	7.19	84.81	8.67	−0.16 (−0.39, 0.08)
Gates McGinitie Passage Comprehension	Comprehension	Treatment	323	80.89	5.11	84.07	8.00	0.48 (0.32, 0.65)
		Control	161	80.80	5.18	84.53	8.92	0.52 (0.29, 0.75)
TOWRE Sight Word Efficiency	Fluency	Treatment	295	81.90	10.97	85.13	12.36	0.28 (0.11, 0.44)
		Control	150	82.86	10.86	84.46	12.09	0.14 (−0.09, 0.37)
TOSRE	Fluency and Comprehension	Treatment	321	11.04	4.52	15.03	7.21	0.67 (0.51, 0.84)
		Control	157	11.34	4.50	15.99	6.38	0.85 (0.61, 1.09)

Note. WJ-III = Woodcock-Johnson III Tests of Achievement (WJ-III; Woodcock, McGrew, & Mather, 2001). TOWRE = Test of Word Reading Efficiency (Torgesen, Wagner, & Rashotte, 1999). TOSREC = Test of Sentence Reading Efficiency and Comprehension (Wagner, Torgesen, Rashotte, & Pearson, 2010)

Note. Effect sizes are posttest – pretest.

Table 6

School-provided reading instruction, researcher-provided intervention, and total intervention time by school and condition.

Site	Group	*n*	Additional Reading Intervention *M (SD)*	Researcher-provided Intervention *M (SD)*	Total intervention time *M (SD)*
Near urban district					
School 1	Control	7	93.2 (74.7)	0	93.2 (74.7)
	Treatment	8	10.3 (29.2)	24.6 (6.5)	34.9 (32.1)
School 2	Control	17	21.4 (41.2)	0	21.4 (41.2)
	Treatment	34	14.7 (24.6)	35.5 (4.5)	50.2 (26.1)
School 3	Control	3	40 (40.5)	0	40 (40.5)
	Treatment	3	93	33.1 (3.4)	126.1 (3.4)
School 4	Control	13	32.8 (41.8)	0	32.8 (41.8)
	Treatment	28	25.3 (45.2)	37.8 (3.7)	63.1 (44.9)
School 5	Control	1	125	0	125 (0)
	Treatment	5	23 (40.6)	39.2 (1.5)	62.2 (41.1)
School 6	Control	5	8.4 (13.1)	0	8.4 (13.1)
	Treatment	10	73.5 (136.1)	38 (2.6)	111.5 (135)
Suburban district				10	20.4 (32.5)
School 1	Control	21	39.4 (69.5)	36.3 (7.1)	75.7 (71.8)
	Treatment	10	27.2 (30.4)	0	27.2 (30.4)
School 2	Control	17	35.9 (56.3)	38.4 (6.7)	74.3 (57.3)
	Treatment	5	83.6 (34.5)	0	83.6 (34.5)
School 3	Control	12	45.8 (39.6)	34.8 (8.3)	80.5 (43)
	Treatment	13	0.8 (2.8)	0	0.8 (2.8)
Urban district			26	5.7 (11.8)	43.1 (3.5)
School 1	Control	7	27.5 (26.1)	0	27.5 (26.1)
	Treatment	16	36 (22.2)	34.4 (4.7)	70.4 (23.2)
School 2	Control	5	45.8 (25.6)	0	45.8 (25.6)
	Treatment	11	30.7 (29.4)	38 (12.9)	68.7 (28.8)
School 3	Control	4	0	0	0
	Treatment	12	4.4 (10.4)	35.7 (7.4)	40.2 (13.4)
School 4	Control	13	6.6 (7.7)	0	6.6 (7.7)

Site	Group	*n*	Additional Reading Intervention *M (SD)*	Researcher-provided Intervention *M (SD)*	Total intervention time *M (SD)*
School 5	Treatment	25	7.4 (8)	39.7 (3.5)	47.1 (7.3)
	Control	14	5.7 (11.6)	0	5.7 (11.6)
School 6	Treatment	28	2.6 (6.6)	36.5 (10.4)	39.1 (11.1)
	Control	5	21 (8.2)	0	21 (8.2)
School 7	Treatment	13	18.3 (20.9)	38.2 (12.7)	56.5 (18.7)
	Control	5	0	0	0
School 8	Treatment	15	6 (15.8)	38.8 (2.5)	44.8 (16.4)
	Control	7	93.2 (74.7)	0	93.2 (74.7)
	Treatment	8	10.3 (29.2)	24.6 (6.5)	34.9 (32.1)

Note. All times reported in hours. Additional reading intervention (ARI) data were collected through teacher report for all students in treatment and control conditions. ARI reported values represent average ARI by school and treatment group. Rearcher-provided intervention time data were collected through attendance logs maintained by intervention tutors. Total intervention time values represent the sum of ARI + researcher-provided intervention time. For students assigned to the control condition, researcher-provided intervention time = 0 and ARI = total intervention time.