Published in final edited form as:

Remedial Spec Educ. 2018 May ; 39(3): 131–143. doi:10.1177/0741932517750818.

Examining the Effects of Afterschool Reading Interventions for Upper Elementary Struggling Readers

Garrett J. Roberts, PhD¹, Philip Capin, MA², Greg Roberts, PhD², Jeremy Miciak, PhD³, Jamie M. Quinn, PhD², and Sharon Vaughn, PhD²

¹University of Denver, CO, USA

²The University of Texas at Austin, USA

³University of Houston, TX, USA

Abstract

We examined the efficacy of an afterschool multicomponent reading intervention for third-through fifth-grade students with reading difficulties. A total of 419 students were identified for participation based on a 90 standard score or below on a screening measure of the Test of Silent Reading Efficiency and Comprehension. Participating students were randomly assigned to a business as usual comparison condition or one of two reading treatments. All treatment students received 30 min of computer-based instruction plus 30 min of small-group tutoring for four to five times per week. No statistically significant reading comprehension posttest group differences were identified ($p > .05$). The limitations of this study included high attrition and absenteeism. These findings extend those from a small sample of experimental studies examining afterschool reading interventions and provide initial evidence that more instruction, after school, may not yield the desired outcome of improved comprehension.

Keywords

learning disabilities; reading; instruction; literacy; intervention research

The prevalence of afterschool programs for students has risen substantially in the last two decades. More than 10 million K-12 students participate in afterschool programs, due in large part to annual investments exceeding \$1 billion of federal funding for the U.S. Department of Education's 21st Century Community Learning Center Program (Afterschool Alliance, 2014). The number of students participating in these programs may continue to grow given their popularity with working parents who value afterschool supervision for their children. The upsurge in afterschool programming coincides with a change in the focus of after-school programs. Many afterschool programs were initially developed to support nonacademic domains (e.g., social skills, performing arts, and athletics). However, the emphasis of many federally funded programs has shifted toward providing additional

Corresponding Author: Garrett J. Roberts, Department of Teaching and Learning Sciences, University of Denver, Morgridge College of Education, 1999 E. Evans Ave., Denver, CO 80210-1700, USA, garrett.roberts@du.com.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

academic opportunities in an effort to narrow the achievement gap (e.g., 21st Century Community Learning). Despite the increased investment and shift in goals of afterschool programs, there are few high-quality examinations of the efficacy of these programs' ability to target academic outcomes, especially for students with learning difficulties.

Five systematic reviews have summarized the effects of afterschool programs on academic outcomes (Durlak & Weissberg, 2007; Kane, 2004; Lauer et al., 2006; Scott-Little, Hamann, & Jurs, 2002; Zief, Lauver, & Maynard, 2006). Although promising findings were noted in all five syntheses, most authors found the need for additional high-quality evaluations to draw sound conclusions. In accordance with these findings, Apsler (2009) and Kremer, Maynard, Polanin, Vaughn, and Sarteschi (2015) documented numerous methodological flaws to further complicate the literature. The largest evaluation of afterschool programs in the United States examined the 21st Century Community Learning Centers and found that students participating in the program showed, on average, no improvement in academic achievement (James-Burdumy et al., 2005). However, several other researchers questioned the value of these findings citing methodological problems in the study (e.g., Bissell et al., 2003; Kane, 2004; Mahoney & Zigler, 2003). To address the large number of studies using nonexperimental designs (e.g., studies without comparison groups, studies with nonequivalent comparison groups), Zief et al. (2006) conducted a synthesis of "well-implemented experimental studies." Consistent with the conclusions of Apsler (2009) and Kremer et al. (2015), only five studies conducted over the previous 20 years met the inclusion criteria for a well-implemented experimental study. Results showed that afterschool programs demonstrated a small, nonsignificant effect on school grades and no positive effect on reading achievement test scores. Of interest, Zief et al. (2006) and other reviews (Kane, 2004; Lauer et al., 2006) noted that the amount and regularity of students' attendance influenced program outcomes.

In the present randomized control trial (RCT), we examine the effects of an afterschool reading program focusing primarily on improving reading comprehension with upper elementary students specifically selected for their reading difficulties. The impetus to study the effects of afterschool reading interventions was twofold. First, although after-school programs are increasingly framed to positively influence academic outcomes for at-risk populations (e.g., Lauer et al., 2006), few high-quality studies have examined the effects of academic afterschool interventions, with even fewer studies targeting at-risk readers. Second, in light of the lack of robust gains resulting from previous upper elementary interventions implemented during the school day (Wanzek, Wexler, Vaughn, & Ciullo, 2010), we sought to test the effects of increasing the amount of reading instruction for struggling upper elementary readers by supplementing the core reading instruction and other regular school day reading interventions with an intensive after-school intervention.

Reading Interventions for Upper Elementary Struggling Readers

Educational researchers have accrued substantial knowledge about reading interventions for readers in the primary grades (K-2; Wanzek & Vaughn, 2007; Wanzek, Vaughn et al., 2016), yet considerably less is known about improving reading interventions in the upper elementary grades (Flynn, Zheng, & Swanson, 2012; Scammacca, Roberts, Vaughn, &

Stuebing, 2015; Wanzek et al., 2010). In a synthesis of reading interventions conducted during the school day for students in Grades 4 and 5 with reading difficulties, Wanzek et al. (2010) identified 24 reading intervention studies and found interventions targeting reading comprehension were associated with moderate to large effects on comprehension measures. However, there were limitations to this corpus of studies. For one, only nine experimental and four quasi-experimental studies were identified. Of the studies using experimental designs, only two utilized norm-referenced measures of reading comprehension (O'Connor et al., 2002; Therrien, Wickstrom, & Jones, 2006). Findings from these two studies are mixed. O'Connor and colleagues (2002) found treatment students made significant increases in reading comprehension when provided a multicomponent intervention. In the Therrien and colleagues (2006) study, they reported students made significant gains in reading fluency, but not reading comprehension after receiving reading and question generation instruction.

Since Wanzek and colleagues (2010) published their synthesis, four reading comprehension studies with upper elementary grade struggling readers were conducted featuring high-quality design elements and standardized reading comprehension outcome measures (Ritchey, Silverman, Montanaro, Speece, & Schatschneider, 2012; Vaughn, Vaughn, Solis, Miciak, Taylor, & Fletcher, 2016; Wanzek, Petscher, et al., 2016; Wanzek & Roberts, 2012). None of the four studies found significant differences between treatment and control conditions on norm-referenced measures of reading comprehension. However, in one study (Wanzek, Petscher, et al., 2016), the authors reported nontrivial effect sizes favoring students in the treatment condition on a standardized measure of reading comprehension (effect size = .14 and .28). On average, students in the treatment condition were provided a multicomponent reading intervention for 45 hr. Considering the challenges of significantly affecting reading comprehension of students in upper elementary grades with significant reading problems, we aimed to conduct an efficacy trial that would permit the investigation of supplemental time on reading instruction by investigating the effects of a reading intervention provided after school hours.

Text-Processing Instruction

As an alternative to reading comprehension strategy instruction, researchers have explored the effects of text-processing approaches (sometimes referred to as content approaches; for example, Barth et al., 2016; Beck & McKeown, 2006; Vaughn et al., 2013). Teachers utilizing text-processing approaches guide students to work through text to build a coherent understanding of the text. Unlike strategy instruction, text-processing approaches do not focus on teaching students to make use of specific routines before, during, and after reading (e.g., making predictions); instead, thinking about and discussing text is the central pathway for developing reading comprehension. Text-processing instruction is grounded in theories of reading comprehension, which focus on explaining how students generate meaning from text while reading (e.g., Kintsch, 1988). For instance, Kintsch (1988) posits that effective readers gather information from text and then integrate this new information into ideas that are then linked with prior learning to develop a coherent mental representation of text. Researchers have operationalized text-processing approaches through instructional programs focused on questioning the author (Beck & McKeown, 2006), engaging in classwide

discussions about text (Applebee, Langer, Nystrand, & Gamoran, 2003), and utilizing classwide discussion and team-based learning to process texts (Vaughn et al., 2013).

Recent investigations of the efficacy of text-processing practices provide encouraging results on researcher-developed measures; however, generalizations based on these findings are constrained by the small number of studies. McKeown, Beck, and Blake (2009) reported a text-processing approach producing larger gains than strategy instruction on researcher-developed measures of reading comprehension for fifth-grade students in general education classes. In a study with eighth-grade students in social studies classes, Vaughn et al. (2013) implemented a text-processing approach to teaching reading comprehension and social studies content. Vaughn et al. reported significant gains on measures of content acquisition and reading comprehension although significant differences were only found on the content acquisition measure in a replication study (Vaughn et al., 2015). Barth et al. (2016) conducted a text-processing reading intervention for middle-school struggling readers and found significant gains on proximal measures of vocabulary and reading comprehension; yet no significant effects were realized on a standardized measure of reading comprehension. These mixed findings led Barth et al. to wonder whether the lack of attention to decoding skills in the intervention may have played a role in the lack of progress on standardized measures of reading comprehension. Barth et al. noted the study's sample demonstrated significant deficits in word reading efficiency, aligning with previous research examining middle school students with reading difficulties (Cirino et al., 2013).

Self-Regulated Learning

To address the broader context of student learning, researchers have suggested embedding self-regulated learning (SRL) strategies into individualized reading interventions (Cirino et al., 2017; Menzies & Lane, 2011; Zimmerman & Schunk, 2011). When students self-regulate their behavior, they can plan for, monitor, and reflect on their performance (Zimmerman & Schunk, 2011). These processes allow students to actively and efficiently make decisions about their own learning by identifying when it is necessary to reread sections of a text to improve comprehension (Menzies & Lane, 2011). More specifically, brief reading comprehension intervention studies incorporating SRL strategies have led to improved reading comprehension outcomes (e.g., Berkeley, Scruggs, & Mastropieri, 2010; Mason, 2013; Zentall & Lee, 2012). Overall, for students who exhibit reading difficulties, the integration of SRL strategies into their reading instruction may hold promise as an effective mechanism to improve reading outcomes (Cirino et al., 2017; Menzies & Lane, 2011; Zimmerman & Schunk, 2011).

Writing to Enhance Reading Comprehension

Researchers have theorized the process of writing leads to improved reading comprehension skills (Ehri, 2000; Neville & Searls, 1991; Tierney & Shanahan, 1991). To examine these theories, Graham and Hebert (2011) conducted a meta-analysis to evaluate whether writing instruction, writing, or increasing the amount of time students write leads to improved reading outcomes. From this meta-analysis, Graham and Hebert (2011) identified 95 studies. Across these studies, when students wrote about a text or received writing instruction, the

mean effect size on norm-referenced measures of reading comprehension was 0.37 and 0.22, respectively. Furthermore, when students had an increase in writing (e.g., set aside time for writing, wrote short passages), the mean effect size on norm-referenced and researcher-designed measures was 0.35. In a similar meta-analysis, Graham et al. (2017) investigated the impact of interventions combining both reading and writing instruction. The results from Graham et al. (2017) also indicated positive findings with the mean effect size on norm-referenced reading outcome measures equaling 0.28. Across both these meta-analyses, findings support the use of writing to improve reading comprehension outcomes.

Current Study

The present study investigated the effects of an afterschool intervention focused on improving students' comprehension of texts using a text-processing approach. We elected to examine the effects of an afterschool intervention in light of previous syntheses and meta-analyses demonstrating reading interventions for struggling readers were associated with only modest gains on standardized outcome measures (e.g., Scammacca et al., 2015; Wanzek et al., 2010). By studying the effects of an afterschool program, we sought to test the impact of intensifying instruction for struggling readers by *supplementing* rather than *supplanting* the reading interventions that occur during the school day. We decided to utilize a text-processing approach (with self-regulation and writing instruction) to reading comprehension due to the lack of robust gains found in previous studies addressing reading comprehension through strategy instruction with upper elementary students (e.g., Ritchey et al., 2012). We were further motivated by the encouraging results of previous text-processing approaches to reading comprehension instruction for students in the upper elementary grades and middle-school grades on researcher-development measures (Barth et al., 2016; McKeown et al., 2009; Vaughn et al., 2013).

The overall question that motivated this study was, "What are the effects of assignment to an afterschool reading intervention on upper elementary struggling readers' reading comprehension?" Due to a previous text-processing intervention study (Barth et al., 2016) that hypothesized students with decoding and comprehension problems may need support in both areas to make substantial gains, we were interested in whether differential effects were present when we tested an intervention featuring foundational reading skills and text-processing to an intervention focused solely on text-processing. By contrasting these interventions, we were able to determine the additive effects of addressing foundational reading skills, an area of weakness for students with poor reading comprehension beyond the primary grades (e.g., Cirino et al., 2013). Last, owing to previous studies noting that students' attendance in voluntary afterschool programs is inconsistent and affects outcomes (Kane, 2004; Mahoney & Zigler, 2006), we investigated the extent to which the effect of treatment varied based on the number of treatment sessions that students attended. To answer these research questions, we designed a randomized control trial with students randomized to one of three conditions: (a) text-processing with foundational reading skills (TP + FS), (b) text-processing (TP), and (c) a business as usual (BaU) comparison group that received no afterschool reading instruction.

Method

Participants

Schools.—There were seven participating schools from two southwestern U.S. school districts. Four of these schools were from near urban areas, and three schools were from urban areas. The enrollment at these seven schools ranged from 435 to 761 students ($M = 654$, $SD = 123$) with 39% to 99% of the students qualifying for free or reduced lunch ($M = 61\%$, $SD = 26\%$).

Student selection.—We screened 2,453 third-, fourth-, and fifth-grade students across seven participating schools using the Test of Silent Reading Efficiency and Comprehension (TOSREC; Wagner, Torgesen, Rashotte, & Pearson, 2010). We used the TOSREC as a screener because of its strong psychometric properties and its efficient administration. Students who scored at or below a standard score of 90 (25th percentile) were eligible for participation in the study. Four hundred nineteen students met the inclusion criteria and consented to participate per the requirements of the Institutional Review Boards at the participating universities. We assigned the group of 419 students (see Table 1) to a business as usual (BaU) condition or to one of two experimental conditions: (a) TP + FS or (b) TP (without foundational reading skills). For purposes of instruction, we created small groups of writing or self-regulation instruction, within condition, yielding a multilevel, partially nested, and potentially cross-classified data structure with BaU-condition students nested in teachers (classroom teachers) and teachers nested in schools and treatment-assigned students nested in tutors and tutors crossed with teachers in schools.

Student characteristics.—Demographics for the intent-to-treat sample ($n = 419$) at Time 1 are summarized in Table 1. There were no statistically significant *pre-treatment* differences across conditions for Limited English Proficiency ($\chi^2 = 0.99$, $p = .91$), special education ($\chi^2 = 1.52$, $p = .82$), or race (p values from .06 for Asian to .97 for Hispanic). In addition, there were no Time 1 differences (see Table 2) on the extended scale score for Gates-MacGinitie Reading Test–Fourth Edition (GM-RT; MacGinitie, MacGinitie, Maria, Dreyer, & Hughes, 2000; $p = .58$), nor were there within-grade treatment-group differences at Time 1 on the GM-RT scale score (p values of .99, .37, .72 for third, fourth, and fifth graders). The Woodcock–Johnson III Passage Comprehension subtest (WJ-III PC; Woodcock, McGrew, & Mather, 2001) was administered at posttest only.

Student attrition.—Four hundred nineteen students consented to participate in the study and were randomized to a condition. Of these, 70 were lost to follow-up because they left the school during the course of the study ($n = 35$) or because their parents asked that their child no longer participate in the afterschool program ($n = 35$). Of these 70, 39 cases left before completing the pretest battery, 37 from one of the two experimental treatment conditions and two students from the BaU condition, yielding a sample of 380 cases with pretest data. We based rates of attrition on the ratio of randomized units and the sample at posttest, independent of pretest status. For TP + FS, TP, and BaU conditions, these rates were .18, .23, and .09, respectively. Sample-wide attrition was .17. The difference in attrition rates for BaU v. TP + FS and for BaU v. TP were .09 and .14, respectively. Attrition from the

combined TP + FS and TP sample differed from BaU by .12. These combinations of overall and differential attrition represent high threats to the internal validity of the study based on standards recommended by the What Works Clearinghouse (WWC; Institute of Educational Sciences, 2014). Accordingly, and per WWC's recommendations, we tested the extent to which the groups at posttest were comparable at pretest on key outcomes, bearing in mind that the majority of cases with missing posttest data were missing at pretest, as well.

A factorial ANOVA crossing condition and attrition status indicated that the BaU and the combined treatment conditions (TP + FS and TP combined) differed on the pretest GM-RT measure, $F(376,1) = 4.39$, $p = .037$. The standardized mean difference (Hedges' g) on the pretest GM-RT was .086. For TP + FS vs. BaU, the effect size was .032. The standardized difference was .131 for TP and BaU. There was no differential attrition between TP + FS and TP, at posttest or at pretest (19 and 18 missing cases at pretest for TP + FS and TP, respectively). These differences at pre-test between the treatment-assigned students and the BaU suggest potential attrition-related selection bias, so we statistically modeled differences at pretest per WWC guidance (Institute of Educational Sciences, 2014). We did not evaluate the effects associated with differential attrition for the WJ-III PC because we administered this measure at posttest only.

Measures—Trained assessment administrators, blind to student condition, conducted all assessments. Prior to the training, all assessment administrators had some experience working with students in a school setting (e.g., paraprofessional, certified teacher). The assessment delivery trainings totaled 18 hr over 3 days and followed all publisher provided protocols. In addition, the training included an overview of the standardization procedures, a review of each assessment, modeling of test administration, opportunities to practice test administration with the lead test administrators, and a reliability test in a mock setting. All test administrators scored 100% reliability in the mock administrations prior to administering the assessments to students. These trainings qualified all assessment administrators to deliver the given measures. We administered the TOSREC as a screener measure prior to randomization, the GM-RT at pre- and posttest, and the WJ-III PC at posttest only. The individually administered WJ-III PC was not delivered at pretest to adhere to school requests to minimize beginning of the school year testing.

Reading comprehension measures.—The TOSREC is a group administered sentence veracity task; alternate-form reliability coefficients exceed .85 across all grades and forms, and correlations with other reading measures, such as the WJ-III PC, exceed .70 (Wagner et al., 2010). The GM-RT is a group administered, multi choice reading comprehension assessment, internal consistency reliability coefficients range from .91 to .93, and Kuder–Richardson reliability statistics range from .92 to .93. Finally, the WJ-III PC is an individually administered cloze procedure assessment of reading comprehension; test–retest reliability coefficients for student ages 8 to 13 range from .89 to .96.

Procedures

Intervention.—Students were assigned to one of two text-processing treatment conditions: TP + FS and TP without foundational reading skills. In both treatment conditions, the aim

was for students to develop reading comprehension by thinking about and discussing text (as compared with strategy instruction with students using specific routines before, during, and after reading). Both the text-processing conditions (i.e., TP + FS, TP) included 60 min of instruction with 30 min of computer-based instruction and 30 min of researcher-delivered small-group instruction (three to six students).

The instruction for both conditions occurred 4 to 5 days per week in an afterschool setting from November to May. Across both treatment conditions, the number of 60 min lessons (i.e., 30 min of computer-based, 30 min of teacher-led small group) students received ranged from 0 to 89 ($M = 44.55$, $SD = 28.11$), with the number of lessons students received in TP + FS ranging from 0 to 89 ($M = 44.10$, $SD = 27.79$) and the number of lessons students received in TP ranging from 0 to 89 ($M = 37.04$, $SD = 28.09$). The large range of lessons the students received across both conditions was due to a broad range of participant attendance in the afterschool program.

Text-processing with foundational reading skills (TP + FS).—In the computer portion of the TP + FS (30 min daily), students utilized the same computer program throughout the entirety of the intervention. Prior to the computer-based instructional lessons, the program assessed each student's reading ability and placed them at a level based on their individual assessment results. Students then received computer-adapted, individualized explicit instruction in areas of need across multiple reading and language domains, including phonemic awareness, phonics, grammar, fluency, listening vocabulary, and/or reading comprehension. Students progressed to more complex reading and language skills as they achieved mastery on the foundational reading skills (e.g., phonemic awareness, phonics). Tutors monitored all student use of the computer program, although the number of minutes students were engaged in the instruction, as compared with off-task (e.g., staring at the screen without engaging in the instruction) and the number of minutes students were logged onto and using the program (students did not always log-off the program at the conclusion of the instruction) were not able to be reliably measured and reported.

Small-group instruction for TP + FS occurred in two phases: Phase 1 (Lessons 1 to 20) and Phase 2 (Lesson 21+). Phase 1 emphasized word reading and reading fluency, with embedded comprehension activities. Lessons included daily instruction in (a) reading fluency with expository or narrative texts (10–15 min); (b) pronouns, context clues, and inference making (10–15 min); and (c) a “Does It Make Sense?” activity (5–10 min), where students evaluated sentences for accuracy based on syntax and semantics. The second phase of the small-group instruction featured more advanced Phase 1 lessons (e.g., more complex vocabulary, longer texts, less scaffolding) in the same 30 min format. Phase 2 also included two instructional days per week with a 30 min “stretch text” or extended reading lesson on a more challenging grade-level text. Stretch text lessons focused on text reading, summarizing ideas, and answering multiple-choice comprehension questions. The stretch text embedded either a 5 min self-regulation or writing component designed as a mechanism for enhanced reading comprehension. Sample downloadable TP + FS lessons are available at the following websites: https://www.texasldcenter.org/files/lesson-plans/TCLD_3–5_T1-TBI-SelfReg.pdf and https://www.texasldcenter.org/files/lesson-plans/TCLD_3–5_T1-TBI-Writing.pdf

Text-processing (TP).—In the daily 30 min computer portion of the TP condition, all students read paper or digital texts followed by completing computer program generated multiple-choice questions. The computer program used with TP had a large assortment of readily available multiple-choice comprehension questions accompanying the paper and digital texts. Immediate feedback was provided to the students on the accuracy of their responses.

The small-group instruction for TP was in a “book club format” with students reading both instructional and gradelevel expository and narrative texts. Books were chosen by the tutors and the students. For the first 20 lessons (Phase 1), the tutor encouraged students to make predictions about the text, prior to reading the text. Whole group (e.g., choral read), partner, and individual readings were utilized for reading the book club texts. Tutors engaged students with comprehension questions throughout the lesson.

The second phase (Lessons 21+) utilized a more advanced text for 2 days per 5-day cycle of instruction. These more advanced lessons included approximately 5 min of either self-regulation or writing instruction. Sample downloadable TP lessons are available at the following websites: https://www.texasldcenter.org/files/lesson-plans/TCLD_3–5_T2-BCI-SelfRegLite.pdf and https://www.texasldcenter.org/files/lesson-plans/TCLD_3–5_T2-BCIWriting.pdf

Procedures for increasing student attendance.—As noted by previous researchers, maintaining regular student attendance is difficult in afterschool programs (Kane, 2004; Lauer et al., 2006; Zief et al., 2006). To promote student attendance, prior to the onset of the study, our research teams met with school administrators, teachers, and other personnel to develop methods to increase the likelihood of student participation. Tutors also tracked and evaluated student attendance through the collection of daily attendance data for each computer-based and teacher-led small-group session. The methods to increase attendance included (a) teachers walking students to the afterschool intervention, (b) weekly attendance report cards sent home to parents, (c) teachers and researchers making phone calls to students’ parents/guardians following three absences in a given week, and (d) incentivizing student attendance. The implementation of these methods varied in conjunction with school-stated preferences (e.g., type of incentives provided to students, teacher vs. researcher calling parents).

Intervention tutors.—Thirty-four tutors (94% female) were hired and trained by the research team to deliver the reading intervention. Of the 34 tutors, 19 held a teaching certificate, 24 held a bachelor’s degree, and 11 held a master’s degree. Tutors completed 6 hr of trainings prior to the intervention, a 2-hr training after Lesson 20, and site-based ongoing observations on an as-needed basis. Tutor trainings focused on implementing the intervention components with fidelity, using effective instructional techniques, promoting active student engagement, and implementing positive behavior management techniques.

Intervention fidelity.—All intervention lessons were audio-recorded. A subset of these lessons was randomly selected for coding, by blocking on tutor and time of year (i.e., first half of lessons and second half of lessons) to identify a mini-mum of 10% of the lessons

delivered per tutor (total lessons coded = 301). Prior to independent coding, the first author established a “gold standard” by coding one lesson for each of the two interventions. Coders ($n = 5$) were then trained and independently coded the same lesson until $> 90\%$ agreement was achieved. Initial agreement was high, and minor disputes were adjudicated through discussion. To protect against rating drift, a second reliability check (following the same procedures) was completed as each independent coder completed half of their assigned lessons.

The coding sheets for both TP + FS and TP rated intervention fidelity across two domains: (a) implementation fidelity and (b) implementation quality. Researchers rated implementation fidelity through adherence to the essential elements of each component and implementation quality through both global quality (e.g., appropriate use of feed-back and pacing) and global fidelity (i.e., holistic evaluation of implementation and success). Across domains, ratings were assigned on a 4-point Likert-type scale, based on the percentage of components (adherence) or elements (quality) present. A 4 represented the presence of all the component or elements, a 3 represented the presence of greater than 50% and less than 100% of the components or elements, a 2 represented the presence of less than 50% and more than 0% of the components or elements, and a 1 represented the absence of the components or elements.

Fidelity of implementation.—For TP + FS, across all components and tutors, the mean fidelity of implementation score was high ($M = 3.79$, $SD = 0.66$, range = 3.00–4). For TP, across all components and tutors, the mean fidelity of implementation score was somewhat lower ($M = 3.34$, $SD = 0.90$), but still in the medium to high fidelity range. Mean scores for TP tutors varied more widely (range = 2.67–4.0).

Quality of implementation.—For both TP + FS and TP, a global quality of implementation score was given in four domains, including quality of (a) instruction, (b) student management, (c) intervention, and (d) student engagement. For TP + FS, quality scores were high across all dimensions: instruction ($M = 3.82$, $SD = 0.38$), student management ($M = 3.91$, $SD = 0.32$), intervention ($M = 3.79$, $SD = 0.41$), and student engagement ($M = 3.96$, $SD = 0.24$). There was little variability on these four dimensions across tutor (range = 3.25–4.0). For TP, quality scores were somewhat lower across all four dimensions: instruction ($M = 3.18$, $SD = 0.81$), student management ($M = 3.62$, $SD = 0.66$), intervention ($M = 2.99$, $SD = 0.91$), and student engagement ($M = 3.72$, $SD = 0.57$). There was greater variability across tutor ratings, as well (range = 2.33–4.0).

Additional reading interventions.—A lead teacher at each participating school completed the additional reading interventions (ARI) survey. This survey was used to collect data on the frequency and duration of the school-based reading interventions each student received during and outside of the core reading instruction. We were able to collect data on 392 of the 419 (93.5%) participating students. Results from the ARI indicated, throughout the entire school year, all schools implemented a commercially available core reading curriculum targeting a combination of phonics, fluency, vocabulary, and/or comprehension. Furthermore, across all conditions, 308 of the 392 students participated in at least one intervention group outside of their core instruction. Six-teen students participated in more

than one school-based reading intervention. These interventions included computer-based instruction during the school day, teacher-led instruction using commercially available curricula (e.g., Achieve3000, 2015) and teacher-, school-, and district created curricula.

Analysis Plan

We modeled two types of effects—intent-to-treat (ITT) and a compliance-adjusted effect. The ITT effect is the difference in outcomes between the group assigned to treatment(s) and the group assigned to comparison or BaU, regardless of participants' actual compliance or receipt of treatment. It is the effect of assignment rather than treatment. A compliance-adjusted effect represents treatment's effect for cases that actually "take up" the treatment. When compliance is binary (did comply or did not comply), local average treatment effects (also called the complier-average causal effect) can be calculated using assignment as an instrumental variable (Angrist & Imbens, 1995). Because participants' reasons for noncompliance can be conceptualized as omitted variables, participation in the treatment is not independent of potential outcomes and confounders (e.g., students who attend more consistently are likely to differ from students who attend less consistently, and these unmeasured or omitted differences rather than the treatment may account for any difference in outcomes). However, assignment *is* independent, and if assignment positively affects the probability of receiving or participating in treatment, it represents an *instrument* for identifying the effects of the treatment on the subpopulation of subjects who comply with assignment. In cases of *one-sided* noncompliance, where compliance is possible among only those assigned to treatment and where noncompliance is binary in the treatment- assigned group, the local average treatment effect (LATE) equals the average treatment effect on the treated (ATET).

Because we delivered treatments after school, we expected that attendance for treatment-assigned students would vary; some would not attend, others would attend all sessions, but most would attend a subset of the available sessions. We did not expect students assigned to the BaU condition to participate in treatment (nor did they). This represents one-sided (non)compliance as described above. However, it also means that compliance *cannot* be fully captured as a binary variable; for treatment-assigned students, compliance, operationalized as number of sessions attended, distributes as a continuous variable and represents an instance of *partial* (non)compliance. The local average treatment effect produces biased estimates of treatment effects in cases of partial (non)compliance. However, Flores and Flores-Lagunes (2013) propose a method for partial identification of the local average treatment effect that conceptualizes this as a causal mediation effect, allowing for the identification of the local average treatment effect within the context of the causal mediation analysis literature (see also Pearl, 2001; Robins, 2003; Robins & Greenland, 1992). We apply this approach, which they label the mediation average treatment effect, to estimate compliance-adjusted effects for the TP + FS and TP treatments.

Results

Preliminary Analyses

We fit unconditional models (i.e., empty means, random intercept) to estimate variance related to partial nesting and cross-classification at different and multiple levels of the models. We began by evaluating partial nesting of tutors within classroom teachers for students in the treated conditions under the assumption that trivial amounts of tutor-level partial nesting would preclude any consequential cross-classification of students from the same classrooms to different tutors. To test the hypothesis that between-groups' variation (the intra-class correlation between tutor groups) did not differ statistically from 0 (within-group deviations were normally distributed), we calculated F -statistics for group effects as $\bar{n}(S^2_{\text{between}}) / (S^2_{\text{within}})$ where $\bar{n}$ is the average cluster size, S^2_{between} is the observed between-group variance, and S^2_{within} is the observed within-group variance (Snijders & Bosker, 1999). The statistic was tested against the F distribution with $N - 1$ and $M - N$ degrees of freedom, N being the number of Level 2 units and M the number of Level 1 units.

Between-group variation was not statistically significant from zero for the GM-RT, $F(35,187) = 1.03$, $p = .43$, or for the W scores on the WJ-III PC, $F(35,187) = 1.04$, $p = .43$. Accordingly, we did not model partial nesting, nor did we model the cross-classification of teacher and tutor for the reason outlined above. Instead, we fit fully nested multi-level models with students nested in classroom teachers (site- and school-level clustering were minimal) and treatment status modeled as a Level 1 fixed effect.

Intent to Treat Effect

We summarize the intent-to-treat effects for the two comprehension outcomes in Table 3. We report two contrasts for each outcome: (a) the treatment conditions combined (TP + FS and TP COMB) and contrasted with the BaU condition and (b) the primary treatments, TP + FS and TP, contrasted with the BaU condition. We modeled the data as multi group in *Mplus* 8. In the model for the GM-RT, pretest scores were included as covariates. The WJ-III PC was administered at posttest only. For clarity's sake, we report differences in simple slopes between each treatment group and BaU condition for model-estimated scale scores at posttest for the GM-RT. Because the statistical significance of the two simple slopes is tested within each group, residual variances can be fixed as equal, meaning that the contrasts in Table 3 represent standardized estimates of the treatment's effect (Raudenbush & Liu, 2001). The WJ-III PC was administered at posttest only. Accordingly, we estimated standardized mean differences in groups' posttest scores as model parameters. The pattern of effects is stable and trivial in magnitude. Across conditions and combinations of conditions and across outcomes, there were no statistically significant differences. There were no moderating effects for race, gender, special education status, or limited English proficiency status.

Compliance-Adjusted Treatment Effect

Figure 1 displays the distribution of sessions attended by students assigned to one of the four experimental treatment conditions. Just over 50 of the treatment-assigned cases did not participate; they attended 0 sessions. The distribution of sessions attended among those who attended at least one session ($n = 235$) is somewhat normal, if slightly left-skewed, with a

mean of 48 and standard deviation of 23.8. It is tempting to remove the 50 or so treatment-assigned cases that did not attend a session; however, as suggested earlier, doing so introduces selection bias to the extent that never-attending students differ from the group of attending students in ways that correlate with the study's outcomes, which is likely. Moving these cases to the BaU group compounds the problem, and represents a particularly poor strategy in the context of randomized designs, which are primarily concerned with minimizing selection bias.

Instead, we modeled the cases according to their original assignment (as in ITT), with assignment as the instrument using the IV language, or as the exogenous variable in the causal mediation context. Number of sessions served as the mediator (or alternative treatment) and the reading out-comes were modeled as dependent variables. We estimated direct (the effect of assignment on outcomes), indirect (the effect of assignment through number of sessions), and total effects (the sum of direct and indirect effects) in *Mplus* 8.0, using bootstrapping (1,000 draws) to correct standard errors and establish confidence intervals. We also estimated the effect of number of sessions (the mediator) on outcomes in the context of mediation. We fit the models as multilevel, contrasting both TP + FS and TP to the BaU condition.

The model makes several assumptions, discussed at length in the literature on instrumental variable models (Angrist & Imbens, 1995). We note several. First, it assumes that assignment (or the instrument) has positive and significant effect on the mediator (or receipt of treatment) and that receipt of treatment is the only means by which assignment can influence outcomes. The former can be tested empirically, and we assume the latter with some confidence given that there are no other apparent (or imaginable) mechanisms through which assignment might influence outcomes.

A second, related assumption, is that observed (potential) outcomes for one individual are unaffected by the assignment of treatments to other persons (Rubin, 1990). Often called the Stable Unit Treatment Value Assumption (SUTVA), it presumes more than simple independence of observations. In the present context, it implies that a student's change in reading outcomes cannot depend on another student's assignment to one or another treatment. We eliminated cross-classification as a variance component in these data; however, doing so may not entirely satisfy SUTVA. It is possible to conceptualize *alternative treatments* that are conditional on a third party's assignment, particularly given the classroom-nested structure of the data and the considerable partial noncompliance.

The mediated average treatment effects (MATE) are not dissimilar from the ITT results. The direct, indirect, and total effects for TP + FS equaled $-.05$ ($SE = .09$, $p = .52$), $.02$ ($SE = .07$, $p = .73$), and $-.03$ ($SE = .04$, $p = .40$), respectively. For TP, the effects were $-.06$ ($SE = .07$, $p = .46$), $.03$ ($SE = .05$, $p = .48$), and $-.03$ ($SE = .05$, $p = .46$) for direct, indirect, and total effects, respectively. The effect of the number of sessions on the GM-RT outcome for TP + FS equaled $.03$ ($SE = .09$, $p = .73$) and for TP equaled $.06$ ($SE = .07$, $p = .47$). These treatment effects are all very small in magnitude and do not differ from 0. Note also that the coefficients for TP + FS and TP are very similar; they performed comparably when contrasted with the BaU condition. As a check on these results, we also treated compliance

as binary, under the assumption that attendance in the population for an afterschool program has a distribution similar to the one in Figure 1. Given this restriction, the IV estimate (the ATET) reduces to the ITT divided by the proportion of compliers. In this sample, 82% of students complied (i.e., attended at least one session). Dividing the coefficients by .82 does not improve the treatment effects appreciably.

Discussion

Using an RCT design, this study examined the relative effects of supplementing typical school day reading instruction with two contrasting text-processing afterschool reading interventions for third-, fourth-, and fifth-grade students with significant reading difficulties. Findings revealed that students receiving an intensive afterschool reading intervention did not outperform students assigned to the BaU condition, despite previously encouraging results from text-processing approaches implemented during the school day (Barth et al., 2016; McKeown et al., 2009; Vaughn et al., 2013). In addition, no statistically significant differences emerged between the two contrasting text-processing reading conditions (i.e., TP + FS, TP) on measures of reading comprehension (i.e., WJ-III PC, GM-RT). The reading comprehension Hedges' g effect sizes between each treatment and the BaU condition ranged from -0.003 to 0.02 , with TP + FS ranging from 0.003 to 0.02 and TP ranging from -0.003 to -0.002 (see Table 3). These effect sizes were similar to the effect sizes found for normal school day extensive reading comprehension interventions with upper elementary students ($M = 0.09$; Wanzek et al., 2010) and out of school time reading interventions with upper elementary students ($M = -0.03$; Lauer et al., 2006).

This study's status as one of the few RCTs of an after-school reading intervention suggests a need for additional research *to the extent that afterschool time represents a legitimate opportunity to intensify instruction for older struggling students* (Apsler, 2009; Durlak & Weissberg, 2007; Lauer et al., 2006; Riggs & Greenberg, 2004). We believe that this latter caveat deserves additional consideration. It may be that reading interventions provided after school to students who are nine to 11 years old are ineffective *precisely* because they are provided after school. At this time of day, it may be reasonable to address domains that offer instruction demanding fewer cognitive resources or to practice recently acquired but still developing skills as compared with the instruction provided in this intervention.

Alternatively, if one assumes that afterschool time *does* represent a legitimate learning opportunity for younger students, the question becomes "how best to use that time?" in this case to improve reading comprehension. Young students struggling to read and comprehend benefit from pro-grams that present sequential instructional activities to promote active, focused, and explicit learning opportunities (Durlak & Weissberg, 2007; Granger, 2008; Granger, Durlak, Yohalem, & Reisner, 2007). This assumes that students are regularly present for instruction. It also suggests that sporadic attendance may exacerbate existing gaps in students' skill set. Our findings indicate that variation in the number of sessions attended by treatment-assigned students did not improve comprehension outcomes, a result that appears not to support this argument.

Limitations and Future Research

In implementing this study, we experienced difficulties with sporadic attendance and high attrition. Across both treatment conditions, 45 students (16.1% of treatment sample) did not attend a single treatment session. Even though these difficulties are not unique, as other researchers have documented low attendance and high attrition in afterschool programs (Apsler, 2009; Durlak & Weissberg, 2007; Riggs & Greenberg, 2004), the low dosage and high attrition make it difficult to know the potential impact of the study.

For future research and practice, the next step for after-school intervention research is to identify strategies for supporting students' attendance in conjunction with or beyond as described elsewhere in this article. One possibility is for future studies to collect student social validity data, which may provide further guidance on how to support such an endeavor of increasing student attendance and outcomes. In addition, Lauer and colleagues (2006) report that one-to-one instruction is associated with higher effects in afterschool programs, with researchers also noting the possibility of ancillary effects on academic outcomes from the use of social/emotional interventions during normal school hours (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011) and in afterschool settings (Durlak, Weissberg, & Pachan, 2010). The findings from these studies do not provide clear guidance about what elements of an afterschool program would need to be adjusted to improve outcomes. However, we think it is worth considering student supports through increased individualization, such as a decreased group size or the inclusion of social/emotional learning. These forms of increased individualization could lead to improved motivation, more consistent attendance, and increased reading outcomes. To test these hypotheses, the current three condition study could be replicated with a fourth condition of smaller group sizes or a social/emotional learning component. Such a design would provide evidence regarding the extent to which these forms of increased individualization impacted reading outcomes.

As noted by previous researchers (e.g., Lauer et al., 2006), future research should also be thorough when describing student characteristics, methods, and analysis of their studies. This represents a practical rather than empirical enterprise, but may be the only way to collect and report best practice related to student attendance in afterschool interventions. Research to date has not been consistent in this respect. Lauer and colleagues (2006) found only five out of 35 studies reported student attendance data.

Finally, our study alerts practitioners to the challenges of providing systematic afterschool reading interventions with low-performing students. For researchers, our experience underscores the importance of collaborating with schools when designing afterschool intervention studies. In both these cases, future work is needed to identify and address questions on how to best optimize afterschool conditions to yield the greatest outcomes.

In summary, we think that this RCT yields valuable data about the potential impact of afterschool programs. Like previous studies, this study indicates that providing more instruction after school for students with learning difficulties may not yield the desired outcome of improved comprehension. However, there are valuable questions to be addressed about afterschool programs that this study guides us to consider: (a) Are there conditions

(e.g., one-on-one instruction) under which afterschool reading programs might be associated with improved outcomes? (b) Are there instructional areas (e.g., early reading instruction, vocabulary development) that are more malleable to afterschool interventions? and (c) Are there grade groupings (e.g., students in Grades K-2) that are better suited to afterschool interventions?

Acknowledgments

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by Award Number P50 HD052117, from the Eunice Kennedy Shriver National Institute of Child Health & Human Development.

References

- Achieve3000. (2015). Achieve3000 Texas Retrieved from <http://www.achieve3000.com/texas>
- Alliance A (2014). Taking a deeper dive into afterschool: Positive outcomes and promising practices Washington, DC Retrieved from <http://www.srgforcesports.com/wp-content/uploads/2016/05/Taking-A-Deeper-Dive-Into-Afterschool.pdf>
- Angrist J, & Imbens G (1995). Identification and estimation of local average treatment effects. *Econometrica*, 62, 467–475. doi:10.3386/t0118
- Applebee AN, Langer JA, Nystrand M, & Gamoran A (2003). Discussion-based approaches to developing understanding: Classroom instruction and student performance in middle and high school English. *American Educational Research Journal*, 40, 685–730. doi: 10.3102/00028312040003685
- Apsler R (2009). After-school programs for adolescents: A review of evaluation research. *Adolescence*, 44, 1–19. [PubMed: 19435164]
- Barth AE, Vaughn S, Capin P, Cho E, Stillman-Spisak S, Martinez L, & Kincaid H (2016). Effects of a text processing comprehension intervention on struggling middle school readers. *Topics in Language Disorders*, 36, 368–389. doi:10.1097/TLD.000000000000101 [PubMed: 28943694]
- Beck IL, & McKeown MG (2006). Improving comprehension with questioning the author: A fresh and expanded view of a powerful approach New York, NY: Scholastic.
- Berkeley S, Scruggs TE, & Mastropieri MA (2010). Reading comprehension instruction for students with learning disabilities, 1995–2006: A meta-analysis. *Remedial and Special Education*, 31, 423–436. doi:10.1177/0741932509355988
- Bissell JS, Cross CT, Mapp K, Reisner E, Vandell DL, Warren C, & Weissbourd R (2003, 5). A statement released by members of the Scientific Advisory Board for the 21st Century Community Learning Center evaluation Retrieved from <http://childcare.wceruw.org/pdf/publication/statement.pdf>
- Cirino PT, Miciak J, Gerst E, Barnes MA, Vaughn S, Child A, & Huston-Warren E (2017). Executive function, self-regulated learning, and reading comprehension: A training study. *Journal of Learning Disabilities*, 50, 450–467. [PubMed: 26746314]
- Cirino PT, Romain MA, Barth AE, Tolar TD, Fletcher JM, & Vaughn S (2013). Reading skill components and impairments in middle school struggling readers. *Reading and Writing*, 26, 1059–1086. doi:10.1007/s11145-012-9406-3 [PubMed: 24000271]
- Durlak JA, & Weissberg RP (2007). The impact of after-school programs that promote personal and social skills Chicago, IL: Collaborative for Academic, Social, and Emotional Learning.
- Durlak JA, Weissberg RP, Dymnicki AB, Taylor RD, & Schellinger KB (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, 82, 405–432. doi:10.1111/j.1467-8624.2010.01564 [PubMed: 21291449]

- Durlak JA, Weissberg RP, & Pachan M (2010). A meta-analysis of after-school programs that seek to promote personal and social skills in children and adolescents. *American Journal of Community Psychology*, 45, 295–309. doi:10.1007/s10466-010-9300-6
- Ehri LC (2000). Learning to read and learning to spell: Two sides of a coin. *Topics in Language Disorders*, 20, 19–36. doi:10.1097/00011363-200020030-00005
- Flores CA, & Flores-Lagunes A (2013). Partial identification of local average treatment effects with an invalid instrument. *Journal of Business & Economic Statistics*, 31, 534–545. doi:10.1080/07350015.2013.822760
- Flynn LJ, Zheng X, & Swanson HL (2012). Instructing struggling older readers: A selective meta-analysis of intervention research. *Learning Disabilities Research & Practice*, 27, 21–32. doi:10.1111/j.1540-5826.2011.00347.x
- Graham S, & Hebert M (2011). Write to read: A meta-analysis of the impact of writing and writing instruction on reading. *Harvard Educational Review*, 81, 710–744.
- Graham S, Liu X, Aitken A, Ng C, Bartlett B, Harris KR, & Holzapfel J (2017). Effectiveness of literacy programs balancing reading and writing instruction: A meta-analysis. *Reading Research Quarterly*. Advanced online publication doi:10.1002/rrq.194
- Granger RC (2008). After-school programs and academics: Implications for policy, practice, and research. *Social Policy Report*, 22, 14–19.
- Granger RC, Durlak JA, Yohalem N, & Reisner E (2007). Improving after-school program quality New York, N.Y: William T. Grant Foundation.
- Institute of Educational Sciences. (2014). What Works Clearinghouse procedures and standards handbook version 3.0 Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_v3_0_standards_handbook.pdf
- James-Burdumy S, Dynarski M, Moore M, Deke J, Mansfield W, Pistorino C, & Warner E (2005). When schools stay open late: The national evaluation of the 21st Century Community Learning Centers Program Washington, DC: U.S Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.
- Kane TJ (2004). The impact of after-school programs: Interpreting the results of four recent evaluations New York, NY: William T. Grant Foundation.
- Kintsch W (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95, 163–182. [PubMed: 3375398]
- Kremer KP, Maynard BR, Polanin JR, Vaughn MG, & Sarteschi CM (2015). Effects of after-school programs with at-risk youth on attendance and externalizing behaviors: A systematic review and meta-analysis. *Journal of Youth and Adolescence*, 44, 616–636. doi:10.1007/s10964-014-0226-4 [PubMed: 25416228]
- Lauer PA, Akiba M, Wilkerson SB, Apthorp HS, Snow D, & Martin-Glenn ML (2006). Out-of-school time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research*, 76, 275–313. doi:10.3102/00346543076002275
- MacGinitie WH, MacGinitie RK, Maria K, Dreyer LG, & Hughes KE (2000). *Gates-MacGinitie Reading Tests* (4th ed.). Itasca, IL: Riverside.th
- Mahoney JL, & Zigler EF (2003). The national evaluation of the 21st-Century Community Learning Centers: A critical analysis of the first-year findings New Haven, CT: Yale University.
- Mahoney JL, & Zigler EF (2006). Translating science to policy under the no child left behind act of 2001: Lessons from the national evaluation of the 21st-century community learning centers. *Journal of Applied Developmental Psychology*, 27, 282–294. doi:10.1016/j.appdev.2006.04.001
- Mason LH (2013). Teaching students who struggle with learning to think before, while, and after reading: Effects of self-regulated strategy development instruction. *Reading & Writing Quarterly*, 29, 124–144. doi:10.1080/10573569.2013.758561
- McKeown MG, Beck IL, & Blake RG (2009). Rethinking reading comprehension instruction: A comparison of instruction for strategies and content approaches. *Reading Research Quarterly*, 44, 218–253. doi:10.1598/RRQ.44.3.1
- Menzies HM, & Lane KL (2011). Using self-regulation strategies and functional assessment-based interventions to provide academic and behavioral support to students at risk within three-tiered

- models of prevention. Preventing School Failure: Alternative Education for Children and Youth, 55, 181–191. doi:10.1080/1045988X.2010.520358
- Neville DD, & Searls EF (1991). A meta-analytic review of the effect of sentence-combining on reading comprehension. *Reading Research & Instruction*, 31, 63–76.
- O'Connor RE, Bell KM, Harty KR, Larkin LK, Sackor SM, & Zigmond N (2002). Teaching reading to poor readers in the intermediate grades: A comparison of text difficulty. *Journal of Educational Psychology*, 94, 474. doi:10.1037/0022-0663.94.3.474
- Pearl J (2001). Direct and indirect effects. In Breese J & Koller D (Eds.), *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence* (pp. 411–420). San Francisco, CA: Morgan Kaufmann.
- Raudenbush SW, & Liu XF (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods*, 6, 387–401. doi:10.1037/1082-989X.6.4.387 [PubMed: 11778679]
- Riggs NR, & Greenberg MT (2004). After-school youth development programs: A developmental-ecological model of current research. *Clinical Child and Family Psychology Review*, 7, 177–190. doi:10.1023/B:CCFP.0000045126.83678.75 [PubMed: 15645707]
- Ritchey KD, Silverman RD, Montanaro EA, Speece DL, & Schatschneider C (2012). Effects of a tier 2 supplemental reading intervention for at-risk fourth-grade students. *Exceptional Children*, 78, 318–334. doi:10.1177/0011440291207800304 [PubMed: 22685347]
- Robins JM (2003). Semantics of causal DAG models and the identification of direct and indirect effects. In Green PJ, Hjort NL & Richardson S (Eds.), *Highly structured stochastic systems* (pp. 70–81). New York, NY: Oxford University Press.
- Robins JM, & Greenland S (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, 3, 143–155. doi:10.1097/00001648-199203000-00013 [PubMed: 1576220]
- Rubin DB (1990). Formal mode of statistical inference for causal effects. *Journal of Statistical Planning and Inference*, 25, 279–292. doi:10.1016/0378-3758(90)90077-8
- Scammacca NK, Roberts G, Vaughn S, & Stuebing KK (2015). A meta-analysis of interventions for struggling readers in grades 4–12: 1980–2011. *Journal of Learning Disabilities*, 48, 369–390. doi:10.1177/0022219413504995 [PubMed: 24092916]
- Scott-Little C, Hamann MS, & Jurs SG (2002). Evaluations of after-school programs: A meta-evaluation of methodologies and narrative synthesis of findings. *American Journal of Evaluation*, 23, 387–419. doi:10.1177/109821400202300403
- Snijders TAB, & Bosker RJ (1999). Testing and model specification. In Snijders TAB & Bosker RJ (Eds.), *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (pp. 86–98). Thousand Oaks, CA: SAGE.
- Therrien WJ, Wickstrom K, & Jones K (2006). Effect of a combined repeated reading and question generation intervention on reading achievement. *Learning Disabilities Research & Practice*, 21, 89–97. doi:10.1111/j.1540-5826.2006.00209.x
- Tierney RJ, & Shanahan T (1991). Research on the reading writing relationship: Interactions, transactions, and outcomes. In Barr R, Kamil M, Mosenthal P & Pearson D (Eds.), *The handbook of reading research* (Vol. 2; pp. 246–280). New York, NY: Longman.
- Vaughn S, Roberts G, Swanson EA, Wanzek J, Fall AM, & Stillman-Spisak SJ (2015). Improving middle-school students' knowledge and comprehension in social studies: A replication. *Educational Psychology Review*, 27, 31–50. doi:10.1007/s10648-014-9274-2
- Vaughn S, Solis M, Miciak J, Taylor WP, & Fletcher JM (2016). Effects from a randomized control trial comparing researcher and school-implemented treatments with fourth graders with significant reading difficulties. *Journal of Research on Educational Effectiveness*, 9, 23–44. doi:10.1080/19345747.2015.1126386 [PubMed: 28491206]
- Vaughn S, Swanson EA, Roberts G, Wanzek J, Stillman-Spisak SJ, Solis M, & Simmons D (2013). Improving reading comprehension and social studies knowledge in middle school. *Reading Research Quarterly*, 48, 77–93. doi:10.1002/rrq.039
- Wagner RK, Torgesen JK, Rashotte CA, & Pearson NA (2010). *TOSREC: Test of sentence reading efficiency and comprehension* Austin, TX: Pro-Ed.

- Wanzek J, Petscher Y, Al Otaiba S, Kent SC, Schatschneider C, Haynes M, . . . Jones FG (2016). Examining the average and local effects of a standardized treatment for fourth graders with reading difficulties. *Journal of Research on Educational Effectiveness*, 9, 45–66. doi: 10.1080/19345747.2015.1116032
- Wanzek J, & Roberts G (2012). Reading interventions with varying instructional emphases for fourth graders with reading difficulties. *Learning Disability Quarterly*, 35, 90–101. doi: 10.1177/0731948711434047
- Wanzek J, & Vaughn S (2007). Research-based implications from extensive early reading interventions. *School Psychology Review*, 36, 541–561.
- Wanzek J, Vaughn S, Scammacca N, Gatlin B, Walker MA, & Capin P (2016). Meta-analyses of the effects of tier 2 type reading interventions in grades K-3. *Educational Psychology Review*, 28, 551–576. doi:10.1007/s10648-015-9321-7 [PubMed: 27594774]
- Wanzek J, Wexler J, Vaughn S, & Ciullo S (2010). Reading interventions for struggling readers in the upper elementary grades: A synthesis of 20 years of research. *Reading and Writing*, 23, 889–912. doi:10.1007/s11145-009-9179-5 [PubMed: 21072128]
- Woodcock RW, McGrew KS, & Mather N (2001). *Woodcock- Johnson III tests of achievement* Itasca, IL: Riverside.
- Zentall SS, & Lee J (2012). A reading motivation intervention with differential outcomes for students at risk for reading disabilities, ADHD, and typical comparisons: “Clever is and clever does.” *Learning Disability Quarterly*, 35, 248–259. doi:0731948712438556
- Zief SG, Lauver S, & Maynard RA (2006). Impacts of afterschool programs on student outcomes. *Campbell Systematic Reviews*, 3, 1–56. doi:10.4073/csr.2006.3
- Zimmerman B, & Schunk DH (Eds.). (2011). *Handbook of self-regulation of learning and performance* New York, NY: Taylor & Francis.

Figure 1.
Histogram of intervention sessions attended by treatment-assigned students.

Table 1.Demographics and Frequencies for the Intent-to-Treat Sample ($n = 419$).

Characteristic	Condition		
	BaU ($n = 139$)	TP + FS ($n = 139$)	TP ($n = 141$)
Grade			
3 ($n = 113$)	37	38	38
4 ($n = 160$)	48	61	51
5 ($n = 146$)	54	40	52
Limited English Proficiency			
Missing	13.5%	16.3%	16.3%
Yes	16.7%	16.3%	16.3%
Special education			
Missing	12.7%	15.6%	15.6%
Yes	17.4%	19.1%	19.1%
Race/ethnicity			
African American	36.0%	36.0%	38.3%
Asian	0.0%	0.0%	2.1%
Hispanic	7.2%	5.0%	6.4%
Native American	1.4%	2.9%	0.7%
White	47.5%	36.0%	49.6%
2 races/ethnicities	7.9%	5.0%	2.8%

Note. BaU = business as usual; TP + FS = text-processing with foundational reading skills; TP = text-processing.

Table 2.Means, Standard Deviations, and Valid *N* at Pre- and Posttest.

Measures at pre- and posttest	Condition								
	BaU			TP + FS			TP		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Pretest									
GM-RT	137	453.10	33.50	120	452.06	29.26	123	448.65	33.44
Posttest									
GM-RT	126	465.24	33.02	114	461.33	33.81	109	458.07	35.65
WJ-III PC	126	485.13	13.53	114	485.91	15.27	109	484.70	14.07

Note. BaU = business as usual; TP + FS = text-processing with foundational reading skills; TP = text-processing; GM-RT = Gates MacGinitie-Reading Test Fourth Edition (MacGinitie, MacGinitie, Maria, Dreyer, & Hughes, 2000); WJ-III PC = Woodcock-Johnson III Tests of Achievement Passage Comprehension subtest (Woodcock, McGrew, & Mather, 2001).

Table 3.

Intent-to-Treat Standardized Effects.

Assessment and condition	Estimate	SE	p value
GM-RT			
TP + FS and TP COMB	.01	.069	.92
TP + FS			
TP + FS	.02	.079	.81
TP	-.003	.083	.97
WJ-III PC			
TP + FS and TP COMB	.001	.006	.89
TP + FS and TP			
TP + FS	.003	.008	.66
TP	-.002	.008	.80

Note. GM-RT = Gates MacGinitie–Reading Test Fourth Edition (MacGinitie, MacGinitie, Maria, Dreyer, & Hughes, 2000); TP + FS = text-processing with foundational reading skills; TP = text-processing; COMB = combined; WJ-III PC = Woodcock–Johnson III Tests of Achievement Passage Comprehension subtest (WJ-III; Woodcock, McGrew, & Mather, 2001).