Intensive Intervention for Upper Elementary Students with Severe Reading Comprehension Difficulties

Jeanne Wanzek

Vanderbilt University

Stephanie Al Otaiba

Southern Methodist University

Christopher Schatschneider

Florida State University

Rachel E. Donegan

Vanderbilt University

Brenna Rivas

Francesca Jones

Southern Methodist University

Yaacov Petscher

Florida State University

Wanzek, J., Al Otaiba, S., Schatschneider, C., Donegan. R. E., Rivas, B., Jones, F., Petscher, Y. (2020). Intensive intervention for upper elementary students with severe reading comprehension difficulties. *Journal of Research on Educational Effectiveness, 13*, 408-429.

Author note

Jeanne Wanzek, Department of Special Education, Vanderbilt University; Stephanie Al Otaiba, Simmons School of Education , Southern Methodist University; Christopher Schatschneider, Florida Center for Reading Research and Department of Psychology, Florida State University; Rachel E. Donegan, Department of Special Education, Vanderbilt University;

Brenna K. Rivas, Simmons School of Education, Southern Methodist University, Francesca G. Jones, Simmons School of Education Southern Methodist University; Yaacov Petscher, Florida Center for Reading Research, Florida State University

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R324A150269 to Vanderbilt University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Correspondence should be addressed to Jeanne Wanzek, Department of Special Education, Vanderbilt University, 110 Magnolia Circle, Nashville, TN 37203.

Email: jeanne.wanzek@vanderbilt.edu

Abstract

This study examined the effectiveness of the intensive implementation of a multicomponent reading intervention (Voyager Passport) with 306 fourth grade students with severe reading difficulties. Students with reading comprehension achievement below the 15%ile in the fall of fourth grade were randomly assigned to the intensive intervention or to typical school services. Students assigned to the intensive intervention were placed in instructional groups of 2-3 students and received daily, 45 min sessions of supplemental, multicomponent reading intervention throughout the school year. Students in the intensive intervention significantly outperformed their peers receiving typical school services in word reading (ES = 0.25) and word reading fluency (ES = 0.19). Study groups performed similarly on measures of reading fluency and comprehension. Ultimately, students with severe reading difficulties accelerated their word reading and word reading efficiency achievement, but the acceleration was not enough to also accelerate their reading fluency or reading comprehension.

Intensive Intervention for Upper Elementary Students with Severe Reading Comprehension Difficulties

As reading instruction progresses to the late elementary grades, unique characteristics begin to emerge that distinguish it from instruction provided in the early grades. Instruction shifts from an emphasis on basic skills common in the early grades to an emphasis on more advanced comprehension and related skills (Kent, Wanzek, & Al Otaiba, 2017). In addition, students are expected to develop the ability to learn content from reading text (Torgesen et al., 2007). This expectation of academic literacy creates an increasing urgency for students with reading difficulties to solidify basic reading skills and accelerate learning in areas crucial to reading comprehension in order to ready them for the transition to middle and high school grades.

Upper Elementary Multicomponent Reading Interventions

Research on reading intervention for students with or at risk for reading disabilities in the upper elementary grades demonstrates generally positive effects (Authors, in review; Flynn, Zheng, & Swanson, 2012; Wanzek Wexler, Vaughn, & Ciullo, 2010). However, a recent meta-analysis demonstrates overall nonsignificant effects on standardized measures of both foundational skills (e.g., word reading, fluency) and reading comprehension (Authors, in review). This research also indicates students who struggle with reading beyond the early elementary grades frequently demonstrate deficits in multiple areas of reading (i.e. phonemic awareness, phonics, fluency, vocabulary, and comprehension) affecting their overall reading performance (e.g. Torgesen et al., 2007), and may require intervention in several reading areas. Multicomponent reading interventions are designed with these students in mind and address a multitude of reading deficits within a single intervention.

The simple view of reading (Gough & Tunmer, 1986), a theoretical model of processes that underlie successful reading comprehension, provides support for multicomponent reading interventions. According to the simple view, reading comprehension results from a multiplicative relationship between decoding and listening comprehension (Gough & Tunmer, 1986). In other words, a students' ability to understand text is inherently linked to his or her ability to accurately and efficiently read words on a page as well as understand that text, and many older readers still struggle with decoding and word recognition abilities (Cho, Capin, Roberts, Roberts, & Vaughn, 2019; Cirino et al., 2013). This makes it reasonable to predict that reading interventions may need to address any deficits in foundational reading skills as well as comprehension to achieve overall reading comprehension. As students progress into middle and secondary grades word reading remains important to student reading comprehension as vocabulary, background knowledge, and inferencing take even stronger relationships with reading comprehension (Ahmed et al., 2016; Cromley & Azevedo, 2007)

Several experimental studies of varying intensities have examined the effectiveness of multicomponent reading interventions for late elementary students. Ritchey, Silverman, Montanaro, Speece, and Schatschneider (2012) examined the effects of a relatively low intensity multicomponent intervention (24, 40 min sessions) for fourth grade students with reading difficulties. The intervention included instruction in reading fluency, vocabulary, and comprehension within informational science text and was provided in groups of two to four students for 24, 40 min sessions. No significant between group differences were noted on standardized measures of reading comprehension, word recognition, or reading fluency, though students who participated in the intervention did outperform students in the control group on

near-transfer measures of comprehension strategy use and science knowledge (ES range = 0.56 to 0.65).

Several upper elementary studies implementing multicomponent interventions and examining standardized reading measures have reported more promising results. These intervention have generally implemented more intensive elements (longer duration (weeks or session length) or smaller group size). However, there are some mixed findings in this regard; and an updated meta-analysis of upper elementary studies demonstrates the difficulties of looking at intensity across studies (Authors, in review). In particular, shorter duration interventions actually had higher effects on reading comprehension outcomes but this finding was limited by the use of mostly unstandardized measures. The longest duration interventions also did not predict better foundational outcomes, but those studies also had more impaired samples of students at the start of the study. Studies with 16-30 hours of intervention did predict better foundational outcomes than studies with less than 16 hours of intervention (Authors, in review).

Examining the experimental studies implementing multicomponent interventions at the upper elementary level and including standardized outcomes we see some promising results for foundational outcomes and mixed results for reading comprehension. Therrien, Wickstrom, and Jones (2006) investigated the effects of a 1:1 reading intervention which included instruction in reading fluency and comprehension, for a sample of struggling readers largely in Grades 4 and 5 who were at least two grade levels behind in reading. Students participated in a total of 50, 15 to 30-min sessions over a 4-month period. The authors noted a significant effect for reading fluency (ES = 0.44) for students in the intervention. No significant differences were noted on a standardized measure of broad reading though a large effect size (ES = 0.69) in favor of the

treatment was found. Vaughn, Roberts, Miciak, Taylor, and Fletcher (2019) implemented a multicomponent intervention with fourth and fifth in larger groups of three to six students. The intervention addressed word reading, reading fluency, and reading comprehension over about 68 sessions (30-45 min sessions). Students randomly assigned to the treatment outperformed the students in the business as usual condition on standardized measures of word reading (ES = 0.58) and reading fluency (ES = 0.46). However, no statistically significant differences were noted for decoding or comprehension. Vaughn and colleagues (2016) also examined a small group (4-5 students) multicomponent intervention with fourth graders administered daily over 16 weeks, but found no significant effects for the treatment on standardized measures of decoding, spelling, reading fluency, or reading comprehension; although the treatment and the business as usual groups made significant gains in reading outcomes over the school year on these measures, suggesting the school provided intervention was equally as strong as the treatment in this study.

In contrast, Vadasy and Sanders (2008) provided an intervention with reading fluency, comprehension, and vocabulary components across 80, 30-min sessions to dyads of fourth and fifth graders who demonstrated below grade level rates of reading fluency. They found significant effects on standardized measures of students' vocabulary and comprehension after participation (ES range = 0.27 to 0.50).

Some of the largest effects for an upper elementary intervention were found in a study conducted by O'Connor et al. (2002) examining a multi-component intervention that included 1:1 instruction in phonological awareness, phonics, reading fluency, and comprehension. Teachers identified fourth and fifth grade struggling readers who were performing below a third-grade level in overall reading achievement. Students selected for intervention participated in about 65 sessions of 1:1 instruction (30 min sessions, 4 times per week). Statistically significant

effects of the intervention were reported on standardized measures of word reading, word decoding, fluency, and comprehension with effect sizes ranging from 0.37 to 2.00.

Recently, Miciak et al. (2018) provided upper elementary students experiencing significant difficulties in reading comprehension a multicomponent intervention for a longer duration. The intervention included phonics, reading fluency, comprehension, and vocabulary instruction and was implemented daily for 30-40 min in groups of 4-6 students for 16 weeks. Students were randomly assigned to receive the intervention for either 1 or 2 years. Students in both the 1- and 2-year intervention groups outperformed students in the control group on standardized measures of word reading. In addition, students who received two years of intervention outperformed students who received one year in word reading and word reading fluency. However, no differences were noted between either treatment group and the comparison group on standardized measures of reading comprehension. The authors asserted the need for continued examination of more intensive interventions for students in the upper elementary grades.

Intensive Interventions

These studies included interventions with varying levels of implementation intensity. Implementation intensity includes factors such as the amount of time in intervention (session length and total number of sessions) provided and the instructional group size (Fuchs & Fuchs, 2015). Interventions that feature longer, more frequent, or more total sessions are usually considered more intense (Fuchs & Fuchs, 2015). The extended time in intervention allows for more instruction as well as more student practice and feedback (Vaughn, Wanzek, Murray, & Roberts, 2012) and may assist students with severe reading difficulties and disabilities in making stronger reading gains (Torgesen et al., 2001). Very small groups or individual instruction is

another feature often found in intensive interventions. Smaller group size affords more opportunities for student response and feedback and, therefore, more potential for individualized instruction (Vaughn, Wanzek, Murray, & Roberts, 2012). Research on ideal group sizes for intensive intervention is still unclear. Studies of early elementary students have demonstrated that students who receive intervention in smaller groups make more progress than those who receive intervention in larger groups (Lou et al., 1996; Vaughn et al., 2003). However, as students get older, the change in instructional group size may not be enough to intensify the intervention (Vaughn et al., 2010).

A traditional approach to multi-tiered system of support for preventing reading difficulties involves sequentially moving students through less to more intensive levels of reading intervention, monitoring progress and increasing intervention intensity when students show a lack of adequate progress. However, this approach may leave some students waiting to receive appropriately intensive intervention. An alternate approach calls for immediate placement in an intensive intervention (Tier 3) for students with the most significant deficits, bypassing less intensive (Tier 2) levels of intervention (Al Otaiba et al., 2014; Compton et al., 2012; Fuchs & Fuchs, 2015). Evidence supporting this approach shows supplemental, Tier 2 intervention data may not be necessary for accurate selection of students in need of intensive intervention. Instead, universal screening data and progress monitoring data gathered during classroom instruction, as well as data from norm referenced assessments can be used to accurately place students with the most significant needs immediately into intensive interventions (Compton et al., 2012; Toste et al., 2014).

Voyager Passport

Many schools implement multi-tiered systems of support for students with reading difficulties (Global Scholar, 2011). A variety of reading intervention programs provide materials and instructional guidelines for two intensity levels of support to align with these multi-tiered systems. One such widely used multi-component program is *Voyager Passport* (*Passport;* Voyager Sopris West, 2008*). Passport* is designed for delivery in small group settings to students who are below grade level in reading. Lessons include daily instruction in phonics, vocabulary, comprehension, and fluency featuring explicit instruction and opportunities for modeling, guided, and independent practice. Options for implementation include a standard 30-min implementation and a more intensive 45-min implementation.

The standard implementation includes 30-min lessons divided between two major components, *Word Works* and *Read to Understand*. During *Word Works*, students are taught strategies to decode and/or determine meaning of words. During the first 6 weeks of the program, instruction in *Word Works* focuses on decoding single syllable words with the majority of the lesson time (20-min) devoted to this instruction. In subsequent weeks, the instruction in this component advances to strategies for decoding and determining meaning of multisyllabic words with instruction time in this component decreasing to 5-min. During the second component, *Read to Understand,* instruction focuses on reading and understanding fiction and nonfiction text. During this component, key vocabulary words are taught, routines for activating prior knowledge are practiced, and strategies for increasing reading comprehension are introduced and practiced, aiming for independent use. Brief fluency practice is also included during text reading. During the first 6 weeks of the program, instruction in this component lasts 10-min. After the first 6 weeks, the program time increases for this component to 25-min.

The effectiveness of the standard implementation (less intensive) of *Passport* in improving the reading achievement of fourth grade struggling readers was recently examined (Wanzek et al., 2017). Students who performed below the 30th percentile in reading comprehension were randomly assigned to groups where they either received the researcher provided intervention or typical school services. Students participating in the intervention significantly outperformed those in the comparison group on standardized measures of reading comprehension (ES = .38). These results were moderated by students' word reading skills with students with lower initial word reading skills experiencing less benefit from the intervention. The relatively brief intensive word study provided at the beginning of the *Passport* intervention may not be enough for students with low word recognition to make the same gains as those entering with higher levels of word recognition. The intervention did demonstrate equal benefit on reading comprehension for English learners (EL) and non-EL students (ES = 0.38 and 0.39, respectively) which was encouraging given the greater risk for ELs of late-emerging reading difficulties after Grade 3 (Kieffer, 2010; 2014) when text demands increase markedly. The findings were in line with Baker et al. (2014), who reported small group multicomponent reading interventions can benefit the word reading and comprehension of younger and older ELs.

In addition, Al Otaiba and colleagues (Al Otaiba, Lan, Petscher, Wanzek, & Rivas, 2018) examined outcomes of the standard implementation for students below the 15th percentile in reading comprehension and noted no differences in reading outcomes between the treatment and typical school services conditions, with only small effect sizes on standardized reading comprehension measures (ES = 0.14 to 0.19). The findings related to the students with these more severe reading difficulties suggested they may need a more intensive approach to intervention in order to reap more benefits.

For students with more significant difficulties, the more intensive implementation of *Passport* includes 45-min lessons and is designed to be provided in smaller groups (e.g., 2-3 students). The lessons include the elements of the 30 min standard lesson plus daily fluency practice using either a repeated reading or reading for expression featuring teacher modeling, and a second activity focusing on either additional practice in previously introduced decoding skills or a writing activity.

Purpose

The purpose of the current study was to examine the effectiveness of this intensive implementation of *Passport* for fourth grade students with severe reading comprehension difficulties and disabilities. Given the previous research on a less intensive implementation of *Passport*, we also sought to examine whether there were differences in outcomes based on students' initial word reading levels or their EL status. Specifically, we examined the effects of an intensive implementation of this multi-component reading intervention with the following research questions in mind:

1. What are the effects of an intensive implementation of *Passport* on word reading, word reading efficiency, and reading fluency for fourth grade students with severe reading difficulties?
2. What are the effects of an intensive implementation of *Passport* on the reading comprehension for fourth grade students with severe reading difficulties?
3. Does initial word reading level or English language learner status moderate the effects of the intervention?

Method

Participant Selection and Assignment to Condition

Study participants were recruited from two cohorts of fourth graders in 15 schools (14 zoned (schools assigned to students based on home addresses) schools and 1 charter school) from three school districts. One school district was located in a mid-size city and two districts were located in a large, urban metropolitan area. The mid-size district serves approximately 89,000 in 75 elementary schools with a student population that is 43% Black, 30% White, 23% Hispanic, and 4% Asian. Seventy-five percent of the student qualify for free or reduced-price lunch. The two other district serve more than 200,000 with more than 77% of the student population eligible for free and reduced priced lunch. Student population demographic for these districts were 66-70% Hispanic, 20-22% Black, 8% Asian in one district, and 5% White in both districts. Schools in the districts were notified of the study from their district administrators and the first schools to express interest in participating in the study were included.

All fourth grade students receiving grade level reading instruction with signed parental consent were screened at the end of the first month of school to determine participation eligibility. Parental consent was obtained for 80-100% of students within each of the fourth grade classes. One student did not assent. Of the 906 students screened, only those 306 students with scores at or below the 15th percentile on a standardized measure of reading comprehension (Gates MacGinitie Reading Tests (GRMT; MacGinitie, MacGinitie, Maria, Dreyer, & Hughes, 2006) were selected for participation and then assigned to condition.

Participants were stratified on screening scores at the classroom level (sorted by initial screening score within each teacher so that similar levels of students were paired for random) and randomly assigned to treatment (Passport Intensive) or to a typical school services comparison group. There were 154 participants assigned to the treatment group (Passport Intensive) and 152 participants assigned to the comparison group (Typical School

Services). School districts provided demographic information for each of the study participants. Of the total sample, 48.3% (n = 148) were female and 52.7% (n = 158) were male. In regards to ethnicity, 29.1% (n = 89) identified as Hispanic. The racial composition of the sample was 36.9% (n = 113) Black, 24.2% (n = 74) White, 2.3% (n = 7) American Indian, 2.6% (n = 5) Asian or Pacific Islander, .3 (n = 1) multiracial, 11.1% (n = 34) did not identify a race in addition to their ethnicity, and 23.5% (n = 72) did not report ethnicity nor race. All schools provided instruction only in English with 21.9% (n = 67) of the sample identified as limited English proficiency and eligible for English as a second language support. Fifteen percent (n = 37) of the sample were identified as having a disability, with the majority classified with a learning disability or other health impairment (35% learning disability; 22% other health impairment; 14% autism; 8% developmental/functional delay; 5% emotional/behavior; 5% intellectual; 3% speech; 3% language; 3% hearing; 3% vision) . The majority of students in the sample received free or reduced-price lunch (86.3%). There were no significant differences between study conditions for gender (χ^2 [306] = .102, p = .95), ethnicity (χ^2 [234] = 3.16, p = .21), race (χ^2 [200] = .20, p = .90), socio-economic status (χ^2 [235] = 1.75, p = .19), English learner status (χ^2 [306] = 1.06, p = .30), or special education eligibility (χ^2 [231] = 3.75, p = .15).

Throughout this study, overall and differential attrition resulted in a tolerable threat of bias under both optimistic and cautious assumptions (What Works Clearinghouse, 2017), though it was on the border of being tolerable under optimistic assumptions. Specifically, a total of 46 students (15.0% of total sample) withdrew from their respective schools during the school year. Attrition was similar across groups; it was 18.0% (n = 28) in the treatment group and 11.8% (n = 18) in the comparison group. Students' pretest performance for students who withdrew compared to those who remained in the schools were not significantly different.

Procedures

Student data collection. At pre- and post-treatment (i.e., fall and spring), tests of word reading, word reading efficiency, reading fluency, and reading comprehension were administered by our research team. Pre-assessments were administered at the end of September and beginning of October; post-assessments were administered at the end of April and beginning of May within two weeks of intervention completion. The pre- and posttest battery was counterbalanced by measure and all assessors were trained and required to demonstrate 100% reliability in administration and in scoring prior to conducting assessments in the field. All testers were blind to the students' assignment to condition.

Student reading measures. To assess word reading, we administered the letter-word identification and word attack subtests of the Woodcock-Johnson III Tests of Achievement (WJ-III; Woodcock, McGrew, & Mather, 2001). The letter-word identification is an individually administered subtest requiring students to read real words of increasing difficulty. The test begins with naming individual letters for the very beginning readers. The word attack subtest is an individually administered test and asks students to read pseudowords increasing in decoding difficulty. Test-retest reliability for fourth grade is reported as .81 for letter-word identification, and .85 for word attack (Woodcock et al., 2001).

To measure word reading efficiency, we administered the sight word efficiency and phonemic decoding efficiency subtests of the Test of Word Reading Efficiency, Second Edition (TOWRE-2; Torgesen., Wagner, & Rashotte,, 2012). Both subtests are individually administered. The sight word efficiency subtest provides students with an untimed practice list of sight words followed by 45 seconds to read words of increasing complexity. The procedure was repeated for the phonemic decoding efficiency subtest with a practice list followed by the timed

reading of a list of phonemically regular pseudowords. The number of words read correctly was calculated for each subtest. Test-retest reliabilities range from .83-.96 on these subtests. For fourth graders, the concurrent validity for sight word efficiency and the word identification subset of the Woodcock Reading Mastery Tests-Revised (WRMT-R) is .89. For phonemic decoding efficiency and the word attack subtest of the WRMT-R, concurrent validity is estimated at .86.

To assess students' oral reading fluency, students were assessed individually using the oral reading fluency (ORF) subtest of the Dynamic Indicators of Basic Literacy Skills -6th Edition (DIBELS; Good & Kaminski, 2002). Students read aloud three grade level passages for one minute each and the number of words read correctly was calculated. Test-retest reliability ranges from .92 to .97 and alternate-form reliability across same level passages is reported as .89 to .94 (Good et al., 2004).

We used two assessments of reading comprehension. The first was the group-administered reading comprehension subtest of the GRMT (MacGinitie et al., 2006). The fall reading comprehension scores were also used to select students for the study. The reading comprehension subtest required students to read passages and answer multiple choice questions designed to assess skills including fact identification, inferencing, and drawing conclusions. Test-retest reliabilities are reported as above .85 and construct validity ranges from .79-.81 (MacGinitie et al., 2006). Additionally, students were assessed individually using the Woodcock-Johnson III Tests of Achievement (WJ-III; Woodcock, McGrew, & Mather, 2001) passage comprehension subtest. Students are presented with cloze text, building from sentences to passages, and asked to provide a missing word in each text. Test-retest reliability for fourth grade is reported as .86 (Woodcock et al., 2001). Median concurrent validity correlations for the

passage comprehension are reported as .62 and .79 with the reading comprehension subtests from the Kaufman Test of Educational Achievement and the Wechsler Individual Achievement Test, respectively.

Following pretesting, students assigned to the Passport Intensive condition were provided with approximately 100 sessions of intervention delivered in daily, 45-minute sessions in small groups of 2-3 students. We collected data throughout the school year on all reading instruction received by the participating students, including core classroom instruction (Tier I), and all supplemental intervention provided to both the treatment and the comparison group. We describe this instruction below.

Instruction and Intervention

School-provided core classroom instruction. All participating schools used *Journeys Common Core* (Templeton et al., 2014) as their classroom reading program. We obtained a sample of classroom reading instruction provided to all fourth graders with one observation in the fall and one observation in the spring of the school year. We averaged these observational samples to document the type, amount, and quality of core reading instruction. However, we acknowledge that we may not be able to represent all of the classroom reading instruction the students received over the year with these observations.

Research staff were trained to use the Instructional Content Emphasis Instrument-Revised (ICE-R; Edmonds &Briggs, 2003), a low-inference reading observation tool, to document the amount of time students received instruction in specific reading domains - fluency and comprehension, for example – and the types of instructional groupings used (i.e., whole group, small group, individual). Observers then assigned a global instructional quality rating (e.g., *uses language that is direct and explicit*, *models many examples*, *provides sufficient and varied*

opportunities for practice) ranging from 4 = excellent to 1 = weak, as well as a global student engagement rating (3 = high engagement to 1 = low engagement). Observations were conducted by trained raters who demonstrated an initial minimum of 90% reliability with our gold-standard (Gwet, 2001) prior to each observation time period. An average reliability of 98.7% was obtained across raters and time periods.

In examining the core classroom reading instruction provided to all students in the study, an average of 71.37 min (*SD* = 26.24) of the school day was dedicated to reading instruction, with the largest amount of time focused on comprehension instruction (*M* = 28.43 min; *SD* = 23.06). Other instructional components covered in the core reading instruction included phonics and word recognition (*M* = 0.44 min; *SD* = 2.57), spelling (*M* = .243 min; *SD* = 1.53), vocabulary (*M* = 3.55 min; *SD* = 6.89), and text reading without other instruction (*M* = 8.65 min *SD* = 10.67). Differentiated instruction wherein students worked on different assignments, was observed for an average of 18.86 min (SD = 22.38). On average, 11.20 min (*SD* = 11.65) of the reading instructional time was used for non-instructional activities (e.g., transitions). Instruction was, on average, conducted with the whole class for 43.42 min (*SD* = 24.83), in small groups for 6.41 min (*SD* = 14.58), or with peer pairs for 2.22 min (*SD* = 5.67). Students spent an average 11.79 min (*SD* = 14.42) in independent work that was not part of differentiated instruction.

Passport intensive intervention. For students assigned to the treatment condition, trained project staff provided instruction using the intensive implementation of *Passport*. We delivered the daily, 45-min sessions in small groups of 2-3 students for approximately 100 sessions across the school year. When possible, students with similar screening scores were placed in a group together. *Passport* was designed as a supplemental reading intervention, organized into 12, 10-day adventures focusing on phonics, word recognition, fluency,

vocabulary, and comprehension in each lesson. To monitor students' mastery of content and progress on oral reading fluency, checkpoints are designed at the fifth and tenth lesson of each adventure. Teachers use these to determine whether students are ready to move on to the next lessons or require any elements of reteaching. The two main components of the curriculum, *Word Works* and *Read to Understand*, made up 30 min of the instructional block, with the remaining 15 min comprised of intensive work around one of the lesson's focal skills and fluency building. As noted earlier, *Word Works* emphasized instruction on basic word reading skills (letter/sound identification, decoding, sight word reading, word families, spelling instruction) in the first six weeks for approximately 20 min of the instructional block. In the following weeks, *Word Works* transitioned to a more advanced word study for 5 min of the lesson.

The *Read to Understand* instructional component is intended to be implemented for approximately 10 min during the first six weeks and 25 min thereafter. As described earlier, the *Read to Understand* component focused on comprehension practices that were organized into before, during, and after reading. The text read was carefully constructed to correspond to the phonics and word recognition students were learning in the *Word Works* component. Vocabulary instruction was delivered through definitions, context, and relationships to other words. This instruction occurred prior to text reading and the vocabulary words were pulled from the upcoming text. Comprehension strategies taught included previewing, setting purpose, text structure and evaluation, making inferences and taking perspectives, drawing conclusions, author's purpose, sequencing, main idea, summarizing, independent reading fix-up strategies, teacher and reader questioning, and making connections within and across texts. Text is constructed to correspond with the phonics and word recognition instruction, both the regular

word and irregular word instruction. Thus, text levels built from words with simple letter/sound correspondences to multisyllabic words, building in text length as well.

The final 15 min of each lesson was dedicated to fluency practice using both literary and informational passages as well as response to reading written activities. Students practiced improving their fluency of reading through repeated reading wherein they read a passage several times to see if they could read with more accuracy and a better rate, or to see if they could read the passage similarly to a teacher model. Students were also given facilitated writing activities to reflect and respond to text they had read during the lesson.

Interventionists and training. A total of 26 interventionists provided the intervention instruction. They all had a bachelor's degree and 17 (65%) had also obtained a master's degree. Sixteen of the interventionists were certified teachers. Two intervention teachers were male. Four teachers identified as Hispanic in ethnicity. In terms of race, 24 (92%) teachers were White and two teachers (8%) were Black. All interventionists participated in approximately eight hours of training for *Passport*. The training included orientation to the project, explicit instruction in the intensive intervention and instructional sequence, and opportunity to practice implementation of lessons with feedback.

Once intervention began, each interventionist was provided with twice monthly on-site coaching on implementation from the project coordinators. The interventionists also attended monthly meetings to discuss implementation and challenges.

Intervention implementation and fidelity. Observations were conducted three times (fall, winter, spring) to examine general intervention implementation using to allow direct comparison of the instructional elements in treatment and comparison. The ICE-R measure

described earlier was used for these observations. In addition, we conducted observations of the direct fidelity of the *Passport* program once per month.

Implementation of intervention occurred, on average, for 44.28 min (*SD* = 4.10), with the most amount of time spent on comprehension instruction (*M* = 16.46 min or 37.17% of time). On average, instruction also included vocabulary (*M* = 6.53 min or 14.75% of time), phonics and word recognition (*M* = 6.01 min or 13.57% of time), fluency (*M* = 5.49 min or 12.40% of time), text reading without other instruction (*M* = 7.60 min or 17.16% of time), spelling (m = 1.07 min or 2.42% of time). Less than one minute (*M* = 0.86 min or 1.94% of time) was spent on non-instructional activities.

Within the treatment condition, fidelity of intervention implementation was monitored to determine interventionists' adherence to the lesson requirements. The fidelity form was adapted from forms used in prior research (Vaughn et al., 2003; Vaughn et al., 2010) to match the specific lesson components of *Passport*. The total number of lessons delivered for the intervention group ranged from 93 to 102 sessions (*M* = 98.2, *SD* = 1.99); the median number of sessions attended was 98. Each interventionist was observed once per month by a trained observer (6 observations per tutor). Observers were required to meet reliability with a gold standard (Gwet, 2001) of at least 90% before conducting observations.

Interventionists were rated for implementation on a scale from 0, indicating the interventionist failed to complete the elements of that lesson component, to 3, indicating all or nearly all of the required elements were completed. Fidelity observations also noted student engagement during the lesson and the quality of instruction including the occurrence of ongoing monitoring, redirection of off-task behavior, positive and corrective feedback, organization of materials, and appropriate selection of content for additional support and practice. Mean ratings

for each tutor on implementation ranged from 1.90 to 3.00 across the lesson components (overall average implementation = 2.81). Similarly, mean ratings of student engagement ranged from 2.46 to 3.00 (overall average engagement = 2.93) and the quality of lesson implementation ranged from 2.35 to 3.00 (overall average quality = 2.83).

School-provided supplemental instruction. Students assigned to the comparison condition received the typical school services provided by their elementary schools. *Passport* had not been purchased or implemented by any participating school, nor was there any evidence of any lesson implementation in the comparison instruction. A total of 117 students (n = 58 treatment; n = 59 comparison) received supplemental reading intervention (with similar interventions for both study groups) provided by the school. Most of the school-provided interventions were scheduled daily in sessions ranging from 21-40 min in small groups ranging from 2-5 students. Most school-provided interventions were provided by certified teachers other than the classroom teacher (e.g., reading specialist).

Observers trained on the ICE-R (described above) observed these interventions in the fall, winter, and spring to describe the instruction. Supplemental intervention occurred, on average, for 37.33 min (SD = 12.31), with the most amount of time spent on phonics and word recognition instruction (M = 8.70 min or 23.31% of time). On average, instruction also included comprehension (M = 6.28 min or 16.82% of time), differentiated instruction with students working on different tasks (M = 6.68 min or 17.89% of time), text reading without other instruction (M = 5.05 min or 13.53% of time), vocabulary (M = 2.23 min or 5.97% of time), fluency (M = 1.22 min or 3.26% of time), or other academic instruction such as assistance with classroom work that was not reading (M = 3.64 min or 9.75%). Some limited intervention time was also spent on phonological awareness (M = 0.26 min) and spelling (M = 0.55 min). On

average, 2.70 min was also spent on non-instructional activities. Seventy-eight percent of the interventions were provided in small groups of 2-5 students.

Data Analysis

We approached the analyses of the impact data via a longitudinal, multilevel structural equation modeling (ML-SEM) framework. There are two primary benefits to this approach: First, a structural equation model approach minimizes the negative impact of measurement error upon the estimation of effects and effect sizes in randomized control trials (RCT). Secondly, this approach also takes the nested structure of the data into account (Goddard, Goddard, Kim, & Miller, 2015), which is critical for the proper estimation of the standard errors.

While ML-SEM in the context of RCTs has benefits, there has been a limitation in their usage when the RCT employs a partially nested design (PN-RCT; Baldwin, Bauer, Stice, & Rohde, 2011; Lohr, Schochet, & Sanders, 2014). In PN-RCT designs, only some individuals are nested within a group while others are not. In this study, the partial nesting occurs when students receiving the intervention were all nested within small groups but the comparison students were not. Multiple research groups have proposed models that take into account the partially nested data within the ML-SEM context (Sterba et al., 2014; Wanzek et. al., 2017). In this paper, we modeled PN-RCT data using *n*-level SEM (*n*SEM; Mehta & Neale, 2005; Wanzek et. al., 2017). Within *n*SEM, observed and latent variables may be used across multiple levels, where the definition of level is far more flexible. In nSEM, a level could mean a hierarchal level where the nested structure is defined, or it could also mean the level of a fixed-effects factor (such as treatment, that may have 3 levels). The ability to flexibly define a level allows for the direct estimation of effects under many complex data structures, including PN-RCT.

The model comparison process for the PN-RCT *n*SEM in this study occurred in two phases with four models in each phase. In the first phase, we constrained the intercepts and one loading to be equal across treatment and control groups. That is, at the child-level, we forced the latent factors to have equivalent intercepts at the beginning. For the loadings, we chose one loading from each of the four latent factors (word reading, word reading efficiency, fluency, and reading comprehension) to be constrained to equality both across groups and across time (pretest and posttest). Building the models this way allows for the latent factors to be on the same metric and allows the difference between latent means for each group to be interpretable. The second phase tests whether the prestest-posttest relation between the latent variables is equivalent for both treatment and comparison conditions. This tests the homogeneity of regression assumption. If the relationship differs by study groups, it implies that the pretest latent variable score moderates the relation between treatment group and the posttest latent variable score. Model comparisons were made using the deviance statistic as well as the AIC and BIC indices. A log-likelihood difference test was used for hypothesis testing of model differences.

Results

Descriptives

Means, standard deviations, number of participants, and correlations among the observed variables are presented in Tables 1 and 2. An inspection of the missing data showed missing data rates varied from 0% to 11.6%. Little's missing completely at random (MCAR) test suggested that all missing data met reasonable assumptions for MCAR [$\chi^2_{(86)} = 92.51$, $p > .30$]; thus, using full information maximum likelihood for model estimation was appropriate.

Tables 1 and 2 provides the student performance results on the individual measures of word reading, word reading efficiency, fluency, and reading comprehension. Table 3 presents

the same descriptive information broken down by study condition. Students' scores on the measures were consistently higher at the spring compared to fall. With the exception of the fall to spring correlation of .28 for GMRT reading comprehension, the stability coefficients from fall to spring were all above $r = .66$ and most of them above $r = .85$, suggesting moderate to high stability in relative rank orders of individuals over time.

_n_SEM Results

The first step in fitting these models was to identify which observed variables to load onto which latent variables. In these models, the WJ-III word reading and word attack subtests comprised the word reading factor, the TOWRE phonemic decoding efficiency and sight word efficiency subtests comprised the word reading efficiency factor, the three ORF passages represented the fluency factor, and the GMRT reading comprehension and WJ-III passage comprehension subtest were loaded onto the reading comprehension factor. These factors were fit for the pretest data and the posttest data. For each of these four factors, a multi-group model was fit where the posttest factor was regressed onto the pretest factor for the treatment group and the comparison group. For the treatment group, we also modeled the nesting of students into small groups, and the classroom-level variance was modeled for both groups. The factor loadings in each of these four models were constrained such that for each observed variable, the pretest-posttest and the across-group loadings were constrained to equality, as were the error variances. Once these models were specified, our first set of model comparisons were conducted to investigate whether the regression weight of the posttest onto the pretest was different between the treatment and comparison groups. This is essentially a test of the homogeneity of regression lines assumption for the groups. In all four model comparisons, the fit of the model was not

significantly depreciated by adding this constraint. Therefore, for all the models, this path was set to equality across the groups.

We also tested for baseline equivalence in the latent pretest means. Two sets of four models were fit to the data. In the first set of four models (one for each of the four latent constructs), the latent pretest means were forced to be equivalent across the two groups, as were the two post-test latent means. The regression line from the pretest to the post-test means were also constrained to be equal across the two groups. These models are referred to as Fixed in Table 4. The second set of four models were identical two the previous four, with one exception: The pretest latent variables means were allowed to vary across the groups. These models are referred to in Table 4 as Pre-Freed models. Comparing these models to their respective Fixed models indicates whether the pre-test means were different across the two groups. The results in Table 4 indicated that after a Benjamini-Hochberg linear step up correction, none of the latent constructs was significantly different at pretest.

To address research questions 1 (related to treatment effects on word reading, word reading efficiency, and reading fluency) and 2 (related to treatment effects on comprehension), we tested the difference between the latent posttest means for the treatment and comparison group. The results of these model comparisons are listed in Table 4. The Fixed models were identical to the Fixed models used to test for latent pretest differences, while in the Post-Freed models the latent means were allowed to be estimated separately for the treatment and comparison groups. Because four different statistical tests were conducted, we adjusted our alpha level using the Benjamini-Hochberg linear step up procedure. In relation to research question 1, two of the three constructs demonstrated a treatment effect where the treatment group was observed to have a significantly higher latent posttest mean than the comparison group: the

word reading factor (-2ll difference = 5.11, 1, p = .014, d = 0.25) and the word reading efficiency factor (-2ll difference = 4.62, 1, p = .018, d = 0.19); no significant differences were found for the reading fluency factor. In relation to research question 2, no significant differences were found between the treatment and comparison groups on reading comprehension. Model fit information is presented in Table 4.

Moderators

In the final set of models to address research question 3, we inspected whether the treatment effects might be moderated by student's initial word reading level or EL status. To test this we used the models described above to test the treatment effect, but added the fall WJIII word identification score or EL status separately to see if the relationship between the potential moderator and outcome differed by study group. In these models, neither the initial WJIII word identification level nor EL status demonstrated a differential relationship with the posttest latent outcome. Therefore, we could not conclude that these indicators had a moderating effect on outcome; in other words the treatment had similar effects regardless of students' initial word reading or their EL status.

Discussion

The purpose of this study was to investigate the effectiveness of the intensive implementation of a multicomponent intensive intervention for fourth grade students with severe reading difficulties. Students with reading comprehension achievement below the 15%ile in the fall of fourth grade were randomly assigned to receive the intensive *Passport* intervention or to continue with typical school services. Students received intervention throughout the school year.

We first examined the effects of the intensive implementation on students' foundational skills that support reading comprehension. We found students receiving the intensive *Passport*

implementation outperformed their peers receiving typical school services in word reading and word reading fluency. Thus, an important implication from this study is that the intensive intervention helped these students to be able to read more words and to recognize these words more efficiently. This is important in light of the previous work that found that students with the lowest word reading skills did not obtain the same reading comprehension benefit from the less intensive intervention as their peers with higher word reading skills (Wanzek et al., 2017). The current sample of students identified with the most significant reading comprehension difficulties demonstrated low word reading skills prior to intervention, and the more intensive intervention helped them to accelerate their learning in these foundational skills. Al Otaiba et al. (2018) noted that students with significant reading comprehension difficulties receiving the less intensive, standard implementation of *Passport* did not significantly outperform peers receiving typical school services in these foundational skills. In contrast, the intensive implementation accelerated students' word reading abilities, one important element of reading achievement. The intensive implementation allowed students to have more time in application of their word reading instruction with additional fluency text reading as well as additional time practicing or reviewing decoding skills. This additional time may have further solidified the phonics and word recognition instruction beyond the standard implementation for these students with more significant needs. Thus, the intensive implementation helped to further student reading abilities in ways that may not have been possible with the less intensive implementation.

In relation to the simple view of reading (Gough & Tunmer, 1986), these basic reading skill improvements may ultimately assist students with improving their efficient reading and understanding of text. Importantly, students in the intensive implementation group did not significantly outperform students in the typical school services in reading fluency, indicating

students may need more foundational skill improvements before gains in fluency can be realized. Notably, students in both groups demonstrated average fluency scores that were still well below expected grade level benchmarks at the end of the intervention, and they were also falling further behind average achieving peers in this area.

We also examined the effect of the intensive intervention on students' reading comprehension. We found these students with significant reading comprehension difficulties who received the intensive intervention did not demonstrate accelerated learning in reading comprehension over typical school services. Instead, students in both study groups made similar gains in reading comprehension during the school year. The accelerated learning in word reading and word reading efficiency noted in this study was not enough to significantly accelerate reading comprehension outcomes beyond the typical school services, despite the extended time in comprehension instruction noted in the *Passport* intervention compared to the school-provided interventions. The previous study examining the less intensive implementation of *Passport* noted significant gains in reading comprehension on average for students beginning fourth grade below the 30th percentile in reading comprehension (Wanzek et al., 2017). However, in that study the authors also noted students with higher initial word reading scores performed better in reading comprehension after the less intensive implementation than those with lower word reading skills. The current study suggests that even the more intensive implementation did not provide enough intensity to accelerate reading comprehension outcomes for this sample of students who averaged lower initial word reading skills. A more intensive or longer duration intervention may be needed before these benefits on reading comprehension can be noted. Notably, the average spring comprehension scores for the current sample mirror the fall scores for the higher level sample in the previous study of the less intensive intervention. In a

study of older students (Grades 5-10), Wang and colleagues (Wang, Sabatini, O'Reilly, & Weeks, 2019) found a decoding threshold below which there no relation with reading comprehension. Wang et al. suggest that improvements in decoding may not immediately lead to reading comprehension progress if the intensity of the intervention is not sufficient to put students above the decoding threshold. Continued intensive intervention may, therefore, assist these students in making further gains including in the area of comprehension now that they have reached these levels.

Our findings generally match the findings of a previous study of fourth grade students with significant reading difficulties (Miciak et al., 2017). Miciak and colleagues noted significant differences on word reading outcomes in favor of a multicomponent reading intervention that was delivered in 30-40 min sessions daily to groups of 4-6 students. Some students received the intervention for 2 school years and these students also accelerated their learning over typical school services in word reading efficiency. The current study implemented an intensive intervention with smaller groups (2-3 students) in 45 min sessions and obtained similar outcomes with one year of intervention. The lack of accelerated learning in reading comprehension in both of these studies with samples of similar initial reading achievement suggests the continued need for research in intensive intervention to improve outcomes for students with the most significant difficulties. Miciak et al. also suggest a possible need to examine more complex theoretical models of reading instruction when it comes to the needs of students with severe reading difficulties. For example, psychosocial elements such as self-regulation, motivation, effort and deliberate practice, or student mindset towards goals may need to be carefully entwined within reading instruction and student reading practice to improve

academic achievement for these students (Dweck, 2006; Okolo, 1992; Page-Voth & Graham, 1999; Robertson, 2000; Snipes et al., 2012).

We followed up our average effects analysis with examination of two moderators of outcomes that have been considered in prior work, initial student word reading level and English learner status. In the study of the standard implementation of *Passport* differences in outcomes by initial word reading were reported (Wanzek et al., 2017). The sample of students with more severe reading difficulties in the current study already demonstrated lower word reading skills on average. We did not find further differentiation on any outcome based on initial word reading levels. Regardless of initial word reading levels, students in the intensive intervention implementation outperformed students in typical school services on word reading and word reading efficiency, but performed similarly to the comparison group on reading fluency and reading comprehension.

Though ELs may struggle in the upper elementary grades with the increase in complexity of text, an encouraging implication from the current study is that similar findings were achieved for both EL and non-EL students in intervention outcomes. EL and non-EL students benefitted similarly from the intensive intervention suggesting the intervention may be appropriate for accelerating word reading achievement for students learning the English language.

Limitations

The findings of this study are relevant to the context of the study. Each of the interventionists in our study worked solely in the schools implementing interventions, but they were hired and trained by our research team. The intervention was implemented with a relatively high degree of fidelity which allowed us to examine the efficacy of the intensive intervention.

We cannot provide information on intervention fidelity for other types of interventionists or determine the effects of the intervention if it is implemented at lower levels of fidelity.

Our sample included schools with very high percentages of students from low socioeconomic backgrounds. As a result, the findings may not generalize to schools with student population from higher socioeconomic backgrounds. Similarly, the vast majority of students identified as ELs were Hispanic and the findings may not generalize to ELs from other language backgrounds. In addition, we did not examine outcomes related to the language comprehension aspect of the Simple View of Reading (listening comprehension, vocabulary, oral language, background knowledge) that could also provide additional insight to the reading comprehension difficulties of the current sample.

Implications and Future Research

Overall, the intensive implementation allowed students with significant reading comprehension difficulties to accelerate their word reading abilities above and beyond typical school services. These findings along with the previous work noting the difficulties of students with low word reading skills to fully benefit from the less intensive implementation, suggest that students with low word reading skills may benefit from the more intensive implementation at least in terms of their foundational skills. As noted previously, the less intensive intervention may not be appropriate for these students, and immediate placement in the intensive implementation may be recommended. Although the intensive implementation for one school year was successful in accelerating students' word reading learning, students still had important deficits in word reading, word reading efficiency, and the acceleration they experienced was not enough to also accelerate their reading fluency or reading comprehension. Thus, a more intensive word reading component than that available in *Passport* may be needed for these

students to accelerate their learning enough to reap the benefits in fluency and comprehension. However, more research in this area to determine appropriate ways to intensify interventions for effective and efficient gains for students with severe difficulties is needed.

In 2010, Vaughn, Denton, and Fletcher suggested that students with the most significant needs may need intensive interventions immediately rather than demonstrating a nonresponse to less intensive interventions first. Our study provides evidence that students with severe reading difficulties may benefit from immediate placement in intensive intervention. It is also clear that the intervention needs to be more intense to meet more of their needs, including implementing a more intensive intervention earlier in their schooling, providing a longer duration in the intensive intervention, and/or increasing the intensity of instruction (e.g., more time, more explicit, more systematic, more opportunities to respond and receive feedback) in one or more of the reading components. It is also possible that these students require additional intervention components, goal-setting, motivation, mindset training, in order to more deliberately apply their improved word learning skills in ways that benefit reading comprehension. Future research examining the key components, combination of components, and embedding of psychosocial aspects into reading intervention is needed to better understand the needs of students with the most significant reading difficulties and disabilities.

References

Ahmed, Y., Francis, D. J., York, M., Fletcher, J. M., Barnes, M., & Kulesz, P. (2016). Validation of the direct and inferential mediation (DIME) model of reading comprehension in grades 7 through 12. *Contemporary Educational Psychology, 44*, 68-82. doi:10.1016/j.cedpsych.2016.02.002

Al Otaiba, S., Lan, P., Petscher, Y., Wanzek, J., & Rivas, B. (2018). I'm not throwing away my shot: What Alexander Hamilton can tell us about standard reading interventions. *Learning Disabilities Research and Practice, 33*, 156-167.

Al Otaiba, S., Connor, C. M., Folsom, J. S., Greulich, L., Wanzek, J., & Schatschneider, C. (2014). To wait in tier 1 or intervene immediately: A randomized experiment examining first grade response to intervention in reading. *Exceptional Children, 81*, 11-27.

Baker, S., Lesaux, N., Jayanthi, M., Dimino, J., Proctor, C. P., Morris, J., . . . Newman-Gonchar, R. (2014). *Teaching academic content and literacy to English learners in elementary and middle school* (NCEE 2014-4012). Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education.

Baldwin, S. A., Bauer, D. J., Stice, E., & Rohde, P. (2011). Evaluating models for partially clustered designs. *Psychological Methods, 16,* 149-165. doi: 10.1037/a0023464

Cho, E., Capin, P., Roberts, G., Roberts, G. J., & Vaughn, S. (2019). Examining sources and mechanisms of reading comprehension difficulties: Comparing English learners and non-English learners within the simple view of reading. *Journal of Educational Psychology, 111*, 982-1000.

Cirino, P. T., Romain, M. A., Barth, A. E., Tolar, T. D., Fletcher, J. M., & Vaughn, S. (2013). Reading skill components and impairments in middle school struggling readers. *Reading and Writing*, 26, 1059-1086.

Compton, D. L., Gilbert, J. K., Jenkins, J. R., Fuchs, D., Fuchs, L. S., Cho, E., … & Bouton, B. (2012). Accelerating Chronically Unresponsive Children to Tier 3 Instruction: What Level of Data Is Necessary to Ensure Selection Accuracy? *Journal of Learning Disabilities*, *45*, 204–216. https://doi.org/10.1177/0022219412442151

Cromley, J., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology, 99*, 311–325. doi:10.1037/0022-0663.99.2.311

Dweck, C. S. (2006). *Mindset: The new psychology of success*. New York, NY: Random House.

Flynn, L. J., Zheng, X., & Swanson, H. L. (2012). Instructing struggling older readers: A selective meta-analysis of intervention research. *Learning Disabilities Research and Practice, 27*, 21-32. https://doi.org/10.1111/j.1540-5826.2011.00347.x

Fuchs, D., & Fuchs, L. S. (2015). Rethinking service delivery for students with significant learning problems: Developing and implementing intensive instruction. *Remedial and Special Education*, *36*, 105–111. https://doi.org/10.1177/0741932514558337

Global Scholar (2011). Response to intervention (RTI) adoption survey. Retrieved April 13, 2018 from learn.globalscholar.com/RTISurvey.html

Goddard, R., Goddard, Y., Kim, E. S., & Miller, R. (2015). A theoretical and empirical analysis of the roles of instructional leadership, teacher collaboration, and collective efficacy beliefs in support of student learning. *American Journal of Education*, *121*, 501-530.

Gough, P., & Tunmer, W. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7,* 6-10.

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

Kent, S. C., Wanzek, J., & Al Otaiba, S. (2017). Reading instruction for fourth-grade struggling readers and the relation to student outcomes. *Reading & Writing Quarterly: Overcoming Learning Difficulties, 33*, 395-411. http://dx.doi.org/10.1080/10573569.2016.1216342

Kieffer, M. J. (2010). Socioeconomic status, English proficiency, and late-emerging reading difficulties. *Educational Researcher, 39*, 484-486. doi: 10.3102/0013189X10378400.

Kieffer, M. J. (2014). Morphological awareness and reading difficulties in adolescent spanish-speaking language minority learners and their classmates. *Journal of Learning Disabilities, 47*, 44-53.

Lohr, S., Schochet, P.Z., & Sanders, E. (2014). *Partially Nested Randomized Controlled Trials in Education Research: A Guide to Design and Analysis (NCER 2014-2000)*. Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncer/pubs/20142000/pdf/20142000.pdf /

Mehta, P. D., & Neale, M. C. (2005). People are variables too: Multilevel structural equations modeling. *Psychological Methods, 10*, 259-284. doi: 10.1037/1082-989X.10.3.259

Miciak, J., Roberts, G., Taylor, W. P., Solis, M., Ahmed, Y., Vaughn, S., & Fletcher, J. M. (2018). The effects of one versus two years of intensive reading intervention implemented with late elementary struggling readers. *Learning Disabilities Research and Practice*, *33*, 24–36. https://doi.org/10.1111/ldrp.12159

O'Connor, R. E., Bell, K. M., Harty, K. R., Larkin, L. K., Sackor, S. M., & Zigmond, N. (2002). Teaching reading to poor readers in the intermediate grades: A comparison of text difficulty. *Journal of Educational Psychology*, *94*, 474–485. https://doi.org/10.1037//0022-0663.94.3.474

Okolo, C. M. (1992). The effects of computer-based attribution retraining on the attributions, persistence, and mathematics computation of students with learning disabilities. Journal of Learning Disabilities, 25, 327-334.

Page-Voth, V., & Graham, S. (1999). Effects of goal-setting and strategy use on the writing performance and self-efficacy of students with writing and learning problems. Journal of *Educational Psychology, 91*, 230-240.

Ritchey, K. D., Silverman, R. D., Montanaro, E. A., Speece, D. L., & Schatschneider, C. (2012). Effects of a tier 2 supplemental reading intervention for at-risk fourth-grade students. *Exceptional Children*, *78*, 318–334. https://doi.org/10.1177/001440291207800304

Robertson, J. S. (2000). Is attribution training a worthwhile classroom intervention for K-12 students with learning difficulties? *Educational Psychology Review, 12*, 111-134.

Snipes, J., Fancsali, C., & Stoker, G. (2012). Student academic mindset interventions: A review of the current landscape. Columbia, MD:Impaq International

Sterba, S. K., Preacher, K. J., Forehand, R., Hardcastle, E. J., Cole, D. A., & Compas, B. E. (2014). Structural equation modeling approaches for analyzing partially nested data. *Multivariate behavioral research*, *49*, 93-118.

Therrien, W. J., Wickstrom, K., & Jones, K. (2006). Effect of a combined repeated reading and question generation intervention on reading achievement. *Learning Disabilities Research and Practice*, *21*(2), 89–97. https://doi.org/10.1111/j.1540-5826.2006.00209.x

Toste, J. R., Compton, D. L., Fuchs, D., Fuchs, L. S., Gilbert, J. K., Cho, E., … & Bouton, B. D. (2014). Understanding unresponsiveness to tier 2 reading intervention. *Learning Disability Quarterly*, *37*(4), 192–203. https://doi.org/10.1177/0731948713518336

Torgesen, J. K., Alexander, A. W., Wagner, R. K., Rashotte, C. A., Voeller, K. K., & Conway, T. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities*, *34*, 33-58. doi:10.1177/002221940103400104

Torgesen, J. K., Houston, D. D., Rissman, L. M., Decker, S. M., Roberts, G., Vaughn, S…., & Lesaux, N. (2007*). Academic literacy instruction for adolescents: A guidance document from the Center on Instruction*. Portsmouth, NH: RMC Research Corporation, Center on Instruction.

Vadasy, P. F., & Sanders, E. A. (2008). Benefits of repeated reading intervention for low-achieving fourth- and fifth-grade students. *Remedial and Special Education*, *29*, 235–249. https://doi.org/10.1177/0741932507312013

Vaughn, S., Denton, C. A., & Fletcher, J. M. (2010). Why intensive interventions are necessary for students with severe reading difficulties. *Psychology in the Schools. Special Issue: Literacy and Disabilities, 47*, 432-444.

Vaughn, S., Linan-Thompson, S., Kouzekanani, K., Bryant, D. P., Dickson, S., & Blozis, S. A. (2003). Grouping for reading instruction for students with reading difficulties. *Remedial and Special Education, 24*, 301–315.

Vaughn, S., Roberts, G. J., Miciak, J., Taylor, P., & Fletcher, J. M. (2019). Efficacy of a word- and text-based intervention for students with significant reading difficulties. *Journal of Learning Disabilities, 52,* 31-44.

Vaughn, S., Wanzek, J., Murray, C. S., & Roberts, G. (2012). Intensive interventions for students struggling in reading and mathematics: A practice guide. Portsmouth, NH: RMC Research Corporation, Center on Instruction.

Vaughn, S., Wanzek, J., Wexler, J., Barth, A., Cirino, P., Fletcher, J. M…., & Francis, D. J. (2010). The relative effects of group size on reading progress of older students with reading difficulties. *Reading and Writing: An Interdisciplinary Journal, 23,* 931-956.

Voyager Sopris Learning (2008). *Passport*. Dallas, TX.

Wang, Z., Sabatini, J., O'Reilly, T., & Weeks, J. (2019). Decoding and reading comprehension: A test of the decoding threshold hypothesis. *Journal of Educational Psychology, 111*, 387-401.

Wanzek, J., Petscher, Y., Otaiba, S. Al, Rivas, B. K., Jones, F. G., Kent, S. C., … & Mehta, P. (2017). Effects of a year long supplemental reading intervention for students with reading difficulties in fourth grade. *Journal of Educational Psychology*, *109*, 1103–1119.

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. https://doi.org/10.1007/s11145-009-9179-5

What Works Clearinghouse. (2014). *Procedures and standards handbook* (Version 4.0). Washington, DC: U.S. Department of Education. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf

Table 1.
Descriptive statistics and correlations for Reading Comprehension and Decoding measures

Variable	1	2	3	4	5	6	7	8
1. Fall GMRT RC	1.00							
2. Fall WJ PC	.24	1.00						
3. Fall WJ LWID	.24	.78	1.00					
4. Fall WJ WA	.22	.66	.81	1.00				
5. Spring GMRT RC	.28	.32	.33	.32	1.00			
6. Spring WJ PC	.23	.65	.67	.50	.33	1.00		
7. Spring WJ LWID	.24	.72	.85	.75	.33	.68	1.00	
8. Spring WJ WA	.21	.54	.69	.72	.26	.50	.77	1.00
Mean	419.72	472.32	471.35	481.97	444.93	481.72	482.16	488.1
SD	21.65	14.12	23.49	19.81	22.89	11.3	21.6	16.32
N	293	292	292	292	260	260	259	260

Note. GMRT RC = Gates-McGinitie Reading Comprehension, WJ PC = WJ-III Passage Comprehension, WJ LWID = WJ-III Letter Word Identification, WJ WA = WJ-III Word Attack

Table 2.
Descriptive statistics and correlations for Reading Efficiency and Fluency measures

Variable	1	2	3	4	5	6	7	8	9	10
1. Fall TOWRE PDE	1.00									
2. Fall TOWRE SWE	.76	1.00								
3. Fall ORF P1	.72	.83	1.00							
4. Fall ORF P2	.76	.86	.92	1.00						
5. Fall ORF P3	.76	.85	.91	.93	1.00					
6. Spr TOWRE PDE	.84	.73	.69	.73	.74	1.00				
7. Spr TOWRE SWE	.69	.83	.78	.80	.79	.76	1.00			
8. Spr ORF P1	.72	.85	.84	.86	.86	.77	.87	1.00		
9. Spr ORF P2	.70	.81	.83	.84	.85	.74	.88	.91	1.00	
10. Spr ORF P3	.73	.77	.85	.86	.84	.77	.84	.89	.91	1.00
Mean	76.8	78.47	64.91	59.46	52.42	80.69	83.15	78.75	78.00	68.53
SD	14.26	13.23	26.77	25.56	25.22	14.93	13.81	28.09	32.29	24.65
N	292	292	292	291	291	260	260	259	260	260

Note. TOWRE PDE= Test of Word Reading Efficiency: Phonemic Decoding Efficiency Subtest. TOWRE SWE= Test of Word Reading Efficiency: Sight Word Efficiency Subtest. ORF= DIBELS Oral Reading Fluency – Passages 1, 2, and 3.

Table 3.
Descriptive statistics of measures by condition

	Passport			Comparison		
Measure	N	M	SD	N	M	SD
Fall GMRT RC scaled score	141	420.15	20.86	152	419.32	22.41
Fall WJ PC W score (standard score)	141	472.41 (80.10)	12.99 (9.84)	151	472.23 (80.00)	15.13 (11.50)
Fall WJ LWID W score (standard score)	141	471.72 (89.29)	22.90 (11.10)	151	471.00 (88.91)	24.11 (11.79)
Fall WJ WA W score (standard score)	141	482.32 (90.94)	21.40 (12.73)	151	481.64 (90.50)	18.27 (11.07)
Fall TOWRE PDE standard score	141	77.52	14.70	151	76.13	13.85
Fall TOWRE SWE standard score	141	78.63	13.17	151	78.33	13.33
Fall ORF P1 raw score	141	65.80	25.81	151	64.08	27.69
Fall ORF P2 raw score	141	60.41	25.91	150	58.58	25.28
Fall ORF P3 raw score	141	52.68	24.43	150	52.18	26.02
Spr GMRT RC scaled score	126	445.08	23.16	134	444.79	22.71
Spr WJ PC W score (standard score)	126	482.60 (84.97)	10.11 (7.64)	134	480.89 (83.69)	12.29 (9.42)
Spr WJ LWID W score (standard score)	125	485.22 (91.05)	20.63 (10.49)	134	479.32 (87.99)	22.16 (11.07)
Spr WJ WA W score (standard score)	126	490.00 (92.13)	15.34 (9.35)	134	486.31 (89.88)	17.05 (10.15)
Spr TOWRE PDE standard score	126	81.96	14.11	134	79.49	15.62
Spr TOWRE SWE standard score	126	84.71	13.63	134	81.67	13.87
Spr ORF P1 raw score	126	80.61	26.39	133	76.99	29.61
Spr ORF P2 raw score	126	81.13	31.68	134	75.05	32.70
Spr ORF P3 raw score	126	71.38	23.78	134	65.85	25.24

Note. GMRT RC = Gates-McGinitie Reading Comprehension, WJ PC = WJ-III Passage Comprehension, WJ LWID = WJ-III Letter Word Identification, WJ WA = WJ-III Word Attack,

TOWRE PDE= Test of Word Reading Efficiency: Phonemic Decoding Efficiency Subtest. TOWRE SWE= Test of Word Reading Efficiency: Sight Word Efficiency Subtest. ORF= DIBELS Oral Reading Fluency – Passages 1, 2, and 3.

Table 4
CFA model fit comparison for latent reading comprehension, word reading, and vocabulary

Outcome	Model	-2LL	*df*	AIC	BIC	Δ-2LL	Δ*df*	*p*	*d*
RC	Fixed	9251.1	10	9271	9321				
	Pre-Freed	9250.5	11	9273	9328	.64	1	.425	
	Post-Freed	9250.3	11	9272	9327	.68	1	.344	.09
Decoding	Fixed	8963.3	10	8983	9033				
	Pre-Freed	8957.9	11	8980	9035	5.39	1	.020	
	Post-Freed	8958.2	11	8980	9035	5.11	1	.014	.25
Efficiency	Fixed	8108.5	10	8128	8178				
	Pre-Freed	8106.5	11	8127	8182	2.00	1	.157	
	Post-Freed	8103.9	11	8126	8181	4.62	1	.018	.19
Fluency	Fixed	13175.8	11	13198	13257				
	Pre-Freed	13175.0	12	13199	13264	.88	1	.348	
	Post-Freed	13172.4	12	13196	13261	3.48	1	.038	.08

Note. -2LL = -2*log likelihood, AIC =Akaike Information Criteria, BIC = Bayes Information Criteria. Effect size d computed by computing the difference between the two latent posttest means and dividing by the pooled observed posttest standard deviations. Pre-Freed is a model where the pre-test latent means were free to vary. Post-Freed is a model where the post-test latent means were free to vary.