

## Examining the Average and Local Effects of a Standardized Treatment for Fourth Graders With Reading Difficulties

Jeanne Wanzek, Yaacov Petscher, Stephanie Al Otaiba, Shawn C. Kent, Christopher Schatschneider, Martha Haynes, Brenna K. Rivas & Francesca G. Jones

To cite this article: Jeanne Wanzek, Yaacov Petscher, Stephanie Al Otaiba, Shawn C. Kent, Christopher Schatschneider, Martha Haynes, Brenna K. Rivas & Francesca G. Jones (2016) Examining the Average and Local Effects of a Standardized Treatment for Fourth Graders With Reading Difficulties, Journal of Research on Educational Effectiveness, 9:sup1, 45-66, DOI: [10.1080/19345747.2015.1116032](https://doi.org/10.1080/19345747.2015.1116032)

To link to this article: <https://doi.org/10.1080/19345747.2015.1116032>



Accepted author version posted online: 08 Jan 2016.  
Published online: 20 Jun 2016.



Submit your article to this journal [↗](#)



Article views: 314



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 8 View citing articles [↗](#)

## Examining the Average and Local Effects of a Standardized Treatment for Fourth Graders With Reading Difficulties

Jeanne Wanzek<sup>a</sup>, Yaacov Petscher<sup>b</sup>, Stephanie Al Otaiba<sup>c</sup>, Shawn C. Kent<sup>d</sup>, Christopher Schatschneider<sup>b</sup>, Martha Haynes<sup>b</sup>, Brenna K. Rivas<sup>c</sup>, and Francesca G. Jones<sup>c</sup>

### ABSTRACT

The present study used a randomized control trial to examine the effects of a widely used multicomponent Tier 2-type intervention, Passport to Literacy, on the reading ability of 221 fourth graders who initially scored at or below the 30th percentile in reading comprehension. Intervention was provided by research staff to groups of 4–7 students for 30 min, 4 days a week throughout the school year ( $M = 90.45$  lessons). Tier 1 instruction was observed to be of generally high quality and intervention fidelity was strong. Findings revealed small, average effects ( $ES = .14-.28$ ) in favor of intervention students on standardized measures of comprehension, but no effects on word reading or fluency measures. Exploratory analyses indicated that intervention effects may differ by students' comprehension abilities. Implications for intervention implementation and directions for future research are discussed.

### KEYWORDS

reading  
intervention  
reading difficulties

There are many students entering fourth grade who struggle significantly with reading. It is estimated that 65% of fourth-grade students cannot read at proficient levels, with 32% of the fourth-grade population unable to read at or above basic levels of understanding (National Center for Education Statistics, 2013). The results are particularly troubling for students who manifest late-emerging reading difficulties (Compton, Fuchs, Fuchs, Elleman, & Gilbert, 2008; Leach, Scarborough, & Rescorla, 2003), putting them at risk for identification with disabilities. Nationally, the number of students served in special education with a learning disability increases by 22% in the upper elementary grades (Office of Special Education and Rehabilitative Services, 2013).

However, the research on reading interventions for upper elementary students is limited in comparison to earlier grades, leaving educators with a dearth of information to make key instructional decisions by fourth grade. A synthesis of the research for students with reading difficulties in fourth and fifth grades located a mere 24 studies published between 1988 and 2006 (Wanzek, Wexler, Vaughn, & Ciullo, 2010). An additional four studies that would have met the synthesis criteria have been published since that time. The large majority of the

**CONTACT** Jeanne Wanzek ✉ [jeanne.wanzek@vanderbilt.edu](mailto:jeanne.wanzek@vanderbilt.edu) 🏢 Department of Special Education, Vanderbilt University, 110 Magnolia Circle, Nashville, TN 37203, USA.

<sup>a</sup>Vanderbilt University, Nashville, Tennessee, USA

<sup>b</sup>Florida State University, Tallahassee, Florida, USA

<sup>c</sup>Southern Methodist University, Dallas, Texas, USA

<sup>d</sup>University of Houston, Houston, Texas, USA

Color versions of one or more of the figures in the article can be found online at [www.tandfonline.com/uree](http://www.tandfonline.com/uree).

© 2016 Taylor & Francis Group, LLC

studies ( $n = 24$ ) examined intervention in a single reading component (e.g., main idea strategy instruction), and most of the studies utilized researcher-developed measures to report effects of the instruction. In fact, only four studies that included comprehension instruction as part of the intervention measured outcomes on norm-referenced tests of comprehension.

Four examinations of multicomponent reading interventions at the upper elementary level have been conducted previously (O'Connor et al., 2002; Ritchey, Silverman, Montanaro, Speece, & Schatschneider, 2012; Therrien, Wickstrom, & Jones, 2006; Vadasy & Sanders, 2008). Three of these multicomponent studies demonstrated moderate to large effects on norm-referenced measures of comprehension. These interventions also demonstrated the largest effects on reading of any study at the upper elementary level, suggesting the potential of multicomponent interventions for improving reading outcomes for struggling readers in these grades. In contrast, Ritchey et al. (2012) implemented a multicomponent supplemental reading intervention for students with reading difficulties in fourth grade and found moderate effects only on the near-transfer measures (science content knowledge and comprehension strategy knowledge and use). However, no significant differences on standardized measures of decoding, word reading, decoding efficiency, word reading efficiency, or comprehension were noted. This most recent study provided a relatively brief (24 sessions) intervention in comparison to the previous work. Thus, there may not have been sufficient time for students to achieve skill mastery that generalized to the broader measures.

Despite the limited research on the impacts of multicomponent intervention for upper elementary students, schools overwhelmingly indicate the use of multicomponent published programs in the interventions they select (e.g., Florida Department of Education, 2010). Few of these programs have been tested for efficacy. Equally problematic is the lack of information on the average effects of upper elementary reading interventions on global reading outcomes such as standardized comprehension measures. Thus, in the current study we conducted a preliminary study of a widely used multicomponent, small-group intervention, Passport to Literacy, and its relationship to various student outcomes, including standardized measures.

## Passport to Literacy

Passport to Literacy is a widely used, supplemental multicomponent intervention program designed to improve the reading outcomes of struggling readers. Passport to Literacy applies principles of behavioral learning theory and cognitive psychology (Flavell, 1992; Palincsar & Brown, 1984). The program provides instruction in a sequential, hierarchical series progressing from the foundational skills to higher level thinking in each lesson. The program includes several practices to address student difficulties in phonological processing, a major cause of reading disabilities (Liberman, Shankweiler, & Liberman, 1989; Ransby & Swanson, 2003). The program is not built upon the assumption that accurate and fluent word reading alone leads to comprehension. The main emphasis of the program is on strategies for gaining understanding, building students' conceptual and background knowledge, and teaching students to interact with the text to gain meaning and monitor comprehension.

Passport is currently used in more than 8,000 schools in every state in the United States with more than one million children with reading difficulties receiving the intervention. The program has also been endorsed by the Council of Administrators of Special Education. Although Passport to Literacy is widely used, there is currently no independent research on

the program's effectiveness, no causal studies have been conducted, and there are no studies examining outcomes on standardized measures of reading. The current study sought to address each of these identified gaps.

## Upper Elementary Response to Intervention Context

Findings from intervention research are most applicable to practitioners when school context is taken into account. Currently, many schools have adopted multitiered models as a framework for implementing reading intervention. In typical RTI models, initial interventions (Tier II) are typically provided to all students who are identified with a reading difficulty (Gersten et al., 2008). These interventions tend to be standardized and multicomponent, and, at various grade levels, have demonstrated the ability to prevent a reading difficulty from becoming more serious for some students (Vaughn et al., 2010; Wanzek et al., *in press*). Passport to Literacy is a standardized treatment protocol Tier II-type intervention.

In typical multitiered models, more intense interventions (Tier III) are then provided for students who do not respond well to the initial interventions. However, this model of implementation is based on research that has largely been conducted at K–3 grade levels. Recommendations at the middle school level are to consider immediate placement in more intensive interventions (without first examining response to less intensive interventions) for students with severe difficulties (Vaughn et al., 2010). The rationale is that in the secondary grades students with severe reading difficulties need intensive interventions immediately to be able to make adequate gains. It is possible that some students in the upper elementary grades would also be better served with immediate placement in intensive interventions. There is currently no data on more local effects of upper elementary interventions to inform which students can be served well through Tier II-type interventions and which students may benefit most from immediate placement in more intensive interventions. In this study, we also sought to systematically examine the variation in student outcomes after the Tier II intervention, specifically for whom the Passport to Literacy intervention was more or less effective. This information can guide decisions regarding which students in the upper elementary grades may benefit most from typical Tier II interventions.

## Study Purpose

The current study was the initial work in a multiyear project to examine the efficacy of Passport to Literacy as a supplemental intervention within an RTI framework. The specific purpose of this study was to provide preliminary data regarding the average and local effects of Passport to Literacy for fourth-grade students with reading comprehension difficulties. Specifically, we examined whether students with reading difficulties receiving the intervention outperformed students receiving typical school services (business as usual) in decoding, word recognition, fluency, or reading comprehension. We also conducted exploratory analyses to examine whether effects were differentiated for students with varying levels of reading comprehension ability. With these preliminary data, we sought to explore whether Passport's emphasis on comprehension instruction would translate into impacts on comprehension outcomes, and whether students with higher reading comprehension levels would differentially benefit from this Tier II-type intervention.

Method

Participants

The participants for this study were 221 fourth-grade students in 10 public elementary schools across four school districts in two states who scored at the 30th percentile or below on the reading comprehension subtest of the Gates-MacGinitie Reading Tests (GMRT; MacGinitie, MacGinitie, Maria, Dreyer, & Hughes, 2006). One school district was located in a large, urban metropolitan area; one district was located in a mid-size city; and two districts were located in rural areas. Female students made up 49.8% of the sample. With regard to ethnicity, 40.3% of the students were identified as Hispanic. The racial composition of the sample was 43.4% African American, 33.9% Caucasian, 21.3% American Indian, 2.7% Asian, and .5% Pacific Islander. The vast majority (91.7%) of students in the sample were considered as low income, 13.5% were English learners, and 18.3% were identified as having a disability. Specific information on the type of disability was available only from three districts and indicated that of those eligible students, 53.3% had a specific learning disability, 36.7% had a speech and/or language impairment, and 10% had an intellectual disability. There were no significant differences between study conditions for gender ( $\chi^2$  [221] = .81,  $p$  = .67), ethnicity ( $\chi^2$  [221] = 1.62,  $p$  = .45), race ( $\chi^2$  [218]  $p$  = .59), socioeconomic status ( $\chi^2$  [181] = 2.00,  $p$  = .16), English learner status ( $\chi^2$  [207] = .30,  $p$  = .59), or special education eligibility ( $\chi^2$  [180] = .74,  $p$  = .39). Sample demographics are provided in Table 1.

Table 1. Participant demographics.

	Treatment group ( <i>n</i> = 111)		Comparison group ( <i>n</i> = 110)		<i>P</i>
	<i>n</i>	%	<i>N</i>	%	
Gender					.668
Male	57	51.4	51	46.4	
Female	53	47.7	57	51.8	
Missing	1	.9	2	1.8	
Ethnicity					.445
Hispanic/Latino	49	44.1	40	36.4	
Non-Hispanic/Latino	61	55.0	68	61.8	
Missing	1	.9	2	1.8	
Race					.689
African American	43	38.7	47	42.7	
Caucasian	35	31.5	35	31.8	
American Indian	27	24.3	19	17.3	
Asian	1	.9	3	2.7	
Pacific Islander	1	.9	0	0	
Multiracial	3	2.7	4	3.6	
Missing	1	.9	2	1.8	
Free or reduced-price lunch					.157
Yes	87	78.4	79	71.8	
No	5	4.5	10	9.1	
Missing	19	17.1	21	19.1	
English learner					.586
Yes	13	11.7	15	13.6	
No	93	83.8	86	78.2	
Missing	5	4.5	9	8.2	
Identified disability					.389
Yes	15	13.5	18	16.4	
No	79	71.2	68	61.8	
Missing	17	15.3	24	21.8	

A total of 20 students (9% of total sample) withdrew from their respective schools during the school year. Attrition was 9.9% ( $n = 11$ ) in the treatment group and 8.2% ( $n = 9$ ) in the comparison group. These rates represent a low level of overall, and differential, attrition (What Works Clearinghouse, 2014). Multiple  $t$  tests revealed higher scores on word attack and oral reading fluency for students who remained in the school, but after accounting for multiple comparisons, there were no significant differences in pretest performance on any of the reading variables for students who withdrew compared to those students who remained in their school for the entire year.

## **Procedures**

### **Screening and Assignment**

All consented fourth-grade students at the 10 schools were screened; one class of students attending a self-contained classroom for students with emotional and behavior disorders at one of the schools was not included by request of the school administration. Students were administered the reading comprehension subtest of the GMRT during the fourth or fifth week of school. All students scoring at or below the 30th percentile on this measure were identified for the study, rank-ordered on the screening measure within school, and then randomly assigned within school to treatment ( $n = 111$ ) or comparison ( $n = 110$ ) using this stratification on the screening measure.

Students assigned to the treatment group were subsequently assigned within school to small groups of four to seven students (a total of 20 groups across schools). Each group received the Passport to Literacy intervention daily for 30 min for 24 weeks. Students assigned to the comparison group received the typical services provided by the school.

### **Data Collection**

Following screening, pretest measures were administered at the end of September and beginning of October. Posttest assessments were administered in early May, within two weeks of the intervention completion. Assessments were counterbalanced by measure and were administered by trained research assistants (RAs) who were blind to condition. Assessment staff were required to demonstrate 100% accuracy in administration and scoring before test administration in the field. This process was completed prior to pretesting and again prior to posttesting. Following administration of assessments at pretest and posttest, all measures were double-scored by a second RA.

To document the type and quality of core reading instruction (Tier 1) received by all students in the study, general education reading classes were observed and coded in the fall and spring by trained RAs using the Instructional Content Emphasis Instrument-Revised (ICE-R; Edmonds & Briggs, 2003). The ICE-R was used to document the content and grouping of instruction. As per the ICE-R guidelines, specific instructional activities were coded if they lasted for at least 1 min. Content categories included phonemic awareness (PA), phonics/word recognition, fluency, vocabulary/oral language development, comprehension, spelling, text reading, and nonliteracy activities (e.g., other academic instruction, noninstructional time). Instructional groupings were coded as whole-class, small-group, pairs, independent activity/assignment, or individualized instruction. Observers also coded student engagement during each instructional activity using a 3-point rubric (3 = high engagement, 1 = low engagement). Finally, a global quality of instruction rating was assigned on a 4-point

Likert scale ranging from weak (rating of 1) to excellent (rating of 4). This global instructional quality variable takes into account the teacher's use of direct and explicit language, modeling, opportunities for practice, specific feedback, monitoring and encouragement of engagement, scaffolding of tasks, and pacing.

A multiple-step training process was utilized to establish inter-rater reliability for the ICE-R (Edmonds & Briggs, 2003). First, each observer was instructed on the meaning of each code/indicator and provided with specific examples. Second, the coding process was modeled by the principal investigator of the project using a short video segment of reading instruction from another project. Third, each observer practiced coding using several novel video segments that were subsequently discussed with the principal investigator. Finally, each observer established 90% or higher coding accuracy with the principal investigator (i.e., gold standard approach) on a separate video segment of reading instruction. An agreement between the coder and the gold standard occurred for each minute of instruction and had to be an exact match (e.g., 2:01 p.m. = spelling instruction). Inter-rater reliability was calculated dividing the number of minutes of agreement divided by the number of minutes of agreement plus disagreements. Observers reestablished reliability prior to spring observations with new video segments. Reliability across coders was 96.4% at both fall and spring time points.

In order to identify supplemental reading instruction/intervention for students in the comparison group, classroom teachers first completed a brief interview with research staff regarding additional reading support received by each student in addition to their core reading instruction (Tier 1). The session time, frequency, grouping, implementer, and implementer's credentials were provided by the teachers each semester. To compare the reading instruction implemented for students receiving the Passport to Literacy intervention and those students in the comparison condition receiving a school-provided reading intervention, audio recordings of instructional sessions in both conditions occurred at three time points during the school year (fall, winter, and spring); recordings of instruction were coded using the ICE-R measure.

In addition, fidelity of implementation of the actual implementation of the Passport to Literacy intervention was monitored monthly via direct observations of lessons. Ratings were collected on implementation, student academic engagement, and quality of instruction for each lesson component. The scale for implementation ranged from 0 (*teacher did not complete elements of component*) to 3 (*all or nearly all required elements completed*), while engagement and instructional quality of each lesson component were also rated from 1 (*weak engagement or quality*) to 3 (*excellent engagement or quality*). Instructional quality indicators included ongoing monitoring, redirection of off-task behavior, positive and corrective feedback, organization of materials, and appropriate selection of additional items for practice when needed. Each observer obtained a minimum reliability of 90% in comparison to a gold standard rating by the project coordinator prior to formal data collection; across three observers, reliability was 93.2%.

## **Description of Instruction**

### **Tier 1**

With the exception of one school, all participating schools utilized Journeys Common Core (Templeton et al., 2014) as their core reading program in fourth grade. The other school



implemented Reading Street Common Core (Afflerbach et al., 2013) for Tier 1 instruction. Data from observations of core reading instruction indicated that the length of reading classes was, on average, 73.62 min ( $SD = 28.03$ ). Within this instruction, activities devoted to reading comprehension and vocabulary development were most prevalent, accounting for nearly 40 min of total instructional time. Instruction devoted to word analysis/decoding was minimal ( $<1$  min), while time spent in reading of connected text and/or oral reading fluency practice was approximately 5 min daily. Of note, approximately 15 min was spent in differentiated instructional activities where students in the class were engaged in different activities simultaneously. The additional 14 min was spent in other types of activities (e.g., transitions). With regard to grouping practices during the instruction, whole-class instruction was predominate (approximately 41 min on average). Just less than 10 min of instructional time consisted of students working independently on the same activity, while approximately 8 min was spent in either small-group or paired instructional activities. Generally, the global ratings of instruction for Tier 1 were suggestive of high average instructional quality ( $M = 3.26$ ,  $SD = .64$ ). Similarly, academic engagement by students during core reading instruction was rated as high ( $M = 2.81$ ,  $SD = .45$ ).

### ***Passport to Literacy Intervention***

Students in the treatment condition received the standard implementation of the Passport to Literacy intervention program at the fourth-grade level. The Passport to Literacy intervention has been developed for use as a supplemental reading intervention in daily, 30-min sessions provided in small groups of four to six students for one school year (120 lessons). Daily intervention sessions were scheduled jointly with the school/teachers and project staff. In many cases, intervention groups were scheduled during the school's designated intervention/enrichment time for all students. For 10 students, however, the only time allowed for the intervention was during the 30 min of Tier 1 instruction devoted to reading centers in their respective classrooms.

Passport to Literacy lessons are organized into 12 10-day adventures addressing phonics and word recognition, fluency, vocabulary, and comprehension in each lesson. Day 1 of each adventure began with an *Adventure Starter* activity (approximately 3–5 min) for building background knowledge and an essential, probing question that linked the lessons/reading in the adventure. Each lesson consisted of two main components: Word Works and Read to Understand. In general, Word Works focused on advanced word study, including working with affixes and roots as well as strategies for reading unknown multisyllabic words. During the initial six weeks, instruction in basic word reading skills was provided, including aspects of letter/sound identification, decoding, sight word reading, word families, and spelling instruction. Instruction for the Word Works component was designed to take 20 min during these initial adventures and then 5 min after the sixth week of the program; in addition, after the sixth week, each lesson contained a brief 2-min Warm-Up where students received additional word study practice through review and application of previously learned letter combinations, sight words, spelling rules, and word endings.

Words introduced in Word Works were typically practiced in context in the Read to Understand component of each lesson. The Read to Understand component was organized into before, during, and after reading comprehension skills and strategies. Students were introduced to new vocabulary daily using definitions, context, relationships to other words, and immediate student practice. A variety of comprehension tools were explicitly taught,



including 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. The texts presented in the Passport to Literacy program at this level included both literary and informational passages. The Read to Understand component was implemented for 10 min during the initial three adventures (approximately six weeks) and then comprised 25 min of the lesson in subsequent adventures for the duration of the intervention. Each lesson also included a short focus on fluency during the text reading.

In addition to the built-in review of each lesson, on Day 5 of each adventure students were administered a quick assessment of skills taught on Days 1 through 4 to provide teachers with information about individual student progress and mastery of the new knowledge. Additionally, this lesson provided time for reteaching specific skills to students who failed to demonstrate mastery. On the final lesson of each adventure (Day 10), a cumulative assessment of specific skills taught was administered. Additionally, the program included global, biweekly oral reading fluency measures built into every 10th lesson in order to monitor progress and inform instruction.

***Intervention Teachers and Training.*** There were nine teachers, hired by the research team, who were responsible for teaching the Passport to Literacy lessons. All of these individuals had a bachelor's degree and three (33.3%) had obtained a master's degree in education. Six of the interventionists were certified teachers; of the other three, one was a certified speech-language therapy assistant and the other two had degrees in noneducation areas. All intervention teachers were female. One teacher identified herself as Hispanic ethnicity. In terms of race, six (66.7%) teachers were Caucasian and three teachers (33.3%) were African American.

Prior to the initiation of the treatment, intervention teachers participated in approximately eight hours of training over the course of two days. Training provided by the project coordinators at each site allowed interventionists to become oriented to the project, familiarize themselves with the Passport to Literacy intervention program and instructional routine, practice implementation of lessons, and discuss positive behavior supports. Once intervention sessions with students were initiated, twice monthly coaching visits were conducted by the project coordinators. These visits allowed teachers to receive feedback on implementation as well as discuss any questions or concerns. Finally, monthly meetings with all intervention teachers were held at each site to provide continued support and ensure fidelity of implementation.

***Intervention Fidelity and Instruction.*** The total number of Passport to Literacy lessons covered for each of the 20 intervention groups ranged from 95 to 106 sessions. For those individual students who remained in the school for the duration of the intervention, the number of lessons attended ranged from a low of 70 sessions to a high of 105 sessions ( $M = 90.45$ ,  $SD = 7.01$ ); the median number of sessions attended was 91. Only five (5%) students attended fewer than 80 intervention lessons.

In terms of direct fidelity of implementation to the Passport to Literacy lessons, mean implementation ratings for each tutor implementation ranged from 2.81 to 3.00 across the lesson components. Similarly, mean ratings of student academic engagement (2.85–3.00)

and quality of lesson implementation (2.88–3.00) for each component were high. Table 2 provides the mean ratings for each component of the intervention.

As noted, each intervention teacher also recorded three intervention lessons during the year and these recordings were coded for reading instructional content and quality using the ICE-R. On average, the treatment session instruction was 26.70 min ( $SD = 4.02$ ) in length. Instruction focused on developing students' reading comprehension ( $M = 10.61$ ,  $SD = 5.51$ ) and vocabulary/oral language ability ( $M = 5.09$ ,  $SD = 4.23$ ). During treatment lessons, students engaged in text reading for 4.48 min ( $SD = 2.89$ ), decoding and word reading activities for 3.91 min ( $SD = 2.56$ ), and spelling practice for just under two minutes ( $M = 1.91$ ,  $SD = 2.76$ ). Explicit instruction in oral reading fluency was observed for .26 min ( $SD = .92$ ), on average. During treatment lessons, less than one minute of time was considered either non-instructional in nature ( $M = .30$ ,  $SD = 1.02$ ) or focused on instruction in another academic area such as writing ( $M = .13$ ,  $SD = .63$ ). Ratings of instructional quality indicated high-average quality ( $M = 3.39$ ,  $SD = .66$ ) and on average, intervention students were engaged during instruction ( $M = 2.87$ ,  $SD = .46$ ).

### Comparison Condition

Students in the comparison condition received the typical supplemental intervention provided by their respective schools. No schools had purchased or implemented the Passport to Literacy intervention. There was no evidence of Passport to Literacy lesson implementation in the comparison instruction. Thirty-five students (32% of comparison condition) received direct, supplemental reading instruction/intervention from a teacher during the school day. There were 75 students in the comparison group whom the school did not provide with supplemental intervention. The group of students chosen by the school for intervention had significantly lower scores on the word-level measures than the students not provided with intervention ( $ps = .001$ – $.009$ ). However, at posttest, those who received intervention were significantly lower than students who did not receive intervention on only word attack ( $p = .003$ ).

Teacher reports indicate that this supplemental reading intervention was most often delivered by classroom teachers (89% of students) or other certified teachers (9% of students), with instruction for one student (3%) implemented by a paraprofessional. All of the school intervention teachers were certified to teach elementary and/or special education; nearly three quarters (71.4%) held a reading endorsement. In terms of group size, 83% of the students received intervention in groups of eight or more students, 11% were in groups of four to five students, and 6% included three or fewer students. The supplemental intervention was most often delivered daily (91% of students), with the other 9% of students receiving intervention three to four times per week. Sessions were generally between 21–40 min

**Table 2.** Fidelity of Passport to Literacy implementation.

Lesson component	Number of observations	Implementation <i>M (SD)</i>	Academic engagement <i>M (SD)</i>	Instructional quality <i>M (SD)</i>
Adventure Starter	13	3.00 (0)	3.00 (0)	3.00 (0)
Warm-up	35	2.89 (.53)	2.97 (.17)	2.97 (.17)
Word Works	48	3.00 (0)	2.85 (.36)	2.88 (.33)
Before reading	49	3.00 (0)	2.94 (.24)	2.92 (.28)
During reading	49	2.98 (.14)	2.92 (.28)	2.94 (.24)
After reading	48	2.81 (.64)	2.91 (.29)	2.93 (.25)

(89% of students) with 6% of students receiving intervention sessions of 50 min or more per session and another 6% receiving sessions of 10 to 20 min. Eight students received two supplemental interventions during the school day.

Based on recordings of this instruction, intervention sessions for the comparison group of students averaged 25.15 min ( $SD = 11.13$ ). Similar to the treatment, the most frequent instructional activities involved those related to comprehension of text ( $M = 9.14$ ,  $SD = 3.48$ ) and vocabulary and oral language development ( $M = 5.90$  min,  $SD = 7.16$ ). Text reading occurred for approximately four and a half minutes ( $M = 4.46$ ,  $SD = 3.14$ ), while on average, students received phonics/decoding instruction for just over 1 min ( $M = 1.37$ ,  $SD = 4.94$ ) and oral reading fluency practice for just under 1 min ( $M = .97$ ,  $SD = 2.91$ ). Minimal instruction was focused on spelling ( $M = .22$ ,  $SD = 1.28$ ) and phonemic awareness ( $M = .08$ ,  $SD = .46$ ). During the additional reading intervention, 3.5 min were spent in other academic instruction and/or noninstruction ( $M = 2.95$ ,  $SD = 3.88$  for other academic instruction;  $M = .50$ ,  $SD = 1.19$  for noninstruction). The mean rating of instructional quality during this supplemental reading instruction was 3.23 ( $SD = .34$ ) and student engagement was also high ( $M = 2.91$ ,  $SD = .22$ ).

### ***Dependent Measures***

Data on students' word reading, decoding, reading fluency, and reading comprehension were collected both prior to, and at the completion of, the intervention.

#### ***Woodcock-Johnson III Tests of Achievement (WJIII; Woodcock, McGrew, & Mather, 2001)***

To specifically assess students' basic reading abilities the word attack and letter-word identification subtests were used. The word attack subtest is a pseudoword test that measures students' decoding skills. Letter-word identification requires students to name individual letters, as well as read real words presented. To assess students' abilities to read and understand connected text, the passage comprehension subtest was also administered. This subtest utilizes a cloze procedure wherein students are presented with several sentences with a missing word(s), and students are asked to supply the missing word. Test-retest reliabilities for these three subtests range from .81–.86 for fourth grade. 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.

#### ***Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1999)***

The TOWRE is a standardized, individually administered timed test of single-word reading fluency wherein students are given 45 seconds to read a list of words. The number of words read correctly within the time is recorded. Two subtests, Sight Word Efficiency (SWE) and Phonemic Decoding Efficiency (PDE) were administered. SWE assesses real-word reading while PDE measures decoding of nonsense words. Test-retest reliabilities range from .83–.96 on these subtests. For fourth graders, the concurrent validity for SWE and the Word Identification subset of Woodcock Reading Mastery Tests-Revised (WRMT-R) is .89. For PDE and the word attack subtest of WRMT-R, concurrent validity is estimated at .86.

### ***Dynamic Indicators of Basic Early Literacy Skills-6th Edition (DIBELS; Good & Kaminski, 2002)***

In order to measure students' abilities to read connected text with speed and accuracy, the Oral Reading Fluency (ORF) subtest from DIBELS was administered. The ORF measure requires students to read three separate passages aloud for one minute. The total number of correct words read per minute is recorded for each passage, and the median score of the three passages is used. Test-retest reliabilities for ORF with elementary-age students range from .92 to .97; alternate-form reliability across passages from the same level is reported as .89 to .94.

### ***GMRT (MacGinitie et al., 2006)***

The GMRT is a group-administered, norm-referenced test. The reading comprehension subtest was administered. Students are presented with multiple paragraph-length reading passages and related multiple-choice questions. Passages include both narrative and expository text. Scores from the GMRT were utilized to screen students in the fall of fourth grade and as an outcome measure in the spring. Test-retest reliabilities are above .85; alternate-form reliability is .86 for the fourth-grade level.

### ***Analytic Approach***

Primary impact analyses were initially evaluated with a set of mixed models to estimate the extent to which the treatment resulted in significant effects across the selected measures. Because the design of the study was a partially nested, randomized controlled trial (PN-RCT; Baldwin, Bauer, Stice, & Rohde, 2011; Lohr, Schochet, & Sanders, 2014), it was necessary that the analytic model appropriately fit the design. Although historical approaches to analyzing data from PN-RCT have ignored the partial nesting component (Baldwin et al., 2011), recent work has provided more robust guidelines for modeling such data (Baldwin et al., 2011; Lohr et al., 2014; Sterba et al., 2014). Subsequently, our model-building process began first with testing an unconditional model to evaluate the extent to which variance in each of the reading posttest scores was attributed to student differences, small-group differences for the intervention group, and school differences. Despite the relatively small sample size of schools ( $n = 10$ ), it was valuable to test whether differences in the posttest scores could be due to school-level nesting. Following the unconditional model, covariates were included to test the impact of the intervention on the reading outcomes controlling for pretest scores. Tests of the intervention effects had a linear step-up correction applied to the result in order to guard against a false-discovery rate (Benjamini & Hochberg, 1995).

In addition to the primary impact analyses, two sets of exploratory analyses were conducted for the measures demonstrating small effect sizes in order to examine the research aim regarding differential benefits of the intervention for students. We first tested the extent to which the relation between posttest scores and treatment effects was moderated by, or conditional upon, the pretest scores. Simple slopes analyses supplemented tests of moderation using methods outlined by Preacher, Curran, and Bauer (2006) to identify regions of significance for the interaction between pretest and the treatment dummy-code variable. As a complementary approach to testing for conditional relations between the posttest and treatment based on pretest scores, the second set of exploratory analyses examined whether conditional relations between the posttest and treatment existed conditional on posttest

scores. The mixed effects models used in the primary analysis are useful for empirically evaluating the average treatment effect, yet this approach based on averages may impose restrictions on interpretations. That is, most linear regression models are rooted in a conditional means approach which, by necessity, provides a conditional mean of  $y$  (e.g., posttest) given a value of  $x$  (e.g., treatment). Although the mean is a desired property for estimating coefficients in a regression analysis, it is possible that associations between variables may vary depending on different points of the distribution of  $y$ . Quantile regression (Koenker & Bassett, 1978; Petscher & Logan, 2014; Petscher, Logan, & Zhou, 2013) is a form of median regression that estimates the relations between  $y$  and  $x$  conditional on the distribution of  $y$ . While traditional linear regression is useful to answer the question, “What is the relation between treatment and average posttest scores?”, quantile regression is useful to answer the question, “Does the relation between treatment and posttest scores vary depending on the posttest score?” Quantile regression has been applied under circumstances where a continuous outcome has been regressed on a dichotomous predictor (Petscher & Logan, 2014), thus a natural extension of that model is to regress continuous posttest scores on a dummy-code variable of intervention effects.

The primary impact questions were analyzed using the mixed package in SAS following guidelines offered by Baldwin et al. (2011) and Lohr et al. (2014) for the partially nested models. Quantile regressions were estimated using the quantreg package in SAS. Hedges’s  $g$  was used as the effect size for the primary impact analysis. In the context of the quantile regression, effect size computation has received little attention. A coefficient of determination has been used in some reports of quantile regression via  $R^1$  (Soyiri & Reidpath, 2013) or  $R^2$  (Petscher et al., 2013), but due to the lack of randomized controlled trials using quantile regression, less research exists in this area. One mechanism for producing a standardized treatment effect indicator is to use Hedges’s  $g$  from an ANCOVA  $F$  test:

$$g = \sqrt{\frac{F(n_1 + n_2)(1 - r^2)}{n_1 n_2}}$$

where  $r$  is the correlation between the pretest and posttest,  $n_1$  is the sample size for the intervention group, and  $n_2$  is the sample size for the control group. At each specified quantile of interest, standard output for quantile regression includes a  $t$  test of coefficients. As such, it is plausible to use  $t^2$  in the Hedges’s  $g$  equation; moreover, the correlation between pretest and posttest at each quantile can be estimated. Thus, the necessary pieces for estimating Hedges’s  $g$  exist in a quantile framework.

## Results

### *Descriptive Statistics and Correlations*

Table 3 presents student performance results on the individual measures of decoding, word reading, fluency, and reading comprehension in the fall and spring for the full sample, as well as by treatment condition. Students had higher average scores in the spring compared to fall, with similar baseline scores between the treatment and comparison groups at the fall assessment period. The sample means in the fall indicate average decoding and word reading accuracy, but deficits of more than one standard deviation are noted in decoding efficiency,

**Table 3.** Student descriptives.

Measure	Full sample			Treatment			Control		
	<i>n</i>	Mean	<i>SD</i>	<i>n</i>	Mean	<i>SD</i>	<i>n</i>	Mean	<i>SD</i>
WJIII WA Fall <sup>a</sup>	199	491.26	17.82	100	488.86	18.94	99	493.68	16.34
WJIII LWID Fall <sup>a</sup>	199	485.70	20.82	100	483.83	21.86	99	487.60	19.63
TOWRE PDE Fall <sup>b</sup>	201	84.35	13.85	100	82.24	13.59	101	86.45	13.86
TOWRE SWE Fall <sup>b</sup>	201	87.15	12.79	100	85.56	12.78	101	88.73	12.67
DIBELS ORF Fall <sup>c</sup>	201	81.23	27.08	100	76.94	26.31	101	85.49	27.28
WJIII PC Fall <sup>a</sup>	199	482.05	12.82	100	480.57	11.83	99	483.54	13.64
GMRT RC Fall <sup>d</sup>	201	439.38	19.46	100	437.72	20.64	101	441.02	18.16
WJIII WA Spring <sup>a</sup>	188	495.36	13.72	92	494.43	13.86	96	496.25	13.59
WJIII LWID Spring <sup>a</sup>	195	492.93	18.66	97	491.72	19.11	98	494.13	18.22
TOWRE PDE Spring <sup>b</sup>	196	86.49	14.66	97	84.82	14.24	99	88.12	14.95
TOWRE SWE Spring <sup>b</sup>	196	90.12	13.34	97	88.79	13.58	99	91.41	13.03
DIBELS ORF Spring <sup>c</sup>	196	96.79	31.01	97	92.72	28.69	99	100.78	32.78
WJIII PC Spring <sup>a</sup>	191	488.01	8.98	95	487.95	9.28	96	488.06	8.72
GMRT RC Spring <sup>d</sup>	191	454.57	20.18	96	456.75	21.24	95	452.36	18.91

Notes. WJIII = Woodcock Johnson III Tests of Achievement. WA = word attack. LWID = letter–word identification. TOWRE = Test of Word Reading Efficiency. PDE = phonemic decoding efficiency. SWE = sight word efficiency. DIBELS = Dynamic Indicators of Basic Early Literacy Skills. ORF = oral reading fluency. PC = passage comprehension. GMRT = Gates-MacGinitie Reading Tests. RC = reading comprehension.

<sup>a</sup>W-score; <sup>b</sup>standard score; <sup>c</sup>raw score; <sup>d</sup>scaled score.

word reading efficiency, and reading comprehension. Fall oral reading fluency scores also averaged below the DIBELS ORF expected benchmark of 93 words correct per minute for fall of fourth grade. Correlations among the measures (Table 4) in the fall ranged from  $r = .30$  between GMRT reading comprehension and TOWRE PDE to  $r = .86$  between DIBELS ORF and TOWRE SWE. Spring associations ranged from  $r = .29$  between WJIII word attack and GMRT reading comprehension to  $r = .87$  between DIBELS ORF and TOWRE SWE. Stability coefficients from fall to spring ranged from  $r = .38$  for GMRT reading comprehension to  $r = .91$  for DIBELS ORF, suggesting moderate to high stability in relative rank orders of individuals over time.

**Table 4.** Measure correlations.

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. WJIII WA Fall	1.00													
2. WJIII LWID Fall	.80	1.00												
3. TOWRE PDE Fall	.78	.75	1.00											
4. TOWRE SWE Fall	.63	.70	.79	1.00										
5. DIBELS ORF Fall	.65	.74	.77	.86	1.00									
6. WJIII PC Fall	.59	.63	.51	.49	.54	1.00								
7. GRMT RC Fall	.35	.39	.30	.32	.37	.38	1.00							
8. WJIII WA Spring	.80	.77	.76	.60	.60	.50	.27	1.00						
9. WJIII LWID Spring	.77	.87	.74	.69	.73	.65	.40	.80	1.00					
10. TOWRE PDE Spring	.76	.75	.86	.76	.76	.44	.30	.77	.71	1.00				
11. TOWRE SWE Spring	.60	.70	.76	.85	.83	.48	.32	.59	.69	.79	1.00			
12. DIBELS ORF Spring	.64	.74	.75	.83	.91	.56	.38	.61	.72	.79	.87	1.00		
13. WJIII PC Spring	.49	.55	.46	.39	.45	.60	.37	.54	.63	.43	.42	.50	1.00	
14. GRMT RC Spring	.35	.38	.30	.38	.46	.44	.38	.29	.39	.31	.40	.49	.45	1.00

Notes. WJIII = Woodcock Johnson III Tests of Achievement. WA = word attack. LWID = letter–word identification. TOWRE = Test of Word Reading Efficiency. PDE = phonemic decoding efficiency. SWE = sight word efficiency. DIBELS = Dynamic Indicators of Basic Early Literacy Skills. ORF = oral reading fluency. PC = passage comprehension. GMRT = Gates MacGinitie Reading Tests. RC = reading comprehension.

**Table 5.** Pretest moderation for reading comprehension.

Outcome	Effect	Estimate	SE	df	t	P
GMRT RC	Intercept	451.90	1.78	95.2	253.41	<.001
	Pretest	0.40	0.07	191	5.81	<.001
	Passport	5.62	2.68	189	2.10	0.037
	Pretest*Passport	0.00	0.14	191	0.00	0.997
WJIII PC	Intercept	487.23	0.74	94	656.15	<.001
	Pretest	0.38	0.06	94	6.58	<.001
	Passport	1.27	1.05	63.2	1.21	0.230
	Pretest*Passport	0.17	0.09	184	1.98	0.049

Notes. GMRT = Gates MacGinitie Reading Tests. RC = reading comprehension. WJIII = Woodcock Johnson III Tests of Achievement.

### Primary Impact Analyses

The initial unconditional models estimated the extent to which variance in posttest scores was due to differences between the intervention clusters, differences between students in the intervention group, and differences between students in the control group. Variance components from the unconditional models suggested little variance due to clustering effects (i.e., <3%), though the ICCs (range .03–.13) suggested that the between-cluster variance, while relatively small, needed to be accounted for in subsequent modeling. The partially nested model was subsequently used for the conditional mixed models (Table 5).

Table 6 provides the fixed effects for each dependent variable. The first conditional model evaluated whether the treatment and comparison groups differed significantly at pretest. Results indicated that groups were statistically equivalent across all measures. Following this

**Table 6.** Fixed effects for dependent variables.

Outcome	Effect	Estimate	SE	df	t	p	G
WJIII WA	Intercept	494.93	0.90	94.80	549.33	<.001	0.07
	Pretest	0.64	0.03	175.00	18.50	<.001	
	Passport	1.02	1.31	50.30	0.78	.441	
WJIII LWID	Intercept	492.43	0.95	96.50	517.42	<.001	0.05
	Pretest	0.78	0.03	190.00	24.48	<.001	
	Passport	0.93	1.33	193.00	0.70	.485	
TOWRE PDE	Intercept	86.07	0.84	100.00	102.53	<.001	0.05
	Pretest	0.92	0.04	185.00	23.56	<.001	
	Passport	0.73	1.15	63.20	0.63	.528	
TOWRE SWE	Intercept	89.84	0.69	99.90	129.60	<.001	0.04
	Pretest	0.89	0.04	181.00	22.85	<.001	
	Passport	0.47	1.01	60.70	0.47	.644	
DIBELS ORF	Intercept	96.06	1.32	98.80	72.72	<.001	0.04
	Pretest	1.04	0.03	183.00	30.04	<.001	
	Passport	1.22	1.93	55.20	0.63	.528	
WJIII PC	Intercept	487.10	0.75	92.90	652.64	<.001	0.14
	Pretest	0.46	0.04	184.00	10.34	<.001	
	Passport	1.30	1.07	63.70	1.22	.227	
GMRT RC	Intercept	451.90	1.78	95.20	253.41	<.001	0.28
	Pretest	0.40	0.07	191.00	5.81	<.001	
	Passport	5.62	2.68	189.00	2.11	.037 <sup>a</sup>	

Notes. WJIII = Woodcock Johnson III Tests of Achievement. WA = word attack. LWID = letter–word identification. TOWRE = Test of Word Reading Efficiency. PDE = phonemic decoding efficiency. SWE = sight word efficiency. DIBELS = Dynamic Indicators of Basic Early Literacy Skills. ORF = oral reading fluency. PC = passage comprehension. GMRT = Gates MacGinitie Reading Tests. RC = reading comprehension.

<sup>a</sup>Not significant after applying Benjamini-Hochberg correction ( $\alpha^* = .05/7$ )



test, the mixed effects model tested for the impact of the intervention. No significant findings were observed for any of the word reading or fluency outcomes (Hedges's  $g$  range = 0.04–0.07). The individual mixed model for the GMRT reading comprehension outcome resulted in a significant, positive effect ( $g = 0.28$ ); however, the application of the Benjamini-Hochberg correction for the seven tests of treatment effects yielded a nonsignificant  $p$  value. No significant effect was observed for WJIII passage comprehension ( $g = 0.14$ ).

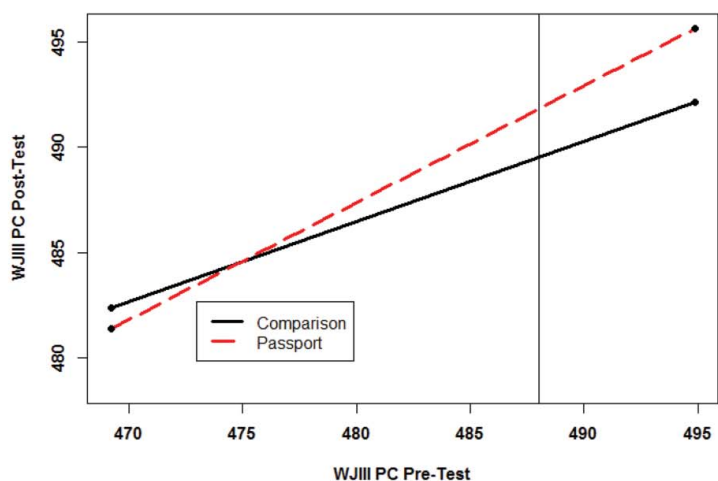
### Exploratory Analyses

The conditional effects of the Passport to Literacy intervention on the reading comprehension measures were explored by first testing the extent to which the relation between posttest, reading comprehension performance, and the intervention was moderated by students' baseline reading comprehension scores. Results suggested that a marginal effect for baseline moderation existed for the WJIII reading comprehension outcome but not the GMRT reading comprehension outcome (Table 5). Probing the interaction terms for the regions of significance on the WJIII reading comprehension measure was conducted using a simple slopes analysis (Preacher, Curran, & Bauer, 2006). The model coefficients for the impact and moderator analyses include centered fall pretest scores. The test revealed that moderation existed when centered pretest scores on the WJIII passage comprehension were greater than 5.70 (60th percentile). In the present sample, a mean WJIII passage comprehension pretest of 482 was observed. Thus, the implication of this moderation test is that the relation between treatment and posttest performance was positively moderated by pretest scores when the pretest was at least 488 (Figure 1a).

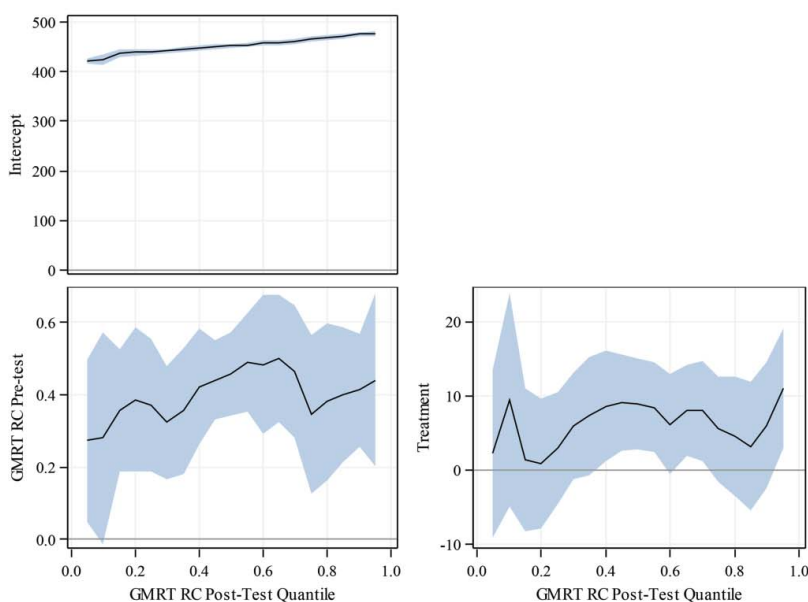
The second exploratory analysis tested for impacts of the Passport to Literacy treatment conditional on the posttest scores. Results from the GMRT reading comprehension quantile regression are displayed in Figure 1b; note that three panels are included, one for the intercept, one for the fall pretest scores, and one for the dichotomous variable representing treatment effects. For explication purposes, the focus will be on the third graph for treatment effects. The quantile regression highlights that the impact of Passport to Literacy ranges across levels of posttest performance on the GMRT reading comprehension. At the .40 quantile of posttest GMRT reading comprehension, the coefficient for the intervention was 8.59 ( $t(1) = 8.59, p < .05$ ) indicating that at approximately the 40th percentile of GMRT reading comprehension the gap in performance between students in the treatment and comparison conditions was approximately nine points in favor of students in the treatment. This result corresponds to a Hedges's  $g$  of 0.28. The panel in Figure 1b illustrates that significant effects for the Passport to Literacy intervention were observed from approximately the .40 quantile to the .70 quantile. Effect sizes within this part of the distribution of the GMRT reading comprehension ranged from  $g = 0.23$  at the .60 quantile up to  $g = 0.38$  at the .50 quantile. When averaged across the .40 to .60 quantile range, the mean effect size was  $g = 0.32$  ( $SD = 0.05$ ).

### Discussion

The primary purpose of this initial investigation, within a multiyear project, was to examine the efficacy of Passport to Literacy as a standardized protocol Tier 2-type intervention for fourth-grade students with reading comprehension difficulties. We sought to explore whether Passport to Literacy's relative emphasis in terms of time on comprehension within



(a)



(b)

**Figure 1.** Exploratory analysis graphs for (a) pretest moderation of WJIII PC with vertical reference line for significant simple slopes and (b) quantile regression of impacts of Passport conditional on GMRT RC post-test scores. The x-axis represents the quantile of GMRT RC posttest scores and the y-axis is the range of values for the labeled effect. The black line represents the parameter coefficient at each estimated quantile. The shaded area represents the confidence interval for the coefficient. WJIII = Woodcock Johnson III Tests of Achievement. PC = passage comprehension. GMRT = Gates MacGinitie Reading Tests. RC = reading comprehension.

a multicomponent intervention would translate into meaningful gains in reading outcomes relative to a business-as-usual comparison group. Specifically, this was the first study to use a randomized control trial to examine whether students with reading difficulties receiving

this widely used Tier 2 intervention outperformed students receiving typical school services on standardized measures of decoding, word recognition, fluency, or comprehension.

We found a positive, nonsignificant overall effect for Passport to Literacy for the norm-referenced GMRT reading comprehension outcome ( $ES = 0.28$ ), which exceeds the effect size criteria of 0.25 for “substantively important” from the What Works Clearinghouse (2014). No significant differences could be detected on the WJIII passage comprehension measure, though a small effect size ( $ES = 0.14$ ) was noted. The analyses revealed no significant differences or practical effects in outcomes across the two conditions for decoding, word recognition, decoding fluency, word recognition fluency, or oral reading fluency. This may not be surprising when considering that on average, Passport to Literacy lessons included less than a minute of fluency instruction and less than 4 min of decoding or word reading instruction. On average, relatively more time was devoted to vocabulary (5 min) and comprehension (11 min) across the school year. Another possible explanation for the lack of significant differences in decoding/word reading and fluency could be due to the relative effectiveness of the Tier 1 intervention; on average, teachers’ instruction was rated highly, as was the degree of student engagement. Furthermore, nearly a third of the students in the comparison group received reading intervention provided by school personnel. Lemons, Fuchs, Gilbert and Fuchs (2014) pointed out that the nature of counterfactuals has changed, with increased emphasis on the need for evidence-based Tier 1 core reading programs and the inability to have true control groups in schools; effect sizes are clearly impacted by the counterfactual.

Differences in the magnitude of effect for these two different comprehension outcomes may be due to the measures capturing slightly different types of comprehension (Cutting & Scarborough, 2006; Keenan, Betjemann, & Olson, 2008). The GMRT reading comprehension measure is a timed test, and it requires students to read relatively long passages and answer questions regarding their reading. The students in the Passport to Literacy treatment may have developed stronger comprehension practices, including monitoring of comprehension in text, than the students in the comparison group, which allowed them to better access the passages in the GMRT measure. In contrast, the WJIII passage comprehension requires reading of shorter passages (largely 1–2 sentences), and students must only supply a missing word rather than answer questions about the passage. It may be that smaller effects were seen between groups on this measure because the treatment and comparison instruction both provided sufficient support for students to gain this type of comprehension ability. The stability ratio for the WJIII passage comprehension was higher than it was for the GMRT reading comprehension, demonstrating that individual students largely maintained their relative rank orders from pretest to posttest on the WJIII passage comprehension measure. In previous research, the WJIII passage comprehension has shown the lowest factor loading of comprehension tests, and variance in student scores on this measure has been predicted better by a student’s decoding ability rather than his or her listening comprehension (Keenan et al., 2008).

In placing these overall reading comprehension findings into the previous research, we note that our findings are similar to findings related to standardized comprehension measures in a recent synthesis of interventions for students in Grades 4–12, where a mean effect size of 0.19 was reported across studies (Scammacca, Roberts, Vaughn, & Stuebing, 2015). Additionally, the effect sizes seen in this study are relatively higher than a synthesis of

extensive reading interventions (defined as 100 or more sessions of intervention) for secondary students, where an effect size of .10 was noted for reading comprehension (Wanzek et al., 2013). In terms of multicomponent intervention implementations for upper elementary students, our effect sizes for standardized comprehension measures are also relatively larger than those reported by Ritchey et al. (2012), yet smaller than two previously reported multicomponent interventions: 0.50 reported by Vadasy and Sanders (2008) and 1.39–1.46 reported by O'Connor et al. (2002). In examining the types of interventions implemented across the multicomponent studies, a more significant fluency emphasis and instruction provided in smaller groups seem to differentiate the interventions in the O'Connor et al. and the Vadasy and Sanders studies from the current study. For example, the two interventions implemented in O'Connor et al.'s study included about 10 min of phonological awareness, word analysis, and spelling instruction, and about 20 min of reading connected text, fluency building, and comprehension. Further, O'Connor et al. noted that students with the lowest fluency showed the weakest response and that they were not able to benefit from instruction on grade-level texts.

To begin to understand for which students Passport to Literacy may be most effective, we further explored the reading comprehension outcomes for evidence of differential effects/benefits for students of varying levels of reading comprehension ability. First, we examined students' incoming comprehension levels as a moderator of their comprehension outcomes. There was no significant moderation found for the GMRT reading comprehension, but we did find significant moderation of pretest levels for the WJIII passage comprehension. Students significantly benefited from the Passport to Literacy intervention if their WJIII passage comprehension began at an above-average level (60th percentile or higher), despite performing below the 30th percentile on the GMRT reading comprehension measure. Thus, students with higher incoming levels on the short passage, cloze measure who still struggled in reading comprehension for lengthier passages and questions benefited most from the Passport to Literacy instruction. For students with lower initial passage comprehension scores, posttest scores on the WJIII passage comprehension were similar in both study conditions, suggesting Passport to Literacy did not have the same benefit for these lower level students. This inadequate response from students in our study with lower initial comprehension scores may be similar to O'Connor et al.'s (2002) students with low fluency scores who also showed inadequate response. Further research is needed to ascertain whether students with the weakest initial skills need a more intensive intervention that is provided in smaller groups or that is differentiated by students' individual needs.

Second, we explored the posttest quantiles of student reading comprehension for differential effects of the treatment. In this exploratory analysis, we found that significant treatment effects for the intervention were observed for students who completed the study between the .40 and .70 quantiles on the GMRT reading comprehension measure. If replicated, this finding—in combination with the moderation findings—provides a unique contribution to the literature on upper elementary interventions, empirically targeting those for whom this intervention may be most beneficial. Importantly, Passport to Literacy seems to be least effective for students with the lowest levels of comprehension ability.

The uniqueness of this type of analysis prevents comparison to previous upper elementary literature, but does suggest that simply reporting average findings of intervention effects may not be sufficient. The findings from this study indicate that one size (or a standardized intervention) does not fit all. An obvious implication is the need for more effective intensive

remedial interventions for students at the lowest achievement levels. It is clear that the standardized implementation of Passport to Literacy did not fully meet these students' needs. One possibility that might deviate from a standard protocol would be to individualize the amount of time in various components of the intervention to align with student needs. Even within a sample of students with reading difficulties (i.e., performing at or below the 30th percentile), some children who ended the study within the lower quantiles may have benefited from relatively more word study and fluency practice as well as more intense instruction in terms of additional time or group size. Further research is needed to determine the active ingredients for these more intense interventions, but the current study provides evidence that many students in the upper elementary grades may not benefit sufficiently from a standardized, multicomponent Tier II-type intervention with an emphasis on comprehension instruction. As we saw across Tier I observations and typical school intervention services, instruction with an emphasis on comprehension is the current norm for all levels of students at the upper elementary level.

These exploratory findings also lead us to several areas of future research. The relatively small sample included in this study for these types of analyses requires that the findings be replicated. Replications of these findings over time could provide important practical implications regarding those for whom the Passport to Literacy intervention is intended. In addition, examining the student characteristics that predict student placement in the posttest quantiles for which the intervention best targets could not only have practical implications for early identification of students who may best benefit from the intervention, but could also inform future research on the use of time and individualization in upper elementary interventions related to students for whom Passport to Literacy may not be most beneficial.

### **Limitations**

As with any school-based research, there are several limitations to our study. First, findings are not directly generalizable to other multicomponent interventions or to other grade levels. Passport to Literacy is a standardized group-administered intervention and findings could differ for smaller group sizes, or for interventions that were more individualized. Second, our participants had relatively weak beginning and end of year fluency scores, but relatively accurate decoding and sight word accuracy standard scores. Thus findings may not generalize to different populations, but the importance of fluency as a requisite for comprehension is consistent with the verbal efficiency theory (Perfetti, 1985). Third, all but one school used the same core, Tier 1 reading program. On average, Tier 1 instruction was high-quality and student engagement was high. Thus, findings may not generalize to other core programs or to lower quality instruction. Fourth, although it is notable that we observed Tier 2 in both conditions, there was considerable variability in the amount and types of Tier 2 intervention provided within the comparison group. Namely, only 32% of the comparison group received supplemental intervention. These interventions were usually provided by their classroom teacher for a similar time as the Passport to Literacy lessons.

In summary, although the small effect sizes on reading comprehension favoring the Passport to Literacy intervention over the comparison condition was promising, findings across other measures were not significant. We did note significant effects for treatment for specific levels of students. The findings do suggest the need for further research to increase the intensity and robustness of interventions for students in the upper elementary grades

who enter with the lowest achievement levels to help educators better understand students' response to intervention.

## Funding

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R324A130262 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.

## ARTICLE HISTORY

Received 21 March 2015

Revised 31 August 2015

Accepted 25 October 2015

## EDITORS

This article was reviewed and accepted under the editorship of Carol McDonald Connor and Spyros Konstantopoulos.

## References

- Afflerbach, P., Blachowicz, C. L. Z., Boyd, C. D., Izquierdo, E., Juel, C., Kame'enui, E. J., ... Wixson, K. K. (2013). *Reading Street Common Core*. Glenview, IL: Pearson.
- 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
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, 57, 289–300.
- Compton, D. L., Fuchs, D., Fuchs, L. S., Elleman, A. M., & Gilbert, J. K. (2008). Tracking children who fly below the radar: Latent transition modeling of students with late-emerging reading disability. *Learning and Individual Differences*, 18, 329–337. doi:10.1016/j.lindif.2008.04.003
- Cutting, L. E., & Scarborough, H. S. (2006). Prediction of comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading*, 10, 277–299. doi:10.1207/s1532799xssr1003\_5
- Edmonds, M., & Briggs, K. (2003). The instructional content emphasis instrument: Observations of reading instruction. In S. Vaughn & K. L. Briggs (Eds.), *Reading in the classroom: Systems for the observation of teaching and learning* (pp. 31–52). Baltimore, MD: Brookes Publishing.
- Flavell, J. H. (1992). Cognitive development: Past, present, and future. *Developmental Psychology*, 28, 998–1005. doi:10.1037/0012-1649.28.6.998
- Florida Department of Education. (2010). *Just Read Reading Intervention Selections*. Tallahassee, FL: Author.
- Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W. D. (2008). *Assisting students struggling with reading: Response to intervention and multi-tier intervention for reading in the primary grades* (NCEE 2009-4045). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Good, R. H., & Kaminski, R. (2002). *Dynamic Indicators of Basic Early Literacy Skills–6th Edition (DIBELS)*. Eugene, OR: Institute for the Development of Educational Achievement.
- Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*, 12, 281–300. doi:10.1080/10888430802132279



- Koenker, R., & Bassett, G., Jr. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society*, 33–50. doi:10.2307/1913643
- Leach, J. M., Scarborough, H. S., & Rescorla, L. (2003). Late-emerging reading disabilities. *Journal of Educational Psychology*, 95, 211–224. doi:10.1037/0022-0663.95.2.211
- Lemons, C. J., Fuchs, D., Gilbert, J. K., & Fuchs, L. S. (2014). Evidence-based practices in a changing world: Reconsidering the counterfactual in educational research. *Educational Researcher*, 43, 242–252.
- Liberman, I. Y., Shankweiler, D., & Liberman, A. M. (1989). The alphabetic principle and learning to read. In D. Shankweiler (Ed.), *Phonology and reading disability: Solving the reading puzzle* (pp. 1–33). Ann Arbor: University of Michigan Press.
- 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). Retrieved from <http://ies.ed.gov/ncer/pubs/20142000/pdf/20142000.pdf>
- MacGinitie, W. H., MacGinitie, R. K., Maria, K., Dreyer, L. G., & Hughes, K. E. (2006). *Gates-MacGinitie Reading Tests* (4th ed.). Rolling Meadows, IL: Riverside Publishing.
- National Center for Education Statistics. (2013). *National assessment of educational progress: The nation's report card*. Washington, DC: U.S. Department of Education.
- 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. doi:10.1037/0022-0663.94.3.474
- Office of Special Education and Rehabilitative Services. (2013). *Thirty-fifth annual report to congress on the implementation of the Individuals with Disabilities Act*. Retrieved from <http://www2.ed.gov/about/reports/annual/osep/2013/parts-b-c/35th-idea-arc.pdf>
- Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1, 117–175. doi:10.1207/s1532690xci0102\_1
- Perfetti, C. A. (1985). *Reading ability*. New York, NY: Oxford University Press.
- Petscher, Y., & Logan, J. A. R. (2014). Quantile regression in the study of developmental sciences. *Child Development*, 85, 861–881. doi:10.1111/cdev.12190
- Petscher, Y., Logan, J. A. R., & Zhou, C. (2013). Extending conditional means modeling: An introduction to quantile regression. In Y. Petscher, C. Schatschneider, & D. L. Compton (Eds.), *Applied quantitative analysis in education and social sciences* (pp. 3–33). New York, NY: Routledge.
- Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interaction effects in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*, 31, 437–448. doi:10.3102/10769986031004437
- Ransby, M. J., & Swanson, H. L. (2003). Reading comprehension skills of young adults with childhood diagnoses of dyslexia. *Journal of Learning Disabilities*, 36, 538–555. doi:10.1177/00222194030360060501
- 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. doi:10.1177/001440291207800304
- Scammacca, N. K., Roberts, G., Vaughn, S., & Stuebing, K. K. (2013). 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
- Soyiri, I. N., & Reidpath, D. D. (2013). The use of a quantile regression to forecast higher than expected respiratory deaths in a daily time series: A study of New York City data 1987–2000. *PLOS ONE*, 8(10), e78215. doi:10.1371/journal.pone.0078215
- 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. doi:10.1080/00273171.2014.882253
- Templeton, S., Lipson, M., Valencia, S. W., Vogt, M., Pikulski, J. J., Chard, D. J., ... Valentino, C. (2014). *Journeys Common Core*. Boston, MA: Houghton Mifflin Harcourt.
- Therrien, W. J., 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
- Torgesen, J. K., Wagner, R., & Rashotte, C. (1999). *Test of word reading efficiency*. Austin, TX: Pro-Ed.



- Vadasy, P. F., & Sanders, E. A. (2008). Repeated reading intervention: Outcomes and interactions with readers' skills and classroom instruction. *Journal of Educational Psychology*, 100, 272–290. doi:10.1037/0022-0663.100.2.272
- 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. doi:10.1007/s11145-009-9183-9
- Wanzek, J., Vaughn, S., Scammacca, N., Gatlin, B., Walker, M. A., & Capin, P. (in press). Meta-analyses of the effects of Tier 2 type reading interventions in grades K-3. *Educational Psychology Review*.
- Wanzek, J., Vaughn, S., Scammacca, N., Metz, K., Murray, C., Roberts, G., & Danielson, L. (2013). Extensive reading intervention for older struggling readers: Implications from research. *Review of Educational Research*, 83, 163–195. doi:10.3102/0034654313477212
- 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(8), 889–912. doi:10.1007/s11145-009-9179-5
- What Works Clearinghouse. (2014). *Procedures and standards handbook* (Version 3.0). Retrieved from [http://ies.ed.gov/ncee/wwc/pdf/reference\\_resources/wwc\\_procedures\\_v3\\_0\\_standards\\_handbook.pdf](http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_standards_handbook.pdf)
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. Itasca, IL: Riverside.