Response to Intervention for Middle School Students With Reading Difficulties: Effects of a Primary and Secondary Intervention

Sharon Vaughn,
The University of Texas at Austin, The Meadows Center for Preventing Educational Risk

Paul T. Cirino,
University of Houston

Jeanne Wanzek,
Florida State University

Jade Wexler,
The University of Texas at Austin, The Meadows Center for Preventing Educational Risk

Jack M. Fletcher,
University of Houston

Carolyn D. Denton,
University of Texas Houston Medical Center

Amy Barth,
University of Houston

Melissa Romain, and
University of Houston

David J. Francis
University of Houston

Abstract

This study examined the effectiveness of a yearlong, researcher-provided, Tier 2 (secondary) intervention with a group of sixth-graders. The intervention emphasized word recognition, vocabulary, fluency, and comprehension, Participants scored below a proficiency level on their slate accountability test and were compared to a similar group of struggling readers receiving school-provided instruction. All students received the benefits of content area teachers who participated in researcher-provided professional development designed to integrate vocabulary and comprehension practices throughout the school day (Tier 1). Students who participated in the Tier 2 intervention showed gains on measures of decoding, fluency, and comprehension, but differences relative to students in the comparison group were small (median $d = +0.16$). Students who received the re searcher-provided intervention scored significantly higher than students who received comparison intervention on measures of word attack, spelling, the state accountability measure, passage comprehension, and phonemic decoding efficiency, although most often in particular subgroups.

Correspondence regarding this article should be addressed to Sharon Vaughn, The University of Texas at Austin, The Meadows Center for Preventing Educational Risk, 1 University Station, D 4900, SZB 228, Austin, TX 78712; srvaughnum@aol.com.

Recognizing the large numbers of students who need academic and behavioral intervention in our schools, educators, policy makers, and researchers have called for school-wide intervention frameworks in which students' response to quality intervention is monitored and used to inform decisions about future intervention and placement (see Fletcher, Lyon, Fuchs, & Barnes, 2007: VanDerHeyden, 2007; Jimerson, Bums, & VanDerHeyden, 2007. However, there is minimal re search-based guidance for effective implementation of tiered interventions for older students (e.g., Grades 4–8) and for effective reading interventions for older students (Kamil et al., 2008).

Edmonds and colleagues (2009) conducted a meta-analysis of 13 experimental and quasi-experimental studies that examined the effects of decoding, fluency, vocabulary, and comprehension interventions on the reading comprehension of students in Grades 6–12. The mean weighted average effect size of these studies on comprehension outcomes was 0.89, in favor of treatment students over comparison students, which suggested that older students with reading difficulties significantly benefited from interventions. Word-level interventions were associated with moderate effect size gains in reading comprehension ($d = 0.34$).

Scammacca and colleagues (2007) extended the Edmonds et al. (2009) meta-analysis to studies that examined reading outcomes in domains other than comprehension. The interventions were conducted with older students with reading difficulties and resulted in a mean effect size of $d = 0.95$ from 31 studies, Several of these studies measured outcomes using researcher-developed instruments; the average effect size was considerably lower when standardized, norm-referenced measures were analyzed ($d = 0.42$). Comprehension and vocabulary interventions were associated with the highest effect sizes, and word study interventions were associated with moderate effect sizes. Interventions implemented by researchers were associated with higher effect sizes than those implemented by teachers; and effects were higher for middle-grade students than for students in high school.

The findings from these two comprehensive syntheses on interventions with older students should be considered in light of several important issues that are not adequately reflected in aggregated effect sizes. First, the effect sizes favoring treatment students may have been inflated if die comparison students were not participating in any reading instruction. Unlike in elementary school where all students receive reading instruction, reading instruction at the middle school level may not be formal and may be represented as part of occasional vocabulary or comprehension activities in the content area. Second, most of the interventions represented in the syntheses were relatively short term (less than 2 months). Finally, insufficient data were available from the studies to determine whether the interventions improved student outcomes relative to grade-level expectations. A moderate or large effect size does not indicate whether the acceleration in reading performance contributed to a meaningful gain in terms of "closing the performance gap" relative to typically developing peers, which is particularly noteworthy with older students because these students are more likely to be multiple grade levels behind the normative sample.

Purpose of This Study

The purpose of this study was to implement and evaluate the outcomes of a comprehensive researcher-provided intervention with older students with reading difficulties. We designed the study to address the gaps in the current research on middle-grade students with reading difficulties. All students in both the treatment and comparison groups benefited from their teachers' participation in a professional development designed to enhance the quality of the core reading instruction (i.e., Tier 1). We also addressed a gap in the research literature by providing an extensive yearlong intervention (i.e., Tier 2) and by using highly reliable and valid measures to determine program efficacy. This study is part of a large-scale, multiyear

study designed to examine the efficacy of increasingly intensive interventions for middle school students with significant reading difficulties. The study reported here represents the first year of implementation in which the intervention and its implementation format were specifically designed to be feasible, given the realities of middle schools.

Our primary research question was as follows: What are the effects of a secondary intervention (Tier 2) provided in relatively large groups (10–15 students) on the reading-related outcomes of individuals with reading difficulties? Based on our previous review of secondary interventions with older students, we hypothesized that the Tier 2 researcher-provided intervention would result in improved outcomes for students relative to other students at risk for reading difficulties, and that Tier 2 students would close the gap with typical readers without reading difficulties over the course of the year.

Method

Participants

School sites—This study was conducted in two large urban cities in the southwestern United States, with approximately half the sample from each site. Sixth-graders from seven middle schools participated in the study, including three schools from a large urban district in one city and four schools from two medium districts in the smaller city. The rate of students qualifying for reduced-cost or free lunch ranged from 56% to 86% across the schools in the larger site and from 40% to 85% in the smaller site.

Criteria for participation—We selected all struggling readers in sixth grade as well as a random sample of typical readers, and used the Texas Assessment of Knowledge and Skills (TAKS: Texas Education Agency, 2004) to identify struggling readers. The participants either obtained a TAKS scaled score below the cutoff of 2,100, or had an obtained TAKS scaled score whose lower bound 95% confidence interval included a failing score. Thus, the sample included students with "bubble" scores (2,100–2,150) because they were potentially at risk of not passing the state achievement test simply because of the measurement error of the test. Students exempted from the TAKS because of special education status and very low reading achievement were also selected. Typical readers scored at least one standard error of measurement above the passing score (i.e., higher than 2150). Students were excluded from participation only if (a) they were enrolled in an alternative curriculum (i.e., life skills class): (b) their performance levels corresponded to a second-grade reading level or lower; or (c) they were identified as having a significant disability (e.g., blindness, deafness) or had individualized education plans that prevented them from participating in a reading intervention.

Student participants—The preliminary sample included 2.034 filth-grade students, who had useable and eligible state test scores in the spring of the 2005–2006 academic year and who were slated to attend one of the seven designated middle schools. These students were designated as either struggling (n = 759) or typical (n = 1.275) readers. The 759 struggling readers were randomly assigned within school in a 2:1 ratio to either researcher-provided Tier 2 intervention (referred to as Tier 2 or Tier 2 treatment; n = 506) or a comparison condition (n = 253). Of the 506 students who received Tier 2. 191 (38%) did not attend their scheduled middle school: of the 253 comparison. 101 (40%) did not attend their scheduled middle school. An additional 25 students assigned to Tier 2 (8%) and 7 students assigned to the comparison (5%) met one of the exclusion criteria outlined earlier (not known at time of randomization). Altogether, the proportion of these students not available and/or excluded did not differ between the two treatment groups (p > .05), and struggling students available to participate (M = 1,881: SD = 328) did not differ from those not available on TAKS scores

(M = 1,892: SD = 306), p <.05. An additional 49 Tier 2 (10%) and 30 comparison (12%) students were not present in their schools at the end of the school year, and these proportions were not significantly different, p > .05. These students did not differ from one another according to treatment group on any pretest measure (p > .05), and as a group they did not differ from those who remained for the duration of the study on any pretest measure (p > .05).

There were 241 Tier 2 students and 115 comparison students. However, 29 Tier 2 students did not receive this intervention, primarily because of an inability to schedule the intervention, a situation not encountered in the case of comparison students. These students did not differ from those who remained in the treatment group on any measure at pretest (all p > .05). Intent-to-treat analyses were performed on all available students, and these results did not differ from the results presented in this article.

Each school contributed between 15 and 97 students to this group of 327. Fifty-two percent of the sample was female, and 79% of the sample qualified for free or reduced-cost lunch (3% did not provide data on free or reduced-cost lunch). One-hundred fifty-two students (46%) were African American, 132 (40%) were Hispanic, 40 (12%) were Caucasian, and 3 (1%) were Asian. The proportions of students from the treatments did not differ in terms of site, sex, free or reduced-cost lunch status, age, or ethnicity (all p > .05).

The sample size for typical students was selected to represent approximately 60% of the original total struggling reader population, and so 468 typical readers were randomly selected from the pool of available typically developing readers. One hundred ninety of these students did not attend the middle school that their fifth-grade feeder pattern data suggested, and no further information was obtained for these students. This left 278 students in the typical group for pretest, and of these, 249 were available at post-test.

Measures

Decoding and spelling

We assessed word reading accuracy for real words and pseudowords with the Letter-Word Identification and Word Attack subtests of the Woodcock-Johnson III Tests of Achievement (WJ-III; Woodcock, McGrew, & Mather, 2001). The WJ-III Spelling subtest was also administered at post-test. Coefficient alphas in the entire sample of 327 struggling readers and 249 typical students who contributed data throughout the year for Letter-Word Identification and Word Attack ranged from .93 to .97: coefficient alpha for Spelling at posttest was .84.

Fluency

Because less is known about measuring fluency in middle school, we obtained multiple assessments of fluency for words and passages. The Sight Word Efficiency and Phonemic Decoding Efficiency subtests from the Test of Word Reading Efficiency (TOWRE: Torgesen, Wagner, & Rashotte, 1999) assessed word list fluency for real words and pseudowords. Alternate-forms reliability for this well-standardized test exceed .90 (Torgesen et at., 1999).

The AIMSweb Reading Maze (Shinn & Shinn, 2002), a 3-min, group-administered curriculum-based assessment, was administered at all five time points. The mean intercorrelation of performances across the five time points in the entire sample of 327 struggling readers and 249 typical readers was .79. Previous research found that the Maze data were sufficiently reliable for instructional decisions and resulted in valid decisions (Shinn & Shinn, 2002).

The Test of Sentence Reading Efficiency (TOSRE: Wagner et al., in press) is a 3-min, group-based assessment that was also used to assess reading fluency. Students are presented with a series of short sentences and are required to assess their veridicality. The mean intercorrelation of performances across the five time points in the entire sample of 327 struggling readers and 249 typical readers was .79 for standard scores and .80 for raw scores.

We designed assessments of Passage Fluency (PF) and Word List Fluency (WLF) specifically for this study (see www.texasld-center.org/outcomes). The PF consists of graded passages administered as 1-min probes to assess fluency of text reading. The passages averaged approximately 500 words each and ranged in difficulty from 350 to 1.400 Lexiles (Lennon & Burdick, 2004). Within each of 10 "Lexile bands," separated by 110 Lexile units, there were 10 passages, 5 of which were expository and 5 narrative. Students were administered the PF at five time points throughout the academic year, including pretest and post-test. At each time point, students read one story from each of the rive Lexile bands for 1 min each. The data that were used as the dependent measure were linearly equated-score averages of the five 1-min reads. The mean intercorrelation of the five stories read at pretest in the entire sample of 327 struggling readers and 249 typical readers was .87, and for the three stories read at posttest was .86.

For the WLF, students read as many words as possible from three word lists that varied in difficulty and the source of the words for I min each. WLF was assed at five time points throughout the academic year, including pretest and post-test. As with PF, the dependent measure utilized for WLF was a linearly equated-score average of the three 1-min word list reads. The mean intercorrelation of the three word lists read in the entire sample of 327 struggling readers and 249 typical readers was .92 at pretest and .89 at post-test.

Comprehension

The TAKS is a criterion-referenced reading comprehension test used for accountability testing in Texas. Students read passages (both expository and narrative) and answer corresponding questions. The internal consistency (coefficient alpha) of the Grade 7 test is .89 (Texas Education Agency, 2004).

We also administered the Passage Comprehension subtest of the Group Reading Assessment and Diagnostic Evaluation (GRADE; Williams, 2001) to assess comprehension. This subtest requires reading a passage and responding to multiple-choice questions. Coefficient alpha for the Passage Comprehension subtest in the entire sample of 327 struggling readers and 249 typicals who contributed data throughout the year was .82 at pretest.

The WJ-III Passage Comprehension subtest, a cloze-based assessment in which students read a passage and fill in missing words, was also used to assess comprehension. Coefficient alphas in the entire sample of 327 struggling readers and 249 typical students were .94 at pretest and .85 at post-test.

Kaufman Brief Intelligence Test—2

The Kaufman Brief Intelligence Test—2 (Kaufman & Kaufman, 2004) is an individually administered intellectual screening measure, used in this study for descriptive purposes. The Matrices subtest was administered at pretest and Verbal Knowledge subtest was administered at post-test. The Verbal Knowledge score was pro-rated for the verbal domain score.

Interventions

Tier 1

The research team provided professional development on evidence-based practices for teaching vocabulary and comprehension (see Denton, Bryan, Wexler, Reed, & Vaughn, 2007) to the content area teachers of all the sixth-grade students. Teachers attended a 6-hr professional development session at the beginning of the school year, and then met in study groups at their respective schools approximately once each month throughout the school year. Study groups consisted of interdisciplinary teams in six of the schools, but one school framed study groups by department area. In-classroom coaching was also provided on request.

The vocabulary component of the professional development was primarily adapted from Beck, McKeown, and Kucan (2002), where teachers learned to (a) select appropriate academic and content-specific vocabulary words to teach: (b) pronounce words part-by-part to assist students in decoding them; (c) provide brief, understandable definitions of the words: and (d) provide (or support students in generating) examples and nonexamples of the words. Teachers also learned to use graphic organizers to provide a framework for vocabulary instruction. The comprehension strategies presented in the professional development included (a) identifying and asking different types of questions, (b) a note-taking guide completed using main idea and summarizing strategies, and (c) identification of text structures and use of graphic organizers. During the monthly study group sessions, teachers worked with a facilitator to apply these strategies while planning lessons in their own content areas. Further information on the Tier 1 professional development can be found at the following Web site: www.meadowscenter.org/

Tier 2

Students were placed in homogeneous intervention groups to the extent that class schedules allowed and received a yearlong Tier 2 intervention. The researcher-provided intervention included three phases of instruction that varied in emphasis.

Phase 1 intervention consisted of approximately 25 lessons taught over 7–8 weeks and emphasized word study and fluency. *Fluency* was promoted by using oral reading fluency data and pairing higher and lower readers for partner reading. Students engaged in repeated reading daily with their partner with the goal of increased fluency. *Word Study* was promoted using the lessons in *REWARDS Intermediate* (Archer, Gleason, & Vachon, 2005a) to teach advanced strategies for decoding multisyllabic words. Progression through lessons was dependent on students' mastery of sounds and word reading. Students received daily instruction and practice with individual letter sounds, letter combinations, and affixes. In addition, students received instruction and practice in applying a strategy to decode and spell multisyllabic words by breaking them into known parts. *Vocabulary* was also addressed each day by teaching the meaning of the words through basic definitions and providing examples and nonexamples of how to use the words. New vocabulary words were reviewed daily, with students matching words to appropriate definitions or examples of word usage. *Comprehension* was addressed by asking students to answer relevant comprehension questions (literal and inferential). Teachers assisted students in locating information in text and rereading to identify answers.

In *Phase II* of the intervention, instruction emphasized vocabulary and comprehension, with additional instruction and practice provided for applying the word study and fluency skills and strategies learned in Phase I. Phase II lessons occurred over a period of 17–18 weeks, depending on students' progress. Word Study and Vocabulary were taught through daily review of the word study strategies learned in Phase I by applying the sounds and strategies

to reading new vocabulary words. After reading words, students were provided with basic definitions for each word and then identified the appropriate word to match various scenarios, examples, or descriptions. In addition, students were introduced to word relatives and parts of speech (e.g., *politics, politician, politically*). Finally, students reviewed application of word study to spelling words. Vocabulary words for instruction were chosen from the text read in the fluency and comprehension component. Interventionists used *REWARDS Plus* Social Studies lessons and materials (Archer, Gleason, & Vachon, 2005b) 3 days each week, and used novels with lesson- developed by the researchers the remaining 2 days each week. Fluency and Comprehension were taught 3 days a week by reading and providing comprehension instruction with expository social studies text (REWARDS Plus; Archer el al., 2005b) and 2 days a week by reading and comprehending narrative text in novels. Students then read the text at least twice for fluency. Students received explicit instruction in generating questions of varying levels of complexity and abstraction while reading (e.g., literal questions, questions requiring students to synthesize information from text, and questions requiring students to apply background knowledge to information in text); identifying main idea: summarizing; and using strategies for multiple-choice, short-answer, and essay questions.

Phase III continued over approximately 8–10 weeks and maintained the instructional emphasis on vocabulary and comprehension. Word Study and Vocabulary in Phase III were identical to Phase II. However, interventionists used fluency and word reading activities and novel units developed by the research team. Fluency and Comprehension were taught through application of strategies for reading and understanding text to both expository science and social studies content and narrative text (novels), with a focus on applying the strategies to independent reading. Students read passages twice for fluency, generated questions while reading, and addressed comprehension questions related to all the skills and strategies learned (e.g., multiple choice, main idea, summarizing, literal information, synthesizing questions, background knowledge, and so on) independently before discussion.

Intervention implementation

Nine interventionists (six female) provided the intervention to students in groups of 10–15 for approximately 50 min per school day from September through May. All interventionists had at least an undergraduate degree, and seven interventionists had a master's degree. Seven of the nine had leaching certification in a reading or a reading-related area such as English/language arts; in addition, one interventionist was certified in English as a second language. Interventionists had between 2 and 39 years of experience, with a median of 13 years ($M = 14.2$; $SD = 12.0$).

The research team provided the interventionists with approximately 60 hr of professional development prior to teaching. This training included sessions related to the standardized intervention, the needs of the adolescent struggling reader, and principles of promoting active engagement in the classroom as well as other features of effective instruction and behavior management. They also received an additional 9 hr of professional development related to the intervention throughout the year and participated in biweekly staff development meetings with ongoing on-site feedback and coaching (once every 2–3 weeks).

Students in the three schools from the larger site received a reading class and an English/ language arts class. The length of these classes was 45 min per school day in two schools and 85 min every other school day for the third. Students attending one of the four schools from the smaller site received an English/language arts class for 50 min per school day. None of the schools in the smaller site offered an additional reading class to all students. Thus, sixth-grade students in the larger site received an additional reading-related class relative to students in the smaller site. Intervention classes were provided during that time in

which the students assigned to treatment condition typically would have received an elective. The average total amount of research intervention received for the 212 students in Tier 2 intervention classes at the end of the year was 99.6 hr (SD = 23.1, range 20.3–126.8, median = 109.9) at the large site and 111 .3 hr (SD = 11 .6, range 60.0–126.7, median = 111.7) at the smaller site.

Student schedules were obtained in December and May to determine if the struggling readers were receiving any additional reading intervention. Of the 212 Tier 2 students, 160 (75%) reported receiving no additional instruction and the information was unavailable for 3 (1%) of the students. Thus, 49 (23%) students received additional intervention, 47 (22%) of which received one additional type of instruction, and two students (1%) received two additional interventions. The type of supplemental instruction varied, but included individual tutoring, resource classes, and state accountability test tutoring. Additional instruction was nearly always delivered by certified interventionists, typically in groups of larger than 10 students. The average amount of time of additional instruction for the 49 Tier 2 students who received it was 107.7 hr (SD = 39.6, range 37.5–141.7, median = 138.8), and for the total group the average amount of time of additional instruction was 24.90 hr (SD = 49.3). Of the 115 comparison struggling readers, 59 (51%) received no additional instruction, 46 (40%) received one additional type of instruction, and 10 (14%) received two. The average amount of time of additional instruction for the 56 students who received it was 140.6 hr (SD = 72.9, range 27.0–283.3, median = 141.7), and for the total group of students the average amount of time of additional instruction was 684 hr (SD = 86.8). A greater proportion of students in the comparison condition (49%) received additional instruction relative to those in researcher-provided Tier 2 (23%), $p < .05$. The proportion of students receiving additional instruction was similar across sites ($p > .05$).

Intervention Fidelity

Project coordinators from the research team observed each interventionist two to three times each month and provided feedback on implementation. Fidelity data were collected throughout the year for each interventionist on up to 5 different instructional days (median = 4; approximately 2%–3% of sessions). Two observers monitored fidelity and consistency of intervention implementation and rotated each month so that both observers saw every interventionist at least once.

Fidelity was coded by rating each of the instructional components on a 3-point Likert-type rating scale ranging from I (*low*) to 3 (*high* see www.meadowscenter.com for a copy of the intervention code sheet). Quality of implementation (e.g., active engagement, frequent opportunity for students' responses, appropriate use of feedback and pacing) was rated on the same 3-point Likert-type scale for each of the instructional components. A score of 3 (excellent) was coded when the interventionist completed all or nearly all of the required elements and procedures. A score of 2 (adequate) was coded when most of the required elements and procedures were completed, A score of 1 was coded if less than half of the required elements and procedures were completed for a given component of the lesson. If an interventionist did not include a required component, then a score of 0 was given.

The mean implementation score across components and across interventionists was 2,53 (SD = 0.32. range 2.00–2.93). The mean quality score across components and across interventionists was 2.68 (SD = 0.30, range 2.00–3.00). The mean total fidelity ranking was 2.60 (SD = 0.27, range 2.00–2.82).

Analyses

Data preparation first involved the evaluation of distributional data both statistically and graphically for skewness, kurtosis, and normality. At pretest, 6 of 11 variables exhibited skewness or kurtosis estimates that exceeded 1.00. However, one case in the comparison group was a (low) multivariate outlier (>3 *SD*) without this case, distributions were significantly improved for three of the six variables, and the remainder were somewhat improved. A similar though less extreme pattern was noted for post-test distributions. Analyses were generally similar with and without this case, but they had undue influence for several measures, so were excluded from the remainder of analyses reported here.

For measures with only two time points, the primary models were analysis of covariance, with the post-test scoring being the de pendent variable and the pretest score the covariate. Measures with several time points were analyzed with growth models that were fit to evaluate performance trajectory. These models generally did not alter the pattern of results substantively, so were not further reported. Most analyses compared the Tier 2 and comparison groups, because students were randomized to these groups. However, performance for typical readers at pretest and post-test is included in tables for visual reference to provide some context regarding the classroom peers of these struggling readers, and the extent to which the gap between struggling and typical readers was closed.

Each model was then extended in multiple ways to consider site as well as covariates/moderators (e.g., age, additional instruction, intervention time, fidelity, group size) and their interactions. The nested structure of the data were also considered, These factors in general were evaluated whenever the primary models showed statistically significant treatment effects, or where the raw effect size was greater than +0.15, favoring the group of students who received Tier 2. We were interested in whether this additional instruction moderated the treatment effect. Additional instruction in the whole sample was modestly related to outcomes (median $r = .20$). Fidelity, intervention time, and group size were evaluated within the context of the Tier 2 treatment students only because these measures were not relevant for other students. Specifically, we were interested in whether (high) fidelity, (more) intervention time, and (smaller) group size were positively related to outcomes of interest. It was the case that the relationships of these variables to outcomes within (he group of students who received Tier 2 intervention were generally weak (median $r = .07$). This may be because fidelity was in general good, intervention time was high, and group size was generally large.

We considered nesting factors in multiple ways. We could not cluster by intervention tutor because there were only a small number of these, but we did group students by classroom reading teacher, by tutoring group for Tier 2 students, or by some combination. In general, the effect of clustering was low (below 10%) and not significant ($z > 0.05$). however clustering was arranged. The largest clustering effects were for measures of spelling and comprehension, particularly TAKS and GRADE, and with clustering according to intervention group. We constructed further models within SAS PROC MIXED to account for additional nesting when treatment effects were evidenced and nesting was significant.

Results

Pretest status is presented first, including additional potential covariates. Next are the primary results of the comparison between treatment and comparison students, considering only pretest performance: these are arranged according to the primary outcome target domains of decoding, comprehension, and fluency. Both inferential statistics and effect size indices are provided, at an alpha level of $p < .05$. Clearly, correcting for multiple

comparisons would reduce the number of significant results, although given the scarcity of data at this age range for randomized intervention studies, we felt it important to highlight what might be potential effects in future studies. It was for this reason that the effect sizes (and their confidence intervals) are also provided. Finally, a variety of follow-up analyses are presented in the next section, to evaluate the potential role of moderators and other factors.

Pretest

Descriptive data, statistical results, and unadjusted effect sizes are presented in Table 1. Performance is presented by group, and variables are organized into measures of decoding and spelling, comprehension, and fluency. Struggling readers in Tier 2 outperformed those in the comparison group on the TOWRE Sight Word Efficiency measure. $F(1.299) = 4.25.p < .04$, and PF, $F(1,307) = 5.27$, $p < .03$, although not on any other pretest measure ($p > .05$). Sites differed in terms of performance on GRADE, TOSRE, TOWRE Phonemic Decoding Efficiency, and WJ-III Letter-Word Identification, Word Attack, and Passage Comprehension (all p values < .05), with students in the smaller city outperforming students in the larger site or city in every case. Moreover, age was negatively related to all outcomes (all p values <.05, range −.24 to −.55, median = −0.30).

Effects of Intervention

Decoding and spelling

As shown in Table 1, there was a significant effect for pretest on the WJ-III Letter-Word Identification subtest, HI, 298) = 656.51, $p < .0001$, but not for treatment, $F(1,298) = 3.74$, $p < .054$, with an effect size of + 0.15. There was an interaction of pretest and treatment group for data from WJ-III Word Attack, $F(1,297) = 4.67$, $p < .0314$, $\eta_p^2 = .015$, as well as a significant main effect for treatment, $p < .009$, with an effect size of $d = +0.15$. The disordinal interaction (with slopes crossing within the observed score range) suggested a stronger overall relation of pretest and post-test scores in the group of students who were in Tier 2 relative to those in the comparison group. Further probing revealed that the treatment effect was not apparent at pretest values below the mean of the sample, but was present at pretest values at or above the mean. The WJ-III Spelling subtest was administered only at post-test, so the Letter-Word Identification was used as the pretest covariate. There was a significant interaction of treatment group with the covariate, $F(1,296) = 5.82$, $p < .0165$, $\eta_p^2 = .019$, as well as a significant main effect for treatment, $p < .0126$, with an effect size of $d = +0.22$, The disordinal interaction suggested a stronger overall relationship of covariate and post-test scores in the students who were in the comparison group, relative to those who received Her 2 intervention. Further probing revealed that the treatment effect was apparent at covariate values at or below the mean of the sample but were not present at covariate values above the mean.

Comprehension

The sample size for TAKS differs from all others because students who qualified for the state-developed alternative assessment in either year are not represented in these analyses. There were significant effects for pretest (assessed in the spring of 2006), $F(1.279) = 114.85$, $p < .0001$, for TAKS data, but not for treatment, $F(1.279) = 1.92$, $d = +0.18$. Because TAKS is a criterion measure, we also evaluated the extent to which treatment differentially increased the chances of passing the test from year to year. There were no differences among treatment groups either among students who met "bubble" criterion prior to intervention, $\chi^2(df = 1, N = 91) < 1.00$, or among those who did not pass the TAKS prior to intervention, $\chi^2(df = 2, TV = 195) = 1.22$, $p > .05$. In the former case, most students (89%) in both comparison and Tier 2 groups continued to meet TAKS benchmark criteria: in the

latter, a majority of students (61%) met TAKS criteria. The proportion of comparison and Tier 2 students did not differ in this regard.

There was a main effect of pretest for GRADE comprehension data, $F(1,323) = 97.96$, $p < .0001$, but not for treatment, $F(1,323) < 1.00$, with a negligible effect size. On the **WJ-III** Passage Comprehension subtest, there was a significant main effect for pretest, $F(1,298) = 588.64$, $p < .0001$, but the main effect for treatment group was not significant, $F(1,298) = 3.26$, $p < .072$, with an effect size of d +0.19.

Fluency

There was a significant main effect for pretest with TOWRE Sight Word Efficiency data, $F(1,298) = 410.64$, $p < .0001$, but not for treatment, $F(I,298) = 1.93$, $p > .05$. The effect size was $d = +0.30$, although as earlier noted, the groups differed similarly at pretest (d - +0.25), so it is unlikely that this difference represents a treatment effect. There was a significant main effect for pretest with TOWRE Phonemic Decoding Efficiency Test data, $F(1,298) = 436.09$, $p < .0001$, but the main effect for treatment group was not significant, $F(1,298) = 3.29$, $p < .071$, with $d = +0.19$.

The remaining fluency measures were each administered on five occasions. For AIM-Sweb Mazes, there was a significant effect for pretest, $F(1,322) = 164.68$. $p < .0001$, although not for treatment group, $F(1,322) < 1$, with a negligible effect size. Similar results were evidenced for the TOSRE (pretest: $F[l, 322] = 241.71$, $p<.0001$: treatment: $F[1,322] < 1.00$, d +0.13), WLF (pretest: $F[1,306] = 942,50$, $p < .0001$; treatment: $F [1,306] = 2.38$, $p > .05$, d +0.14), and for PF (pretest: $F[l, 306] = 778.83$, $p <.0001$: treatment: $F[1.306] < 1.00$). An effect size of $d = +0.24$ was found for PF data, although as was the case with TOWRE Sight Word Efficiency, the groups differed at pretest as well ($d = +0.27$).

Typical Students

Table 2 presents performances of the typical readers. With the exception of the TOSRE, standard scores showed little change from pretest to post-test, with performances very near the normative average. For both standard score measures and raw score measures, gains within the group of typical students were comparable to those evidenced in the comparison group. Effect size differences of students who received Tier 2 and typical students are also presented. Table 2 shows that typical students outperformed those who received Tier 2 intervention on all measures, and similarly so at pretest (median d —0.95) and posttest (median $d = -0.94$).

Follow-Up Post-Test Results

Follow-up results considered the relation of the primary results when site and age were added to models. Where relevant, moderators particular to the students who received Tier 2 intervention (additional instruction, intervention time, fidelity, group size), and the nested structure of the data, were also considered. In addition, interactions among these variables were considered.

Adding site and age did not change the interpretation of the primary results discussed earlier for the comprehension measures from the GRADE, and for the fluency measures from the Mazes TOSRE, TOWRE Sight Word Efficiency, and WLF. On some occasions, these variables exerted main effects, but there were no interactions, and the treatment effects did not change. Interactions of site and/or age with treatment were noted for the remaining measures. The most complex results were evident for WJ-III Word Attack, where there was a significant four-way interaction of pretest, age, site, and treatment, $F(2,285) = 8.56$, $p < .0002$, $\eta_p^2 = .050$; the treatment effect remained, $p < .036$, Follow-up analyses within site

revealed that in the larger site, neither age nor treatment group was significant over pretest (both $P > .05$). In the smaller site, a three-way interaction of pretest, age, and treatment was evidenced, F(2,146) = 7.04, $p < .0012$, $\eta_p^2 = .088$, and the treatment main effect remained, $p < .0013$: specifically, students in Tier 2 outperformed those in comparison, when pretest means were at or above the mean of the sample, at ages at or below the mean. Similar results (the same four-way interaction) were obtained when additional school-provided instruction was included in the models. Within the group of students in Tier 2, neither group size, total intervention lime, nor additional instruction time was significantly related to outcomes. The overall pattern indicated that Tier 2 students in the smaller site made gains relative to comparison students, particularly where students were younger and had higher pretest scores.

Three-way interactions of age, site, and treatment were noted for WJ-III Letter-Word Identification, F(2.287) = 3.68, $p < .026$, $\eta_p^2 = .015$; for TAKS, $F(2,264) = 3.90$, $p < .0213$, $\eta_p^2 = .021$; and for WJ-III Passage Comprehension, F(2,287) = 5.72, $p < .004$, $\eta_p^2 = .038$. Although significant treatment effects were not observed for these three measures in the primary analyses in the context of effect sizes in the range of + 0,15 to + 0.19, the interactive effects suggest effects for subgroups of students on some measures.

For Letter-Word Identification, the effect for age indicated that age was negatively related to performance (younger students outperformed those who were older), $p < .0001$, though only in the smaller site. The pattern did not change with the inclusion of additional instruction to the model, and within the group of students in Tier 2, neither group size nor total intervention time was significantly related to outcomes. For TAKS, follow-up revealed that in the larger site, a main effect of treatment was observed, $p < .003$ (but not for age): similarly, in the larger site, students who received Tier 2 were more likely to pass (76%) TAKS than those in the comparison group (57%), $\chi^2(df = 1, N = 96) = 4.03$, $p < .05$. In the smaller site, there was a main effect for age, $p < .0001$, but not for treatment. Consideration of other components (additional instruction, nesting structure, and the role of group size and instruction time in the treatment groups) did not change these results.

For Passage Comprehension, follow-up within site revealed that in both sites there were interactions of age and treatment (larger site: F[1,141] = 4.78, $p < .031$, $\eta_p^2 = .033$; smaller site: F[1.148] = 4.32, $p < .04$, $\eta_p^2 = .028$); however, the effects were in different directions. *Older* students who received Tier 2 from the smaller site outperformed those in the comparison group, whereas *younger* students who received Tier 2 from the larger site outperformed those in the comparison group. Including additional instruction did not change results for the smaller site. In the larger site, there was a four-way interaction of site, age, additional instruction, and treatment, F(2,271) = 3.68, $p < .027$, $\eta_p^2 = .026$, with follow-up suggesting that students who received Tier 2 outperformed those in the comparison group when the amount of additional instruction was small *and* students were either younger or scored lower at pretest. Considering other variables such as group size, total intervention time, or cluster did not alter conclusions.

For WJ-III Spelling and for PF, there were interactions of site and treatment, $F(1,292) = 7.90$, $p < .006$, $\eta_p^2 = .026$ and F(,293) = 8.01, $p < .0050$, $\eta_p^2 = .027$, respectively. When additional school-provided instruction was included in the models for spelling, there was a significant four-way interaction of additional instruction, covariate, site, and treatment, $F(2,270) = 4.42$, $p < .013$, $\eta_p^2 = .032$. In the larger site, the effect of treatment was not significant, and in the smaller site, there were interactions of the covariate and treatment, F(1,146) = 5.96.$p < .016$.η_p^2.039 (as in the primary analyses), and of additional instruction and treatment, $F(1,146) = 8.54$. $p < .004$, $\eta_p^2 = .055$. These interactions suggested that in the smaller site. Tier 2 students outperformed those in the comparison group when covariate

performances were low or with low levels of additional instruction. Adding the nesting component to models did not substantively alter conclusions. Within the group of students in Tier 2, group size was not significantly related to outcomes, but total intervention lime was positively related to outcomes, and additional instruction time was negatively related to outcomes. Follow-up did not reveal any significant differences between Tier 2 and comparison students in PF data, but students in the larger site who received Tier 2 intervention outperformed those in the smaller site who received Tier 2 intervention, $p < .002$.

Finally, no interactions were noted for TOWRE Phonemic Decoding Efficiency when only site and age were considered, but when additional instruction was also considered, there was a three-way interaction of pretest, site, and treatment, $F(1,279) = 3,39. p <.036, \eta_p^2 .024$. There were no significant effects in the larger site, but there was a significant interaction of pretest and treatment in the smaller site, $F(1,146) = 4.23, p <.0415, \eta_p^2 = .028$. The disordinal interaction suggested a stronger overall relationship of pretest and post-test scores in the group of students who were in Tier 2, relative to those in the comparison group. Further probing revealed that the treatment effect was not apparent at pretest values below the mean of the sample, but were present at pretest values at or above the mean. Within the group of students in Tier 2. group size, additional instruction time, and total intervention time were not significantly related to outcomes. The overall pattern indicated that Tier 2 students in the smaller site made gains relative to comparison students, particularly where students had higher pretest scores.

Discussion

We evaluated the effectiveness of a large-scale middle school reading intervention provided within the context of a response to intervention framework for struggling readers in sixth grade. As expected, students who received Tier 2 intervention outperformed those in the comparison condition on several measures, including word attack, spelling, comprehension, and phonemic decoding efficiency. In most cases, gains were small and were more apparent in particular subgroups of students (at a given site or at certain levels of pretest performance or age). Other relations involving treatment group were noted for sight word fluency and passage fluency, but it was difficult to attribute differential gains for these outcome measures directly to the intervention. Measures that tapped both fluency and comprehension did not reveal treatment effects. Except in two instances (spelling in the smaller site and passage comprehension in the larger site), the pattern of these results did not change substantively according to whether additional instruction was delivered. Within the students who received researcher-provided Tier 2 intervention, group size, time in treatment, and additional instruction did not substantially relate to outcome achievement.

The findings from this study revealed that the goal of closing the gap between at-risk sixth-grade students who received Tier 2 intervention and students not at risk in the beginning of the school year may be overly ambitious. Findings for intervention students were positive but did not change substantially over the course of the year. On the other hand, performance did not decline over the course of the school year. It was clear from evaluation of pretest to post-test means of raw score data within groups (Table 2) that students' proficiency increased in these domains. We consider several factors that are likely to have affected these results.

Although much is known about effective instruction to assist young students' transition from nonreaders to readers, less is known about how to effectively remediate struggling readers at the secondary level. This disparity is likely to be particularly true for older readers who are from high-poverty, low-resource settings, Edmonds et al. (2009) and Scammacca et al.

(2007) found overall intervention effects that were higher than those revealed in the current study, even when considering only standardized measures in the outcomes. Although it is not entirely clear why the findings of this study differ from the findings presented in those syntheses, the current study's sample is much larger than most intervention studies, which tends to attenuate effect sizes. In addition, three critical features of the current study differ from much of the previous research: (a) the duration of the intervention (Tier 2), (b) the instruction provided to the Tier 2 comparison group, and (c) the context of an response to intervention framework that provided enhanced Tier 1 instruction to all students.

Most of the interventions synthesized by Edmonds et al. (2009) and Scammacca et al. (2007) were provided for less than 2 months. It is a frequent finding in intervention research that interventions of shorter duration report higher effects than interventions of longer duration (Elbaum, Vaughn, Hughes, & Moody, 2000). perhaps because of an initial boost in learning from the addition of instruction or even the novelty of the intervention. Two previous intervention studies with older students provided more extensive interventions (70 sessions by Anderson, Chan, & Henne, 1995, and 80 by Bryant et al., 2000). The other studies ranged from 2 to 40 sessions, considerably less than the amount of intervention provided in the current study (more than 150 sessions). None of the studies were large-scale, randomized studies that took all eligible students in the schools. We think that this large-scale, long, and school-based model is likely to be associated with lower effects, as are many of the efficacy studies reported by the Institute for Education Sciences (e.g., Kemple et al., 2008).

This study of middle school students with reading difficulties is the first to be conducted within the context of a response to intervention framework in which all students (treatment and comparison) are provided instructional enhancements. To clarify, Tier 1 professional development aimed at enhancing vocabulary, word reading, and comprehension was provided to all content area teachers serving all sixth-grade students, with the goal of enhancing instructional practices in reading for all students. Although struggling readers who were randomly assigned to the comparison group did not receive the Tier 2 intervention from the research team, all students in all classrooms may have benefited from professional development introduced to their content area teachers. In an effort to align key instruction for students, some of the same vocabulary and comprehension strategies that were taught in the Tier 2 intervention were provided to the content area teachers for use with all students, as applicable to each specific content area. Thus, it may be that some of the gains made by both the Tier 2 treatment group and the comparison group in "closing the gap" with typical students could be from this professional development. Because we could not study the effects of the Tier 1 professional development separately, we are unable to confidently determine the relative effects of this element of the treatment.

Many of the comparison struggling readers also received additional reading instruction beyond their content area classes. Because of the increased pressure for accountability, schools are motivated to improve the poor performance of their students who struggle to read and to show progress on state accountability tests. Thus, many of these students were targeted for extensive assistance within the context of their classroom experience. Further, in the larger site, all students received an additional reading class. The results of this study may also indicate that the Tier 2 intervention provided was as effective as the reading interventions provided to the comparison students.

Results from a recent report on The Enhanced Reading Opportunities (ERO) Study (Kemple et al., 2008) provide a more positive perspective regarding the findings of this study. The ERO study gave supplemental literacy instruction for one period each day for a full year to a very large sample ($N = 2.916$) of ninth-grade students who were performing 2 or more years

below grade level. Students were randomly assigned to one of three groups (two intervention approaches): a '"flexible fidelity" individualized intervention, a standardized intervention, or school practice. Therefore, the standardized intervention and school practice groups in ERO were similar to the Tier 2 groups provided in this study, with the distinction that our groups were also provided with enhanced classroom instruction, likely weakening the effects of the Tier 2 intervention.

Both ERO treatments produced a 0.9 standard score point increase on the GRADE reading comprehension subtest. This corresponds to an effect size of 0.09 standard deviation and is statistically significant. A larger effect size was found on the GRADE comprehension subtest for the current study ($d = 0.17$), although this was not statistically significant. Despite the fact that the ERO study had a larger sample and the flexibility to respond to student need in the ""flexible fidelity" problem-solving protocol group, the current study still produced higher effects on the same standardized measure of comprehension. Furthermore, the comparison group for this study was not strictly a comparison group but instead a comparison group that was provided enhanced instruction through Tier 1, thus providing a more rigorous test of the treatment. Neither our study nor the ERO study yielded results suggesting that a Tier 2 intervention provided over 1 school year was robustly effective, especially in terms of closing the gap relative to typically achieving peers.

Most students in this study were able to decode text at a basic level, but the efficiency with which they read and comprehended text varied widely. Although students in this study had varying needs, the intervention package designed for the current study was a standardized intervention that addressed basic decoding skills, fluency, and comprehension; it was aimed at meeting the needs of the group and with less focus on individualization or responsiveness to students' specific needs. In addition, the fact that these students were selected based on low performance after many previous years of instruction indicates that the group as a whole may be less amenable to change—at least over I school year. Expecting significant growth as a result of a yearlong intervention with little flexibility for responding to individual needs in medium-sized groups for students who have continued to struggle into the upper grades may have been overly ambitious.

It is practically and logistically difficult to implement an intensive reading intervention in the high-poverty middle school settings where this study was conducted, which may have contributed to the results. In some schools, it was difficult to arrange students' schedules to allow students to consistently attend intervention. Thus, although the intervention was implemented with fidelity, the composition of the reading groups continuously changed. There were students who began the intervention class late (e.g., because of difficulties in scheduling), changed intervention classes (e.g., because of schools changing students' schedules), and exited intervention early (e.g., because of attrition), interrupting the "flow" of the intervention. These changes represent typical circumstances in middle schools, but may have affected the power of the intervention.

Implications

Determining effective school-wide practices that will influence outcomes for students with significant difficulties is challenging for most school psychologists. The model we implemented provides a framework to consider for implementing school-wide reading g practices across content areas and an instructional reading class for students with more extreme difficulties. Of course, it may be necessary to provide even more intensive intervention for some students (e.g., longer time, smaller groups, intervention more specifically focused to meet students' needs).

From this study, we can form hypotheses about effective ways to remediate this population. One area of further study is the intensity of intervention these struggling students may need. We are currently investigating Tier 3 interventions for the students in this study who demonstrated minimal response to the sixth-grade intervention. We have identified the minimal responders from the sixthgrade intervention and are providing them an additional year of intervention (currently underway). The Tier 3 intervention is being provided in small groups of 3–5 students, increasing the intensity of the intervention. We are also examining the effects of continued standardized intervention (as reported in this study) in these smaller groups versus a more individualized, responsive approach to remediation. In the individualized intervention, interventionists have more flexibility with respect to materials selection, lesson structure, and overall application of instruction to respond to the varying needs of the students in their small groups.

The findings from this large-scale intervention raise questions about the extent of the effect based on the expenditure of implementing the intervention. The question of whether this intervention is worth the cost is a difficult question to answer. Based on the small effect sizes resulting from this study, it seems reasonable to argue that using resources to focus on enhancing Tier 1 and perhaps even more intensive interventions for students with severe reading problems (Tier 3) may be worth evaluating. However, our confidence in this recommendation is limited to the data at the end of one year of treatment and it is not possible based on the study design to adequately determine the overall effect of the Tier 1 (primary) intervention. The effects on the students over time with respect to reading and even dropout prevention may be worth examining before policy recommendations are forthcoming. This study also raises questions to policy makers about how to provide school supports and what kind of outcomes to expect. For example, perhaps it is unreasonable to expect that students who have been significantly behind for many years would compensate and close the gap by being provided one 50-min reading class a day. It may be that significantly more intensive interventions and perhaps even different types of interventions are necessary to achieve this outcome.

Limitations

As might be expected in any school-based intervention study, there were several limitations of the current data. First, some of the control students received secondary intervention by the schools. Although we were able to document and examine the relative effects from these students assigned to a secondary intervention, none of the students in the comparison group would have received a secondary intervention. This intervention was not provided by the research team and instead by the school district. Second, the lack of flexibility and movement for participating students between Tiers 2 and 3 could also be a limitation of the data. Based on multitiered intervention approaches at the early elementary level, students are typically moved between tiers, determined by their progress. In this study, we were interested in examining the relative effects of a Tier 1 intervention with and without a Tier 2 intervention. This allowed us to make clearer causal claims about these interventions. If students were allowed to move between tiers, we would have increased the number of groups in the study and application.

In summary, we responded to the need for additional research on secondary students with reading difficulties by designing and implementing a multiyear intervention. Findings from the first year of the intervention examining the relative effects of Tier 1 plus Tier 2 compared with Tier t alone are described in this article. Research with secondary students and response to intervention is limited, so we focused on two of the critical elements of response to intervention primary (Tier 1) and secondary (Tier 2) instruction. However, our

data suggest that additional research is needed before policy implications can be confidently derived.

Acknowledgments

References

Anderson, V.; Chan, KK.; Henne, R. The effects of strategy instruction on the literacy models and performance of reading and writing delayed middle school students. In: Hinchman, KA.; Leu, DJ.; Kinzer, CK., editors. Perspectives on literacy research and practice: Forty-fourth yearbook of the National Reading Conference. Chicago: National Reading Conference; 1995. p. 180-1891.

Archer, AL.; Gleason, MM.; Vachon, V. REWARDS intermediate: Multisyllabic word reading strategies. Longmont, CO: Sopris West; 2005a.

Archer, AL.; Gleason, M.; Vachon, V. REWARDS plus: Reading strategies applied to social studies passages. Longmont, CO: Sopns West; 2005b.

Beck, IL.; McKeown, MG.; Kucan, L. Bringing words to life, Robust vocabulary instruction. Upper Saddle River, NJ: Pearson; 2002.

Bryant DP, Vaughn S, Linan-Thomason S, Ugel N, Hamff A, Hougen M. Reading outcomes tor students with and without reading disabilities in general education middle-school content area classes. Learning Disabilities Quarterly. 2000; 23:238–252.

Denton, C.; Bryan, D.; Wexler, J.; Reed, D.; Vaughn, S. Effective instruction for middle school students with reading difficulties: The reading teacher's sourcebook. Austin: University of Texas System/Texas Education Agency; 2007.

Edmonds MS, Vaughn S, Wexler J, Reutebuch CK, Cable A, Tackett K, et al. A synthesis of reading interventions and effects on reading outcomes for older struggling readers. Review of Educational Research. 2009; 79:262–300. [PubMed: 20072704]

Elbaum B, Vaughn S, Hughes MT, Moody SW. How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. Journal of Educational Psvchologv. 2000; 92:605–619.

Fletcher, JM.; Lyon, GR.; Fuchs, LS.; Barnes, MA. Learning disabilities: From identification to intervention. New York: Guilford Press; 2007.

Jimerson, S.; Bums, MK.; VanDerHeyden, A. Handbook of response to intervention: The science and practice of assessment and intervention. New York: Springer; 2007.

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

Kaufman, AS.; Kaufman, NL. Kaufman Brief Intelligence Test, Second Edition (K-BIT-2). Minneapolis, MN: Pearson Assessment; 2004.

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

Lennon, C.; Burdick, H. The Lexile Framework as an approach for reading measurement and success. 2004 [Retrieved January 7, 2010]. http://www.lexile.com/research/1/

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

Shinn, MR.; Shinn, MM. AlMSweb training workbook: Administration and scoring of reading maze for use in general outcome measurement. Eden Prairie, MN: Edformation; 2002.

Texas Education Agency. TAKS: Texas Assessment of Knowledge and Skills, Information booklet: Reading, grade 7— Revised. 2004 [Retrieved December 19, 2007]. http://www.tea.state.tx.us/student.assess-ment/taks/booklets/reading/g6e.pdf

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

Wagner, RK.; Torgesen, JK.; Rashotte, CA.; Pearson, NA. Test of sentence reading efficiency (TOSRE). Austin, TX: PRO-ED; in press

Williams, KT. The group reading assessment and diagnostic evaluation (GRADE)Teacher's scoring and interpretive manual. Circle Pines, MN: American Guidance Service; 2001.

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

Biographies

Sharon Vaughn, PhD, holds the H E. Hartfelder/Southland Corp. Regents Chair in Human Development. She is the executive director of The Meadows Center for Preventing Educational Risk. She is the author of numerous books and research articles that address the reading and social outcomes of students with learning difficulties. She is currently the principal investigator or coprincipal investigator on several Institute for Education Science. National Institute for Child Health and Human Development, and Office of Special Education Programs research grants investigating effective interventions for students with reading difficulties and students who are English language learners.

Paul T. Cirino, PhD, is a developmental neuropsychologist whose interests include disorders of math and reading, executive function, and measurement. He is an associate professor in the Department of Psychology and the Texas Institute for Measurement, Evaluation and Statistics at the University of Houston.

Jeanne Wanzek, PhD. is an assistant professor in special education at Florida State University and on the research faculty at the Florida Center for Reading Research Her research interests include learning disabilities, reading, effective instruction, and response to intervention.

Jade Wexler, PhD. is a senior research associate at The Meadows Center for Preventing Educational Risk at The University of Texas at Austin. Her research interests are interventions for adolescents with reading difficulties, response to intervention, and dropout prevention.

Jack M. Fletcher, PhD, is a Hugh Roy and Lillie Cranz Cullen Distinguished Professor of Psychology at the University of Houston. For the past 30 years, as a child neuropsychologist, he has completed research on many issues related to reading, learning disabilities and dyslexia, including definition and classification, neurobiological correlates, and intervention. He is the principal investigator of a NICHD-funded Learning Disability Research Center. He was the 2003 recipient of the Samuel T. Orton award from the IDA and a corecipient of the Albert J. Harris award from the International Reading Association in 2006.

Carolyn Denton, PhD, is an associate professor in the Children's Learning Institute, part of the Department of Pediatrics at the University of Texas Health Science Center at Houston. Her research interests include reading intervention, response to intervention models, and coaching as a form of professional development.

Amy E. Barth, PhD, is an assistant research professor at the University of Houston. Texas Institute of Measurement, Evaluation, and Statistics, and Texas Center on Learning Disabilities. Her interests include the identification and treatment of students with language and learning disabilities.

Melissa A. Romain, PhD, is a research assistant professor at the University of Houston. Her primary research interests have been in the areas of early reading intervention and prevention of reading difficulties: reading intervention with middle school students; and school, teacher, and student variables associated with the effective inclusion of students with disabilities in general education classrooms. She is also coordinator of a new campus-wide community advancement initiative at the University of Houston.

David J. Francis, PhD, is the Hugh Roy and Lillie Cranz Cullen Professor and Chair of the Department of Psychology and Director of the Texas Institute for Measurement. Evaluation, and Statistics at the University of Houston. His research interests include applied psychometrics, latent variable and longitudinal models, developmental disabilities, reading and language acquisition, and educational evaluation.

Table 1

Pretest and Post-Test Performance on Reading Measures for Treatment Groups

Measure	Group	n	Pre M	Pre SD	FU M	FU SD	d	CI	F	p
Letter-Word Identification	COMP	106	92.80	12.0	94.35	12.1	+0.15	−0.09 to +0.38	3.74	.054
	Tier 2	195	93.04	11.3	96.06	11.4				
Word Attack[a]	COMP	106	96.67	10.8	96.44	9.5	+0.15	−0.08 to +0.39	6.85	.009
	Tier 2	195	96.01	10.2	98.00	10.4				
Spelling[a,b]	COMP	106	92.80	12.0	92.75	15.2	+0.22	−0.01 to +0.46	6.31	.013
	Tier 2	194	93.10	11.3	95.94	13.6				
TAKS	COMP	106	2,012.7	102.0	2,150.7	174.6	+0.18	−0.06 to +0.43	1.92	.167
	Tier 2	176	2,020.0	94.5	2,182.6	171.8				
Reading Comprehension	COMP	114	88.37	8.8	88.32	8.4	+.06	−0.16 to +0.29	<1	ns
	Tier 2	212	89.39	9.1	88.87	8.4				
Passage Comprehension	COMP	106	87.04	12.0	87.47	11.6	+0.19	−0.03 to +0.42	3.26	.072
	Tier 2	195	87.83	9.4	89.35	8.9				
Sight Word Efficiency	COMP	106	91.74	10.9	93.61	11.2	+0.30	+0.06 to +0.54	1.93	.166
	Tier 2	195	94.30	10.0	96.85	10.6				
Phonemic Decoding Efficiency	COMP	106	94.42	15.4	94.87	13.2	+ 0.19	−0.04 to +0.43	3.29	.071
	Tier 2	195	95.38	13.1	97.47	13.5				
Mazes	COMP	114	15.05	6.1	24.51	8.5	+ 0.09	−0.14 to +0.32	<1	ns
	Tier 2	211	16.12	6.3	25.32	9.2				
Sentence Reading Efficiency	COMP	114	85.25	11.7	91.20	12.7	+0.13	−0.10 to +0.36	<1	ns
	Tier 2	211	86.82	10.1	92.90	13.0				
Word List Fluency	COMP	109	67.83	28.3	77.65	28.7	+0.14	−0.09 to +0.37	2.38	.124
	Tier 2	200	69.21	22.6	81.28	24.6				
Passage Fluency	COMP	109	101.64	30.0	124.33	34.4	+0.24	+0.00 to +0.47	<1	ns
	Tier 2	200	109.68	29.1	131.95	31.0				

Note. COMP =comparison group; Tier 2 =intervention group: Pre M = mean of pretest performance: FU M = unadjusted mean of post-test performance: *d* effect size difference between COMP and Tier 2 at post-test, based on unadjusted means; CI= confidence interval of effect size; *F* = statistical test of the difference between adjusted means of COMP and Tier 2 subgroups (*df* [1,*N*-3], where *N* is comp +Tier 2); Letter-Word Identification, Word Attack, Spelling, and Passage Comprehension =subtests of the Woodcock-Johnson—III Tests of Academic Achievement, where reported means are standard scores; TAKS = Texas Assessment of Knowledge and Skills, the state accountability measure, where means are scaled scores (a score of 2,100 is passing): Reading Comprehension = subtest of the Group Reading Assessment and Diagnostic Evaluation where reported means are standard scores; Sight Word Efficiency and Phonemic Decoding Efficiency - subtests of the Test of Word Reading Efficiency, reported in standard scores: Mazes = softest from AIMSweb, reported in terms of the number of target words correctly identified in 3 min; Sentence Reading Efficiency = Test of Sentence Reading Efficiency, where

scores are reported in standard scores; Word List Fluency and Passage Fluency = investigator-created measures, reported in terms of words correctly read in 1 min, equaled for form effects, and averaged over all the stories read (five passages were read at pretest, three passages were read at post-test, and three word lists were read at both pretest and post-test).

[a] F values are for the (centered) treatment effect in the context of the observed significant interaction of pretest and treatment group.

[b] Pretest means are for the covariate utilized (Letter-Word Identification standard score). Interactions involving site, age, and additional instruction are not reflected above, but are reported in text.

Table 2

Pretest and Post-Test Performance on Reading Measures for Typical Students

Measure	n	Pre M	Pre SD	FU M	FU SD	d Pre	d Post	CI (Post)
Letter-Word Identification	231	106.34	13.0	107.64	12.3	−1.08	−0.97	−1.17 to −0.77
Word Attack[a]	231	103.87	10.9	105.78	11.5	−0.74	−0.71	−0.90 to −0.51
Spelling[a,b]	231	—	—	108.66	12.2	—	−0.99	−1.19 to −0.79
TAKS	243	2,298.2	133.5	2,411.7	213.1	−2.34	−1.16	−1.37 to −0.95
Reading Comprehension	248	101.25	11.4	101.53	12.5	−1.14	−1.17	−1.37 to −0.97
Passage Comprehension	231	99.16	11.0	100.32	10.0	−1.10	−1.15	−1.36 to −0.94
Sight Word Efficiency	231	102.90	12.4	105.58	12.3	−0.76	−0.76	−0.95 to −0.56
Phonemic Decoding Efficiency	231	105.14	14.3	106.49	13.6	−0.71	−0.66	−0.86 to −0.47
Mazes	247	23.30	8.4	34.17	10.4	−0.95	−0.90	−1.09 to −0.70
Sentence Reading Efficiency	247	100.04	12.5	109.49	16.9	1.15	−1.09	−1.28 to −0.89
Word List Fluency	233	86.64	24.2	96.34	26.3	−0.74	−0.59	−0.78 to −0.40
Passage Fluency	233	138.48	31.3	162.39	36.1	−0.95	−0.90	−1.10 to −0.70

Note. Awe, Pre *M* mean of pretest performance: FU M = unadjusted mean of posttest performance: CI = confidence interval: Letter-Word Identification, Word Attack. Spelling, and Passage Comprehension = subtests of the Woodcock Johnson—III Tests of Academic Achievement, where reported means are standard scores: TAKS - Texas Assessment of Knowledge and Skills, the state accountability measure, where means are scaled scores (a score of 2.100 is passing); Reading Comprehension = subtest of the Group Reading Assessment and Diagnostic Evaluation where reported means are standard scores: Sight Word Efficiency and Phonemic Decoding Efficiency = subtests of the Test of Word Reading Efficiency, reported in standard scores; Mazes = subtest from AIMSweb, reported in terms of the number of target words correctly identified in 3 min: Sentence Reading Efficiency = Test of Sentence Reading Efficiency, where scores are reported in standard scores; Word List Fluency and Passage Fluency = investigator-created measures and are reported in terms of words correctly read in 1 min, equated for form effects, and averaged over all the stories read (five passages were read at pretest, three passages were read at post-test, and three word lists were read at both pretest and post-test): Tier 2 = intervention group The *d* values presented are for difference at pretest and at post-test between students who received Tier 2 versus those in the typical group, with negative values reflecting means that favor the typical group (CIs provided for post-test only). Means and standard deviations for students in Tier 2 are presented in Table 1.

[a]*F* values are for the (centered) treatment effect in the context of the observed significant interaction of pretest and treatment group.

[b]Pretest means are for the covariate utilized (Letter-Word Identification standard score) Interactions involving site, age, and additional instruction are not reflected above, but are reported in text.